

METODOLOGÍA Y ESTADÍSTICA AVANZADA EN LAS CIENCIAS DEL MOVIMIENTO HUMANO

José Moncada Jiménez
Pedro Carazo Vargas
Jessenia Hernández Elizondo
Isaura Castillo Hernández
Gerardo Araya Vargas



METODOLOGÍA Y ESTADÍSTICA AVANZADA EN LAS CIENCIAS DEL MOVIMIENTO HUMANO

José Moncada Jiménez, Ph.D.
Pedro Carazo Vargas, Ph.D.
Jessenia Hernández Elizondo, Ph.D.
Isaura Castillo Hernández, Ph.D.
Gerardo Araya Vargas, Ed.D.



**UNIVERSIDAD DE
COSTA RICA**

Licencia de Creative Commons: Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional



© Derechos reservados de los autores.

Las fotografías que ilustran la portada del libro y los capítulos 1 y 2 fueron proporcionadas gratuitamente por la empresa Pixabay (<https://pixabay.com/>), bajo su licencia "[Pixabay License](#)" (Free for commercial use, No attribution required). La fotografía de la contraportada fue cedida gratuitamente por el M.Sc. Manrique Rodríguez Campos.

370.78
M737m

Moncada Jiménez, José
Carazo Vargas, Pedro
Hernández Elizondo, Jessenia
Castillo Hernández, Isaura
Araya Vargas, Gerardo

Metodología y estadística avanzada en las Ciencias del Movimiento Humano. / José Moncada Jiménez; Pedro Carazo Vargas; Jessenia Hernández Elizondo; Isaura Castillo Hernández; Gerardo Araya Vargas. – 1 ed. digital –. San José, Costa Rica: 2023.
315 p. : Recurso en línea: ilus.

ISBN: 978-9968-03-198-1

Contiene: Archivo de texto PDF – Gráficos – Edición digital
<INVESTIGACIÓN> <MOVIMIENTO HUMANO> <EDUCACIÓN FÍSICA> <ESTADÍSTICA> <METODOLOGÍA>
<DATOS ESTADÍSTICOS>

Presentación

Este libro es la culminación de un proyecto para el mejoramiento de la labor docente y estudiantil inscrito en la Vicerrectoría de Docencia de la Universidad de Costa Rica con el código PD-EF-1503-2021. Los autores exponen argumentos basados en evidencia científica con el propósito de estimular investigación de alta calidad en las Ciencias del Movimiento Humano. A la vez, agradecen el apoyo de la Escuela de Educación Física y Deportes y del Centro de Investigación en Ciencias del Movimiento Humano de la Universidad de Costa Rica, así como a colegas investigadores con quienes han podido compartir y publicar sus investigaciones.

Esta es una obra que requiere conocimientos básicos en diseños de investigación y análisis estadístico; por lo que se presenta como un material de consulta de nivel intermedio a avanzado, que será especialmente útil en programas de maestría y doctorado. La obra se presenta en dos capítulos; uno acerca de avances metodológicos, y otro, que presenta avances estadísticos. El primer capítulo introduce conceptos actualizados referentes a metodologías de investigación utilizadas en estudios publicados en revistas científicas de alta calidad, de manera que los lectores puedan apreciar metodologías de vanguardia que se utilizan para responder preguntas de investigación relevantes para las Ciencias del Movimiento Humano.

El segundo capítulo introduce técnicas novedosas de análisis de datos que se han incorporado al universo de los investigadores para poder afrontar la cada vez más compleja cantidad de información recopilada en los estudios. Además, la integración de ambos capítulos ofrece nuevas oportunidades para el trabajo inter, multi y transdisciplinario, aceptando con humildad que un proceso de investigación se ve claramente favorecido por las diferentes perspectivas de profesionales unidos para resolver los principales problemas que afronta la sociedad en la búsqueda del conocimiento.

Agradecimientos

Se agradece a la Escuela de Educación Física y Deportes de la Universidad de Costa Rica por haber brindado el tiempo para preparar esta obra, así como a nuestros estudiantes de posgrado y colegas, quienes leyeron versiones preliminares del libro y nos brindaron excelentes comentarios para mejorarlo.

Dedicatoria

"A la Dra. Marta E. Picado Ramírez, quien me inspiró a explorar el riguroso pero gratificante camino de la ciencia y la academia" – José Moncada Jiménez

"A mi esposa Irene y mis hijos Ignacio, Daniel e Isabel, inspiración para avanzar. – Pedro Carazo Vargas

"A quienes han sido mi inspiración y refugio en el arduo pero hermoso camino académico" – Isaura Castillo Hernández

"A mi esposa Éricka y a mi hija Samira, mis motores de aprendizaje" – Gerardo A. Araya Vargas

Contenido

Presentación.....	7
Agradecimientos.....	9
Dedicatoria	11
Introducción	17
Capítulo 1: Metodología avanzada	21
1.1. Revisión y síntesis de la evidencia científica	25
1.2. Revisiones de literatura.....	28
1.2.1. Bibliométricas	28
1.2.2. Metodológicas	34
1.2.3. Realistas	36
1.2.4. Clínicas	38
1.2.5. Narrativas.....	45
1.3. Revisiones sistemáticas de literatura	47
1.3.1. Vivas.....	47
1.3.2. De alcance.....	49
1.3.3. Sombrilla	54
1.3.4. De métodos mixtos	59
1.4. Revisiones meta analíticas	65
1.4.1. Meta análisis	65
1.4.2. Meta análisis armonizado	66
1.4.3. Meta análisis ²	67
1.4.4. Meta análisis en red.....	69
2.1. Evaluación de calidad y riesgo de sesgos	77
3.1. La variabilidad en la investigación.....	88
3.1.1. Estudios en gemelos	88
3.1.2. ¿Por qué se estudia la variabilidad?.....	93
3.1.3. Variabilidad entre diferentes individuos	98
3.1.4. Variabilidad dentro del mismo individuo	100
3.1.5. Respondedores vs. No respondedores	106
3.2. Diseños de investigación con alto potencial	113
3.2.1. Estudios de factibilidad, piloto y prueba de concepto	115
3.2.2. Dosis-Respuesta	120

3.2.3.	Entrenamiento-Desentrenamiento.....	125
3.2.4.	Ensayos controlados aleatorizados por conglomerados	130
3.2.5.	Longitudinal prospectivo, retrospectivo y análisis secundario de datos .	133
3.2.6.	Estudios observacionales	140
3.2.7.	Estudios de no inferioridad	146
3.2.8.	Sujeto único-estudio de caso y ABA.....	150
3.2.9.	Estudios econométricos	153
3.3.	Placebo y nocebo en la investigación.....	162
3.4.	Modelos animales	168
3.5.	Diseño y publicación de protocolos de investigación.....	173
Capítulo 2:	Estadística avanzada.....	179
2.1.	Los enfoques frecuentista y Bayesiano	187
2.1.1.	Alternativas y complementos de $p < 0.05$	191
2.1.1.1.	Riesgos relativos y razón de momios	194
2.1.1.2.	Precisión de las estimaciones.....	200
2.1.1.2.1.	Intervalos de confianza	200
2.1.1.2.2.	Intervalos de predicción.....	203
2.2.	Interpretación de tamaños de efecto	205
2.3.	Significancia estadística vs. clínica vs. práctica.....	208
2.4.	Validez y confiabilidad.....	211
2.4.1.	Evidencia de validez	212
2.4.2.	Evidencia de confiabilidad.....	215
2.5.	Tamaño de la muestra y remuestreo	223
2.6.	Datos perdidos en la investigación.....	227
2.7.	Transformación de datos	230
2.8.	Técnicas de análisis de datos poco utilizadas.....	231
2.8.1.	Análisis robustos de datos.....	232
2.8.1.1.	Inteligencia artificial	233
2.8.1.2.	Redes neuronales.....	235
2.8.1.3.	Aprendizaje de máquina	239
2.8.2.	Meta regresión.....	242
2.8.3.	Análisis de supervivencia.....	247
2.8.4.	Alometría	249

2.8.5.	Análisis composicional y de sustitución isotemporal	252
2.8.6.	Análisis de mediación.....	255
2.8.7.	Análisis de varianza multivariante permutacional.....	260
2.8.8.	Análisis de Rasch	261
2.8.9.	Análisis discriminante espectral	263
2.8.10.	Series de tiempo	264
2.8.11.	Análisis del perfil latente	266
2.9.	Administración de la ciencia	271
2.9.1.	¿Cómo comunicar los hallazgos de la investigación?	271
2.9.2.	Graficación de los hallazgos de la investigación	273
2.9.3.	Curva ROC	278
2.10.	Herramientas para la investigación.....	281
2.10.1.	Programas informáticos gratuitos.....	281
2.11.	<i>Errare humanum est</i>	290
Referencias	293	
Acerca de los autores	313	

Introducción

En términos generales, el propósito principal de la investigación científica es la producción de nuevo conocimiento para el beneficio de la sociedad. Las Ciencias del Movimiento Humano, en sus diversas áreas, también contribuyen con la generación de nuevos conocimientos, los cuales son diseminados por diferentes medios, ya sean impresos, orales o los más populares actualmente, los electrónicos. Sin duda alguna, hoy en día, tanto científicos como el público en general tienen acceso a millones de documentos, videos y otras fuentes de información que les permite adquirir conocimiento.

A pesar del vertiginoso avance de la ciencia impulsado por factores como un mayor desarrollo científico y tecnológico, la calidad de la información que se difunde puede y debe ser cuestionable. Es decir, el ojo entrenado del personal científico debería distinguir entre información de calidad e información cuestionable y carente de sustento científico. Eso es precisamente lo que diferencia al personal científico del público general. Sin embargo, aún dentro del gremio del personal científico en las Ciencias del Movimiento Humano, existen prácticas arraigadas que atentan contra la calidad de las investigaciones que se publican, por lo que las revistas prestigiosas proporcionan recomendaciones y consejos útiles para preparar y enviar manuscritos a revistas para la revisión de pares académicos^[637].

Los consejos abarcan desde la conceptualización hasta la aceptación del manuscrito. Estas sugerencias son importantes porque incluso en estudios publicados en los que existe una revisión de pares, se encuentran criterios diferentes, tanto metodológicos como de análisis de datos. Por ejemplo, Saco-Ledo et al.^[625] realizaron un meta análisis (i.e., una síntesis estadística de muchos estudios) acerca de los efectos del ejercicio sobre el tejido adiposo epicárdico, en el que encontraron que la intervención de ejercicio físico resultó en una reducción significativamente mayor en el tejido adiposo epicárdico entre personas con sobrepeso u obesidad en comparación con un grupo control que no realizaba ejercicio.

Sin embargo, Tian^[697] realizó una fuerte crítica a la metodología empleada por los investigadores. Entre sus argumentos, indicó que los autores solamente reportaron haber buscado artículos en dos bases de datos, y muestra evidencia que indica que eso no es suficiente porque no se alcanza una cobertura adecuada, lo cual obviamente tiene un impacto negativo en la cantidad de artículos meta analizables. También critica el método de efectos aleatorios utilizado para calcular los tamaños de efecto (i.e., la unidad de análisis en el meta análisis), ya que no es apropiado cuando se meta analizan pocos estudios (i.e., $n = 7$), por lo que muestran literatura que apoya el uso de otros métodos. Por último, Tian^[697] critica la técnica estadística utilizada para estudiar el sesgo de publicación, ya que también demuestra que pierde potencia estadística cuando se aplica a una pequeña cantidad de estudios. En síntesis, la crítica está bien fundamentada y además muestra las deficiencias en el proceso de revisión de pares de la revista, a pesar de haberse publicado en una revista prestigiosa.

Diversos investigadores recomiendan tomar en consideración elementos metodológicos para mejorar el diseño, ejecución y reporte de las investigaciones^{[224][232][584]}, incluso cuando se utiliza un paradigma de investigación cuantitativo o cualitativo^[761]. Uno de los consejos más relevantes se refiere a la planificación de cualquier estudio, en donde se incluye: a) generar una hipótesis comprobable, que tenga mérito y que sea sólida, b) evitar factores de confusión y otras formas de sesgo^[767], c) comprender la "unidad experimental", d) comprender los principios de la prueba estadística que se utiliza para analizar los datos, e) presentar resultados positivos y negativos

convincientes, f) es útil tener una idea de cómo serán los datos, g) si primero puede imaginar los resultados, luego puede planificar el análisis, h) mostrar los datos de la forma más clara y sencilla posible, y i) describir el experimento con suficiente detalle para que otros puedan repetirlo. Estos elementos son de especial interés tomando en consideración si el propósito del estudio es describir, predecir o generar inferencias de causalidad (i.e., "X" causa "Y", causa → efecto)^[534].

También existen excelentes recomendaciones para el análisis de datos en las Ciencias del Movimiento Humano^[348]. Algunas de éstas incluyen usar la precisión de la estimación para inferencias sobre los efectos de la población en referencia a las pruebas de hipótesis nulas, que son inadecuadas para evaluar la importancia clínica o práctica. Incluso, actualmente se discute el uso de pruebas de significancia no basadas en hipótesis (NHST, por sus siglas en inglés) y cada vez se utilizan más los enfoques de análisis Bayesianos^[593]. Otras recomendaciones se refieren a justificar el tamaño de la muestra a través de una precisión o confianza aceptable para las decisiones clínicas en lugar de una potencia adecuada para la significación estadística. También se recomienda mostrar la desviación estándar en lugar del error estándar de la media con el propósito de comunicar mejor la magnitud de las diferencias entre las medias y la falta de uniformidad del error.

Otra recomendación se refiere a evitar los análisis puramente no paramétricos que no pueden proporcionar inferencias sobre la magnitud y son innecesarios. Con respecto a los estudios de validez, se prefiere utilizar estadísticas de regresión en lugar de los límites de concordancia, que son poco prácticos y sesgados. Otra recomendación se refiere a utilizar más los métodos cualitativos para enriquecer los proyectos cuantitativos basados en muestras y buscar la aprobación ética para el acceso público a los datos anonimizados sin procesar de un estudio para satisfacer la necesidad de un mayor escrutinio de la investigación y realizar mejores meta análisis. Finalmente, se recomienda usar covariables en los modelos lineales para ajustar los factores de confusión, tener en cuenta las diferencias individuales e identificar los posibles mecanismos de un efecto; utilizar la transformación logarítmica para tratar la falta de uniformidad de los efectos y el error; identificar y eliminar valores atípicos; y lidiar con el sesgo que surge de problemas con el muestreo, la asignación, el cegamiento, el error de medición y los prejuicios de los investigadores.

Reconocer, estudiar y combatir el sesgo es determinante para la realización de investigaciones de calidad. El sesgo se puede definir como "cualquier proceso, en cualquier etapa de inferencia, que tiende a producir resultados o conclusiones que difieren sistemáticamente de la verdad, p. 1."^[767]. Algunos de los principales son: a) sesgo de patrocinio, b) sesgos de selección, ejecución y detección, y c) sesgos de publicación (e.g., informe selectivo y no reportar resultados). Estos sesgos tienen el potencial efecto dañino de la supresión de evidencia crítica, de mostrar publicaciones que informan lo que probablemente sean hallazgos falsos positivos, así como de la divulgación de hallazgos inexactos o irreproducibles.

Cada uno de esos sesgos puede ser combatido revelando la relación financiera de los investigadores, por medio de la revisión por pares, listas de verificación de la revista y pautas para el investigador. Estas medidas también se complementan con las herramientas de sesgo de informe para su uso en revisiones sistemáticas, como por ejemplo, a través de escalas y listas de verificación como "**CON**solidated **S**tandards **O**f **R**eporting **T**rials" (CONSORT), "**E**nancing the **QUAL**ity and **T**ransparency **O**f health **R**esearch" (EQUATOR), "**P**referred **R**eporting **I**tems for **S**ystematic **R**eviews and **M**eta-**A**nalyses" (PRISMA), "**T**emplate for **I**ntervention **D**escription and **R**eplication" (TIDieR), y "**S**tandard **P**rotocol **I**tems for **R**andom**I**zed **T**rials" (SPIRIT)^[498].

En el caso de que ya se haya completado un estudio y se desee publicar, Pyne^[584] indica que existen al menos tres áreas claves para garantizar un manuscrito de alta calidad: a) la formulación y el diseño del proyecto, b) aspectos metodológicos y analíticos, y c) el estilo, claridad y calidad del manuscrito. También indica que se deberían realizar análisis de las respuestas individuales de los participantes del estudio bajo la noción de los “respondedores” individuales a las intervenciones de ejercicio, lo cual sin duda mejoraría la calidad de los manuscritos enviados a revisión de pares a las revistas científicas. Otro aspecto relevante se refiere a los mecanismos de los efectos, así como a su identificación y descripción; es decir, la medida en que una variable mecanicista media un efecto a través de mediciones que vinculan el predictor con la variable dependiente que se suma al impacto de un estudio, por lo que se debe profundizar en la sección de “Discusión” de los resultados.

En esta obra se pretende mostrar evidencia de alta calidad acerca de avances metodológicos y estadísticos para avanzar las Ciencias del Movimiento Humano, en donde confluyen diversas áreas disciplinares, interdisciplinares y transdisciplinares. El impacto esperado de esta obra será un mejor diseño, planeamiento, ejecución y reporte de investigaciones en la salud y el rendimiento físico humano. Esta obra también contribuye con el conocimiento en idioma español, lo cual significará un impacto para una generosa parte de la población de habla hispana, lo cual eventualmente permitirá un mayor avance en la calidad y pertinencia de las investigaciones publicadas en revistas de prestigio. Esperamos que esta obra pueda ser diseminada en los cursos de grado y posgrado impartidos en las universidades, así como en los centros e institutos de investigación relacionados con la generación del conocimiento en las Ciencias del Movimiento Humano.

Capítulo 1: Metodología avanzada



Introducción

Todo proceso serio de búsqueda del conocimiento requiere de un detallado planeamiento para finalmente responder a una pregunta de investigación. El detalle lógico y sistemático de cada paso del proceso permite cumplir con transparencia los requisitos de calidad impuestos por la comunidad científica internacional. Como decimos popularmente, para “jugar en las grandes ligas”, se requiere poseer y dominar herramientas metodológicas que vayan más allá de reclutar participantes y medirles alguna variable de interés. En este capítulo, se presentan algunos de los avances metodológicos más recientes en las Ciencias del Movimiento Humano que se observan en las publicaciones de alta calidad.

Uno de los aspectos metodológicos más recientes se refiere a la inscripción o registro de los protocolos de investigación^[4]. Esto quiere decir que el protocolo (i.e., antecedentes, metodología) se envía a una revista o a algún sitio especializado para la revisión de pares académicos, quienes realizan una serie de observaciones para mejorar la metodología propuesta del estudio. Es decir, tal y como se haría con un manuscrito finalizado, el protocolo es revisado exhaustivamente por miembros de la comunidad internacional. Esta es una práctica que le permite a los investigadores proponentes poder contar con un mejor protocolo, y a la vez, una mayor probabilidad de publicar el manuscrito una vez finalizado el estudio.

Una forma para comprender el tipo de estudio puede ser a través del “Modelo de Investigación Aplicada para las Ciencias del Deporte” (ARMSS, por sus siglas en inglés)^[78]. El ARMSS describe tres etapas amplias de investigación: a) la descripción, b) la experimentación, y c) la implementación. A la vez, éstas se subdividen en 8 etapas que van desde estudios descriptivos hasta estudios de implementación, lo cual permite identificar con mayor precisión el tipo de estudio que se desea realizar. Las revistas científicas también cada vez recomiendan seguir guías de buenas prácticas específicas para el diseño de investigación que aumenten el rigor metodológico, como, por ejemplo, las guías STROBE para diseños observacionales o las guías CONSORT para estudios intervencionales, y PRISMA y PERSiST para revisiones sistemáticas y meta análisis en temas de ejercicio, rehabilitación, medicina deportiva y ciencias del deporte^{[4][34]}.

Las revisiones de literatura y sus diferentes indicaciones de uso (e.g., sistemática vs. de alcance) cada vez cobran mayor relevancia debido a la extensa cantidad de información que se encuentra disponible de forma gratuita en bases de datos y repositorios^{[518][519]}. La búsqueda, acceso y organización de la información utilizando gestores o administradores de referencias (e.g., EndNote®, Zotero®, Mendeley®) es una buena práctica^[479]; incluso, actualmente se está recurriendo cada vez más al uso de inteligencia artificial para que, por medio de algoritmos computacionales, se puedan realizar búsquedas de literatura científica automatizadas para posteriormente resumirlas en revisiones sistemáticas y meta analíticas^{[41][135][562][575][782]}. A pesar de que no se han reportado en publicaciones en las Ciencias del Movimiento Humano específicamente, es posible que muy pronto se puedan consultar artículos en los que se haya aplicado esta tecnología.

Un consejo para mejorar la calidad de los estudios se refiere al fundamento de la propuesta de investigación. Es decir, los estudios de alta calidad recibidos en las revistas proporcionan un flujo de información lógico y conciso relacionado con la base teórica que lo sustenta, mostrando una cadena causal y las implicaciones prácticas que surgen de los datos analizados^[4]. Por consiguiente, los manuscritos que carecen de una justificación basada en la teoría o en preguntas de investigación claras basadas en esa justificación teórica van a tener pocas posibilidades de ser publicados – al menos, en revistas de reconocido prestigio –.

Otros aspectos relacionados con la calidad metodológica y la transparencia se refieren a la búsqueda de colaboración entre grupos de investigación nacionales e internacionales, lo cual brinda una perspectiva más integral a los manuscritos enviados a las revistas. Por ejemplo, en temas relacionados con la fisiología, se ha estimado que de 1976 al 2016, la cantidad de autores en un artículo pasó de 2 a 5 autores^[280]. También se promueve el uso de paradigmas mixtos (i.e., cuantitativo, cualitativo), el uso de muestras de participantes heterogéneas (i.e., poblaciones poco representadas o estudiadas), el cálculo y reporte del tamaño de la muestra, la disponibilidad de los datos recolectados sin procesar para los lectores, el compartir códigos de análisis como los utilizados con los programas como *R* o IBM-SPSS®, mostrar estadísticas complementarias a $p < 0.05$ (e.g., intervalos de confianza, tamaños de efecto, intervalos de predicción), y mostrar los patrocinadores del estudio, así como reportar las limitaciones del estudio y sus fortalezas^{[4][280]}.

Todos estos elementos y otros más permiten reducir la denominada crisis de reproducibilidad, en la que se ha mencionado que una gran parte de los hallazgos publicados en las revistas científicas pueden ser falsos^[363]. Bajo esa premisa de falsedad, se recomienda reproducir los métodos para determinar si los hallazgos reportados son consistentes o no. Eso ayuda a validar los resultados reportados en las investigaciones, y por lo tanto, representan un avance metodológico importante.

1.1. Revisión y síntesis de la evidencia científica

La generación y difusión del conocimiento han aumentado exponencialmente en la última década debido, en parte, al avance de la ciencia *per se*, la tecnología y el internet, así como al uso de las redes sociales (e.g., Facebook®, Twitter®, Instagram®) y a un mayor aprovechamiento de los recursos para la generación de nuevo conocimiento (Figura 1). Se podría afirmar que una extensa cantidad de la producción científica del s. XXI está acompañada ineludiblemente de la comunicación de ésta; los medios de comunicación se han diversificado, así como las estrategias del personal científico para difundir sus hallazgos y alcanzar a una mayor cantidad de lectores. Se cuenta con una suficiente cantidad de información en temas de particular interés, pero con insuficiente tiempo para leer cada una de las publicaciones para formarse un criterio y plantear una nueva pregunta de investigación o para tomar decisiones acerca del tipo de prescripción del ejercicio, entre otras. Es más, la sobrecarga de literatura limita la productividad científica^[623]. Consecuentemente, es indispensable resumir, analizar y sintetizar la evidencia científica disponible para poder tomar decisiones acerca de si se continúa con nuevos estudios o si se redirigen los recursos y esfuerzos hacia una nueva línea de investigación diferente de ese tema de interés^[382].

Figura 1. El conocimiento académico generalmente se obtenía en lugares especiales, conocidos como bibliotecas, que reunían una amplia cantidad de libros, revistas y otros materiales de consulta. El proceso para acceder a la información era formal y administrativo. Con la aparición y el desarrollo de las computadoras de escritorio y portátiles, así como con el inicio del internet, la información científica se puso al alcance de más personas. Con el acelerado desarrollo de la telefonía celular y de nuevas redes de internet, ahora hasta en redes sociales se difunde conocimiento científico, lo que permite un mayor acceso a información reservada anteriormente sólo para unos pocos.



La comunidad mundial de investigadores constantemente revisa y resume la evidencia científica para presentar a sus colegas un nuevo concepto, reunir o combinar dos temas por primera vez, y para mejorar la comprensión acerca de cuáles investigaciones se han llevado a cabo en un área específica. También resumen las investigaciones para responder a las prioridades expresadas por quienes formulan políticas, para mejorar la exactitud o precisión de una estimación recopilando datos de estudios cuantitativos similares, o para combinar una gran cantidad de estudios que examinan los impactos de las intervenciones. Otra razón para resumir los hallazgos principales de un cuerpo de evidencia es para evaluar si una intervención específica es generalizable, es decir, si se puede esperar que tenga un efecto deseado similar en un contexto novedoso^[308].

Existen distintas formas aceptadas por la comunidad científica internacional que se utilizan para resumir objetivamente decenas de investigaciones. Esto facilita la traducción de los conocimientos a la práctica. Sus objetivos son distintos y el grado de profundidad o rigurosidad también. Por ejemplo, los "usuarios de evidencia" representan una categoría de personas cuyo interés primordial es resumir un amplio cuerpo de evidencias para responder a preguntas muy generales y cuantificar la incertidumbre^[308]. Los usuarios "metodológicos" por su parte, maximizan la rigurosidad, la transparencia y la profundidad de las revisiones de literatura, por lo que este tipo de revisiones minimizan la susceptibilidad a tener sesgo en sus conclusiones. Finalmente, la categoría de "autores de revisiones" representa a investigadores que desean publicar buenas revisiones de amplia relevancia para la sociedad y su objetivo primario es estimular la atención en nuevos temas, algunas veces de un interés y alcance muy reducido.

Actualmente, se reconocen diversos tipos de revisiones sistemáticas de literatura, entre las que se encuentran las de efectividad, experienciales, evaluación de costo-económicas, prevalencia e incidencia, precisión diagnóstica de una prueba, etiología o riesgo, opinión de experto, psicométrica, pronóstica y metodológica^{[294][519][630]}. Algunas de esas metodologías incluyen el meta análisis (i.e., armonizado, de previos meta análisis, en red o "network"), la bibliometría, las revisiones sistemáticas vivas ("living"), la revisión metodológica, la síntesis realista, la revisión clínica, las revisiones narrativas, las revisiones sistemáticas de alcance ("scoping review"), las revisiones sombrilla ("umbrella review"), el cuasi meta análisis y la revisión de métodos mixtos^{[95][502][603]}. También existen revisiones de evidencia en las que se utilizan metodologías cualitativas^[17].

Estas metodologías pueden organizarse en una jerarquía y permiten responder a preguntas prácticas, como, por ejemplo, ¿cuál es el mejor tipo de estrategia de entrenamiento para mejorar la fuerza, la potencia, la velocidad, la capacidad aeróbica, reducir el tiempo en un maratón, mejorar la imagen corporal, la autoconfianza, etc.?^[125]. Otras permiten conocer quiénes son los investigadores más prolíficos o las instituciones y países que lideran la investigación en un tema en particular. A pesar de que existen diversas metodologías para reducir y sintetizar la información proveniente de múltiples fuentes (e.g., artículos, libros, tesis, literatura gris), también se advierte acerca de la información publicada en lo que se conoce como revistas depredadoras o "*predatory journals*" en inglés^[609], pues los artículos podrían contener mayores sesgos que los publicados en otras revistas. Por ejemplo, en una revista depredadora puede existir una escasa o nula revisión de pares, aspecto esencial en la fiscalización de la ciencia y que es totalmente reconocido mundialmente como una práctica indispensable para la credibilidad de la comunicación de esta. También se sabe que las revistas depredadoras representan un importante negocio con ganancias millonarias al cobrar altas cantidades de dinero por la publicación de manuscritos. Un sitio web en el que se puede observar una lista no exhaustiva de casas editoriales y de revistas consideradas como potencialmente depredadoras es <https://beallist.net/>

Cabe mencionar que actualmente existe una forma inescrupulosa de apropiación de información, por medio de la cual algunas revistas obtienen artículos publicados en otras revistas y los reproducen en las suyas, haciéndolos pasar por manuscritos enviados por los autores y revisados por pares académicos, lo cual es falso. Dos de los autores de este libro pudieron observar la forma en que una revista se apropió indebidamente de un artículo que fue publicado en otra revista. El artículo original completó positivamente todos los procesos editoriales de la *Revista Educación* y fue publicado en el año 2021^[505], y fue tomado sin permiso, traducido del español al inglés y publicado en otra revista en el año 2022^[506]. Esto, además de ser ilegal, puede representar un problema de naturaleza ética para los autores del manuscrito porque se podría considerar, erróneamente, que enviaron el artículo simultáneamente a dos revistas, lo cual provocaría una doble publicación.

Dejando de lado las formas nefastas para la difusión de información, en esencia todas las estrategias poseen una finalidad única, sintetizar el conocimiento en un área específica del conocimiento, y actualmente se estimula a las personas investigadoras a inscribir sus protocolos de revisión en plataformas públicas y utilizar procedimientos de reconocida calidad que permitan ser replicados por otros grupos de investigación^{[389][497]}. La inscripción o registro permite economizar tiempo y recursos al saber que existen otros grupos intentando sintetizar un tema particular, así como también posicionar a su institución e investigadores en un tema particular. En esta sección, se presentarán algunas de las estrategias de síntesis y análisis del conocimiento que han recibido especial atención recientemente, las cuales, en su mayoría, ha sido desarrolladas debido a las herramientas tecnológicas como las computadoras y programas informáticos de análisis y graficación de datos. Se presenta una descripción metodológica del tema, así como evidencia recientemente publicada para explicar de una manera práctica y real la forma en que se utiliza dicha herramienta.

1.2. Revisiones de literatura

1.2.1. Bibliométricas

En las áreas del conocimiento de las ciencias de la información, bibliotecología, filología y ciencia de las ciencias, se nombran tres constructos que algunas veces se consideran sinónimos, pero que, sin embargo, poseen definiciones y orígenes propios. Así, los términos *cienciometría*, *infometría* y *bibliometría* se refieren al análisis de documentos para obtener métricas de la producción académica de autores o áreas de investigación (i.e., producción científica); sin embargo, la *bibliometría* pertenece a las ciencias de la bibliotecología y la documentación, la *cienciometría* a la ciencia de las ciencias y la *infometría* a las ciencias de la información^[766]. Estas disciplinas pueden utilizar herramientas de análisis matemáticas, algoritmos de detección de relaciones, minería de texto, redes de proximidad, agrupamiento jerárquico y análisis afines a las ciencias de la información, como, por ejemplo, análisis de texto de red y redes de referencia^[404].

La *bibliometría* es un tipo de "mapeo" de la evidencia científica, la cual incluye un procedimiento de análisis de datos que permite catalogar la evidencia científica (i.e., literatura publicada) para comprender sus tendencias (i.e., distribución cronológica), enfoques y contribuciones a la academia (i.e., impacto) desde diversas aristas; como, por ejemplo, un área del conocimiento específica (e.g., nutrición deportiva, gestión deportiva), un país, una institución, un autor, o una revista, lo que a su vez facilita la valoración para determinar qué tan popular o relevante ha sido un tema o si ha existido algún "pico" histórico de publicaciones^{[404][578]}. Se podría decir que esta técnica aplica métodos matemáticos, estadísticos y heurísticos (i.e., descubrimiento) para publicaciones científicas. Se le considera un insumo importante de información para la toma de decisiones, justificación y persuasión necesaria en la búsqueda de financiamiento para la investigación, ya que ofrece evidencia cuantitativa objetiva sobre la calidad y cantidad del trabajo científico en parámetros de actividad, impacto y vínculo.

La *bibliometría* permite conocer dónde podrían estar los vacíos o brechas de conocimiento en un tema de investigación (e.g., grupos de edad, demográficos, diseños de estudio, padecimientos), y a la vez facilita y orienta al lector para que conozca las revistas y áreas más afines a la publicación del tema de interés, laboratorios y palabras clave que podrían utilizarse para buscar evidencia publicada en revistas. Esto a su vez promueve una asignación óptima de recursos de investigación intelectuales y materiales, los cuales suelen ser escasos. A diferencia de los meta análisis, los cuales en general pretenden demostrar efectos de una intervención en términos de magnitud y dirección de esa magnitud, las revisiones bibliométricas no tienen como objetivo resumir el efecto ni la magnitud de alguna intervención. Sin embargo, ambos tipos de revisiones podrían considerarse, en efecto, estudios o técnicas complementarias; es decir, se podrían sugerir estudios bibliométricos a partir de revisiones sistemáticas de literatura con o sin meta análisis y viceversa.

Como se indicó, la *bibliometría* permite conocer el origen y ubicación de autores, instituciones y países más prolíficos en la generación del conocimiento en un tema específico de las Ciencias del Movimiento Humano. Este tipo de estudios permiten resaltar el impacto de la investigación previa y tendencias actuales en el área, mientras que a la vez ofrece dirección y recomendaciones para futuras investigaciones. Por ejemplo, Memon et al.^[488], ofrecen recomendaciones concretas en el tema de la actividad física y el sueño, en donde indican que los países en vías de desarrollo necesitan hacer más investigación en esta área, y para ello, ofrece una serie de medidas que podrían tomarse al respecto.

Este tipo de orientaciones basadas en la evidencia proporcionan seguridad a la comunidad científica y guían los esfuerzos para la búsqueda de nuevo conocimiento.

La bibliometría es considerada como uno “artículo de revisión”, y una de sus ventajas es que no requiere de aprobación de un comité de ética para su elaboración, pues las unidades de análisis son los artículos publicados y no las personas. Otra de las ventajas es que requiere de conexión a internet para acceder a las bases de datos y no es necesaria demasiada potencia en el hardware para procesar los datos, por lo que una computadora regular es suficiente. Así, mediante la aplicación de modelado teórico de red estadística de las referencias, similitud de texto y relaciones de autoría de la literatura, el mapeo de la ciencia a través de la bibliometría facilita la exploración de la estructura temática-conceptual, las tendencias y la dinámica del área de conocimiento.

Una forma para realizar una revisión bibliométrica consiste en revisar bases de datos (e.g., PubMed, **Web of Science** [WoS], SCOPUS, Redalyc) con el propósito de extraer la información en algún formato práctico, como, por ejemplo, en archivos de Excel o CSV (**C**omma-**S**eparated **V**alues), los cuales brindan mayor versatilidad para el manejo de grandes cantidades de información. Una vez obtenidas frecuencias y porcentajes, así como índices de calidad, se procede a resumir los datos y presentarlos en mapas de visualización, rankings y tablas de frecuencias. En la tabla 1, se presenta un ejemplo hipotético del perfil de los primeros 10 autores que publicaron investigaciones acerca del ejercicio y la imagen corporal en atletas.

El análisis bibliométrico puede utilizar únicamente estadísticas cuantitativas descriptivas, pero también se puede apoyar de técnicas avanzadas de minería de textos, análisis de redes de referencia semántica (i.e., el significando de las palabras, frases y textos), algoritmos de identificación y análisis de “cluster” o conglomerados, entre otras.

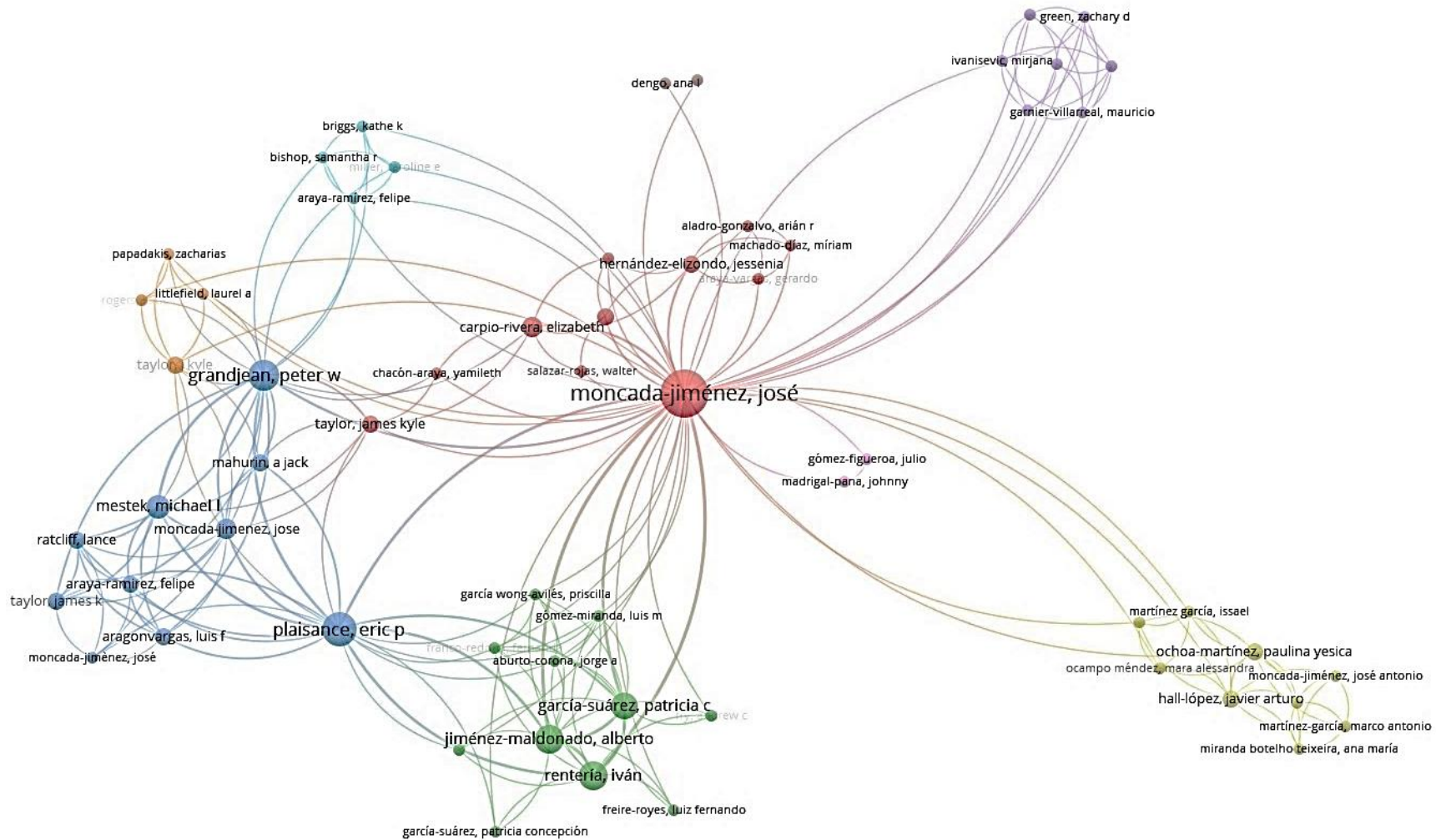
La esquematización gráfica y diversa utilizada en el análisis bibliométrico permite resumir visualmente un elevado volumen de información objetiva y cuantitativamente, lo cual facilita a los diferentes grupos de investigación a nivel mundial, tomar decisiones relacionadas con las implicaciones de estudios futuros en un tema específico y su evolución temporal. Así, un propósito importante de los estudios bibliométricos es comunicar resumida y fácilmente la información, y para ello pueden utilizar programas gráficos diversos; como, por ejemplo, algunos disponibles de acceso o código abierto gratuitos (e.g., *VOSviewer*[®]). En la figura 1, se muestra un ejemplo gráfico de la asociación de colaboración entre varios autores. Se puede observar que, en algunos de los nódulos o círculos de reducción de datos, se encuentran nombres repetidos. Eso se debe a que el nombre de los autores ha sido indizado de forma diferente en las bases de datos, por lo que se recomendaría realizar una limpieza de las bases de datos para uniformar la información antes de graficar. También se sugiere a los investigadores, ser consistentes en la forma en que reportan sus nombres y apellidos cuando envían manuscritos a publicar. Algo importante, es que en las revistas anglosajonas, en general, se utiliza solamente un apellido, pero para la mayoría de los hispanoamericanos, se utilizan dos apellidos; por tal motivo, se recomienda usar un guion entre los apellidos (e.g., Araya-Vargas) cuando se envían manuscritos a ese tipo de revistas

Tabla 1. Perfil de los primeros 10 autores en el área de ejercicio e imagen corporal en atletas. El ejemplo es hipotético, pero por lo general se presentan estadísticas asociadas a cada autor, así como su institución de origen al momento del análisis de las publicaciones. Esta información permite identificar líderes en el campo de estudio, lugares y laboratorios prolíficos en la producción científica.

Puesto	Autor	Número de artículos	(%)	Número de citas	Citas por artículo	Índice h	Afiliación institucional
1	Grandjean, P.	35	2.64	274	17.1	15	University of Mississippi, EE.UU.
2	Araya, G. A.	32	2.47	308	20.5	7	Universidad de Costa Rica, Costa Rica
3	Chacón, Y.	23	2.14	264	20.3	9	Universidad Pablo de Olavide, España
4	Carazo, P.	22	2.13	771	59.3	10	Universidad de Castilla-La Mancha, España
5	Castillo, I.	21	1.98	262	21.9	9	University of Georgia, EE.UU.
6	Taylor, K.	20	1.81	260	23.9	9	Auburn University, EE.UU.
7	Rentería, I.	19	1.65	195	36.5	11	Universidad Autónoma de Baja California, México
8	Jiménez, A.	18	1.48	164	24.4	11	Universidad Autónoma de Baja California, México
9	Hernández, J.	15	1.47	130	23.5	8	Universidad de Granada, España
10	García, P.	12	1.46	125	22.8	12	Universidad Autónoma de Baja California, México

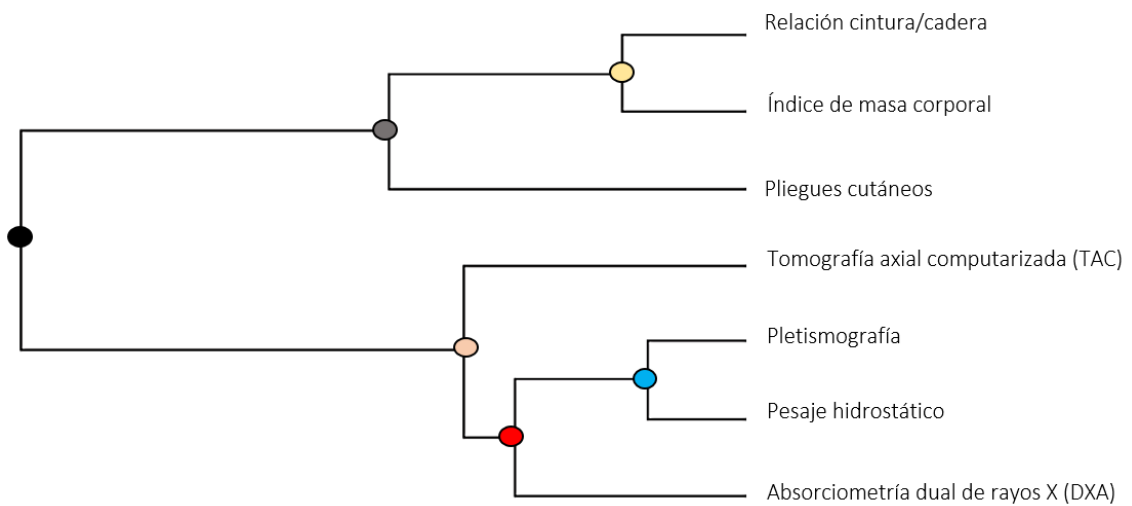
Este mapeo bibliográfico permite visualizar la investigación académica en forma de panoramas científicos. Los panoramas científicos, a diferencia de las revisiones convencionales, que se centran en preguntas de investigación, extraen información en varios niveles y capturan la relación entre diversos temas y conceptos, creando una especie de "síntesis de revisión".

Figura 1. Ejemplo hipotético de visualización de información de las relaciones de colaboración entre varios investigadores utilizando el programa gratuito VOSviewer®, versión 1.6.16. Las líneas conectan a los autores en sus trabajos de colaboración, y el tamaño del círculo, nodo o “cluster” representa la magnitud de las colaboraciones de esa persona en particular; entre mayor sea el tamaño del círculo, también será mayor el grado de colaboración o coautoría.



Otro estilo para presentar los resultados gráficamente es mediante dendrogramas y matrices. Los dendrogramas son representaciones gráficas de datos en forma de árbol que permiten agrupar la información en varias subcategorías, que, a su vez, se van subdividiendo hasta llegar a un nivel de detalle específico. Se les obtiene por medio de análisis estadísticos de conglomerados jerárquicos, y tienen como intención brindarle una importancia o jerarquía a un grupo de palabras, variables o datos y a la vez mostrar la distancia que existe entre éstos^[468]. Por ejemplo, si se realizara una revisión bibliométrica del tema de medición de la composición corporal humana, y se identificaran publicaciones al respecto, posiblemente se les podría graficar en un dendrograma (Figura 2).

Figura 2. Ejemplo hipotético de visualización de un dendrograma acerca del tema de medición de la composición corporal. Los círculos en colores representan nódulos en los que se agrupan los datos.



Las posibles asociaciones poseen una magnitud (i.e., conceptos densamente relacionados), la cual se puede observar en el grosor de las líneas que los unen o en el tamaño de los "clusters", dependiendo de la forma preferida de comunicación de los autores y del software utilizado. Así, la figura 2 permite comprender asociaciones entre los indicadores o variables que se miden comúnmente, y se lograría agruparlas y asignarles una taxonomía a esos temas; como, por ejemplo, en técnicas de campo, técnicas de laboratorio, métodos indirectos o doblemente indirectos. Los diversos indicadores o variables se agrupan desde sus raíces (derecha) hasta conformar el tronco del árbol (izquierda). En la intersección de cada indicador se forman nódulos (diferentes colores), que representan constructos a los cuales se les puede asignar un nombre o categoría descriptiva. Esos nombres pueden ser arbitrarios o basados en la teoría, dependiendo del área de estudio y de la producción del conocimiento existente. De esta forma se puede tener mucha mayor claridad acerca de las publicaciones, su cantidad, impacto, su agrupación y posibles relaciones. Esta información permite replantear estrategias de investigación y posiblemente maximizar los recursos disponibles.

Kiss et al.^[404], indican que ya se han publicado estudios bibliométricos en psicología del deporte, gestión deportiva, economía del deporte, actividad física y envejecimiento, y en nutrición deportiva, por lo que se espera que los estudios bibliométricos tomen mayor auge en los próximos años. Recientemente, se publicaron estudios bibliométricos que involucran el ejercicio físico y el dolor neuropático^[164], dolor lumbar^[736], calidad de vida en niños^[770] y fútbol^[403]. Así, se podrían realizar múltiples estudios bibliométricos, como, por ejemplo, en temas de actividad física y control del dolor, en entrenamiento físico y sueño, en deportes específicos y sus métodos de entrenamiento, entre otros. Incluso, se pueden realizar revisiones bibliométricas que podrían utilizarse para guiar políticas de salud pública que involucren la actividad física^[422].

Estos temas podrían desarrollarse como parte de los estudios de licenciatura o de maestría para orientar a los estudiantes acerca de un tema particular, y así lograr investigaciones más dirigidas basadas en la evidencia existente. De hecho, la bibliometría podría ser un paso previo a la realización de una revisión sistemática de literatura con o sin meta análisis, por lo que no puede ser considerada como algo de menor calidad metodológica.

1.2.2. Metodológicas

Para avanzar en el conocimiento de un tema específico es necesario reconocer que no todos los estudios son diseñados de la misma manera. Es decir, cada investigación publicada puede haber utilizado diferente metodología para responder a la misma pregunta, lo que obviamente provoca que las conclusiones deban ser interpretadas y comprendidas en un contexto particular, y muchas veces con cautela. Esto hace que la síntesis de los estudios sea complicada, más allá del propio proceso de búsqueda y selección de cada artículo. Por este motivo, algunas veces, el personal investigador enfoca sus esfuerzos en conocer cuáles son los factores metodológicos que podrían explicar los hallazgos generales obtenidos en las revisiones sistemáticas con o sin meta análisis. Esta disección metodológica realizada preferiblemente por expertos en la temática puede colaborar en la comprensión de variables que no se han considerado previamente o que pueden estar causando ruido o interferencia en los resultados de las revisiones sistemáticas y meta analíticas (Figura 1).

Figura 1. Las revisiones metodológicas se realizan para conocer detalles de la metodología interna de los estudios que pueda ser sujeta a mejoras.



Se debe recordar que la técnica meta analítica proporciona un tamaño de efecto global; es decir, muestra en unidades estandarizadas el efecto de una intervención sobre una variable de interés o variable respuesta. Sin embargo, no se enfoca en comparar diferentes metodologías que se usaron en los diferentes estudios más profundamente de lo que se reportaría con un análisis de variables moderadoras o una meta regresión. De hecho, una crítica común a los meta análisis es la escasa descripción metodológica, en donde se omiten o se describe someramente detalles básicos de cualquier prescripción del ejercicio, como por ejemplo la frecuencia, intensidad, duración de la sesión y de la intervención, modalidad, nivel de condición física inicial e incluso los valores de la variable principal de interés^[147].

Es por ello, que se ha recomendado realizar una revisión metodológica (en inglés, "methodological review"), que, en términos concretos, propone evaluar el campo de estudio al comparar exhaustiva y críticamente los aspectos del diseño, recolección y análisis de datos de una serie de estudios acerca de una temática particular. Así, el campo de estudio avanza, y para lograr ese objetivo, la persona o grupo de investigación debe poseer un profundo conocimiento teórico del tema en revisión para poder comprender relaciones teóricas que puedan ser afectadas por aspectos metodológicos; es decir, tiene que coexistir el conocimiento conceptual y procedimental.

La revisión metodológica puede catalogarse como breve o extensa. Por ejemplo, un grupo de investigadores de Canadá realizó una breve revisión acerca de la termorregulación durante el ejercicio en niños en edad prepuberal^[535]. Para lograr su objetivo, organizaron su revisión explicando aspectos conceptuales y teóricos acerca de la regulación térmica en los niños y los adultos, para posteriormente presentar aspectos metodológicos presentes en las investigaciones que deberían ser considerados cuando se diseñen nuevos estudios. En ese trabajo en particular, se describen y discuten aspectos metodológicos como el tamaño y el volumen corporal de niños y adultos, la selección de la intensidad del ejercicio, e incluso consideraciones especiales como el tipo de análisis estadístico necesario para comparar niños y adultos. Este tipo de información y recomendaciones son valiosas para avanzar el campo de estudio y permiten a los grupos de investigación poder plantear proyectos con mayor solidez conceptual y metodológica.

El grado de experiencia y conocimiento de los miembros del grupo de investigación permite detectar inconsistencias y limitaciones metodológicas que permiten evaluar críticamente la literatura, con el propósito ulterior de mejorarla. Así, la heterogeneidad entre los diferentes estudios se puede explicar – más allá de la variabilidad biológica que existe entre los seres humanos –, por aspectos tan básicos como la propia definición conceptual de la variable de interés. Por ejemplo, en el tema del esfuerzo percibido durante el ejercicio físico, se han utilizado diversas definiciones que complican sintetizar resultados al existir dudas acerca de la validez del constructo bajo escrutinio. Así, al menos una revisión metodológica ha señalado esa problemática^[315] y se han ofrecido dos soluciones puntuales que pueden mejorar el avance del conocimiento en esa área: a) reducir la cantidad de definiciones de esfuerzo percibido, escalas e instrucciones, y b) medir adicionalmente constructos de apoyo como afecto, fatiga e incomodidad, entre otras. Los investigadores indican que al cumplir esas dos simples recomendaciones se espera aumentar la validez de las mediciones y se podrá profundizar en el objeto de estudio^[315].

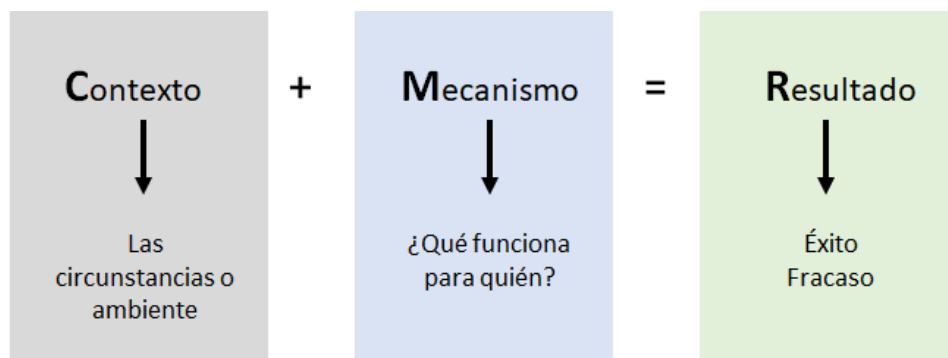
Otras críticas metodológicas que se han reportado en la literatura incluyen la utilización inconsistente de variables manipuladas en los estudios, como la cantidad de repeticiones, la intensidad de estas, la duración de las sesiones de ejercicio, o el descanso entre las sesiones de entrenamiento^[509]. Otros aspectos críticos para obtener un resultado consistente y replicable son el tipo de prueba física que se utilice (e.g., máxima, submáxima), el instrumento y el protocolo de medición (e.g., nuevo o estandarizado, con propiedades psicométricas de validez y confiabilidad), hora del día en la que se realizan las pruebas para evitar un posible efecto circadiano en la variable de interés, la alimentación previa, durante o posterior a la medición de alguna cualidad física o cognitiva, el tamaño de la muestra y la potencia estadística, la modalidad de entrenamiento, la familiarización con la prueba de rendimiento y su comodidad, el cegamiento al tratamiento o intervención, y el grado de involucramiento muscular de esta (e.g., cuerpo total, tren inferior o tren superior)^{[293][373][509]}.

1.2.3. Realistas

La síntesis realista, también llamada revisión realista (en inglés, “realist synthesis”), representa un enfoque cualitativo guiado por la teoría que se fundamenta en la filosofía de la ciencia realista y que considera la interacción entre el contexto, el mecanismo y el resultado^[760]. En términos simples, el tipo de pregunta que trata de responder la síntesis realista es “¿Qué funciona para quién, bajo qué circunstancias, cómo y por qué?” (p. 2)^[760]. Se considera que la revisión realista es una estrategia relativamente nueva para sintetizar la investigación y que tiene un enfoque más explicativo que crítico como el que se obtendría de un meta análisis o de un estudio aleatorizado controlado.

Las revisiones realistas son procesos menos homogéneos y mucho más repetitivos que proporcionan un análisis explicativo del mecanismo de trabajo que sustenta una intervención social compleja^[428]. Así, la síntesis realista trata de aclarar la relación “contexto-mecanismo-resultado” (CMR) (las siglas CMO en inglés para “Context”, “Mechanism”, y “Outcome”, Figura 1), lo que permite explicar casos de éxito, fracaso y diversas eventualidades intermedias. A este tipo de explicaciones teóricas se les llama “teorías de rango intermedio” y permiten diseñar posteriores pruebas empíricas. Aunque no existen estándares para llevar a cabo una síntesis realista por ser un método de revisión relativamente novedoso, se han realizado esfuerzos y actualmente se recomienda utilizar la guía de **Realist And MEta-narrative Evidence Syntheses: Evolving Standards (RAMESES)**^[760]. También se puede consultar los pasos planteados por otros autores^{[533][554]}.

Figura 1. Fundamento de la síntesis realista.



Este tipo de revisión de literatura aplica la filosofía realista a la síntesis de hallazgos de estudios primarios que tienen relación con una sola pregunta de investigación o con un conjunto de preguntas. Si se compara el enfoque meta analítico para la síntesis de evidencia con la síntesis realista se puede contrastar que, desde el punto de vista de la perspectiva realista, una determinada intervención (e.g., programa de ejercicio) no explica un tamaño de efecto en una variable dependiente; sino más bien, cómo la intervención cambia el contexto, lo que a su vez activa mecanismos que producen tanto resultados deseados como no deseados. Es decir, la intervención podría funcionar bien en un contexto, pero mal en otro contexto^[760].

Desde un punto de vista metodológico, las personas que realizan una síntesis realista comienzan extrayendo de la literatura las principales ideas que fundamentaron algún tipo de intervención (e.g., programa de ejercicio físico aeróbico). Tanto la pertinencia como la eficacia de cada una de esas ideas se somete a prueba utilizando fuentes de evidencia relevante (e.g., cualitativa, cuantitativa, administrativa) provenientes de las fuentes de literatura primaria sobre ese tipo de ejercicio. En la prueba, se reformula cada idea que calza dentro de la teoría del programa y se conceptualizan en términos realistas. Así, para cada idea, los revisores buscan las influencias contextuales (C) que se supone desencadenan los mecanismos relevantes (M) para generar los resultados (R) de interés^[760].

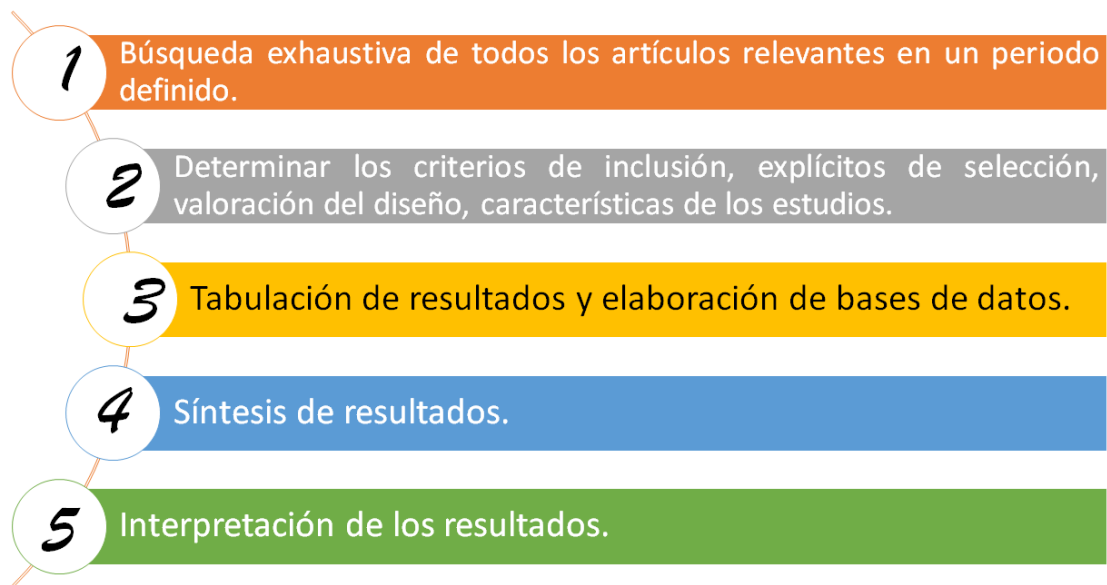
De esta forma, la síntesis de la evidencia consiste en comparar “¿Cómo se suponía que operaba el programa?” con la “evidencia empírica sobre la actualidad en diferentes situaciones”, a lo largo de las líneas C-M-R. La comparación analítica es el resultado de la capacidad del grupo investigador para describir y comprender las múltiples contingencias que afectan la probabilidad de que las intervenciones generen los resultados esperados. Esto, a su vez, proporciona orientación sobre lo que los responsables de la formulación de políticas o los profesionales podrían poner en práctica para cambiar el contexto o proporcionar recursos, de tal manera que probablemente activen los mecanismos correctos para producir el resultado deseado^[760]. En otras palabras, la configuración de CMR está diseñada para brindarles a los tomadores de decisiones una comprensión más profunda de la intervención y cómo hacer que funcione de manera más efectiva, por lo que brinda a los profesionales y a los formuladores de políticas una comprensión detallada y altamente funcional de las intervenciones sociales complejas^[428].

En conclusión, existe un creciente interés en la síntesis realista como método alternativo de revisión sistemática. Este enfoque ofrece la posibilidad de ampliar la base de conocimientos en áreas relevantes para la formulación de políticas; por ejemplo, explicar el éxito o el fracaso de determinadas intervenciones para la promoción de la actividad física^[760]. Por lo tanto, se espera que dentro de poco tiempo se publiquen más revisiones realistas en diferentes áreas de las Ciencias del Movimiento Humano.

1.2.4. Clínicas

Las revisiones sistemáticas consisten de estudios cuya población procede de artículos de tipo casos-controles previamente publicados; es decir, se trata de un estudio cuya materia prima son muchos otros estudios realizados con heterogeneidad de muestras y diseños de investigación. De esta forma, en una revisión sistemática se recopila la información generada por investigaciones de un tema determinado. Por otra parte, cuando se pretende valorar cuantitativamente los resultados de esos estudios y llegar a conclusiones sobre el porcentaje de efectividad de una intervención sobre alguna variable de interés, se utilizan técnicas más complejas, como por ejemplo los meta análisis. Algunos de los pasos previos para conseguir extraer la información de tantos estudios, se podrían resumir en la figura 1.

Figura 1. Pasos recomendados para la extracción de información en revisiones clínicas.



Estos diseños son también llamados diseños eficientes, ya que permiten generalizar resultados, incrementando el poder estadístico y la precisión de una estimación, haciendo además una evaluación estricta de la información publicada disponible^[464].

Las revisiones de literatura pueden ser de naturaleza práctica o básica y pueden llevarse a cabo en diferentes contextos, según sea el objetivo. Uno de esos contextos y tal vez de mayor relevancia es el ámbito clínico. Un "Clinical Review" o Revisión Clínica (RCLi) analiza desde el punto de vista narrativo diferentes documentos de naturaleza científica y médica. Aun cuando es sabido que cualquier revisión narrativa puede conllevar consigo previsibles sesgos en el proceso de recopilación, valoración crítica y síntesis, probablemente sean los documentos médicos con mayor difusión y utilización entre los profesionales sanitarios internacionalmente^[447].

De acuerdo con Siwek y Barton^[653], su alta utilización entre el personal de salud se debe a diferentes factores, entre los que se pueden enumerar la relativa facilidad en su elaboración y uso (en parte debido a la disponibilidad de publicaciones integradas o secundarias, como revisiones sistemáticas o guías de práctica clínica), la participación activa de los clínicos que las utilizan posteriormente, y la independencia de expertos en metodología (i.e., no es necesario tener conocimientos sobre diseños de investigación avanzados).

Usualmente, los tópicos más apropiados para llevar a cabo una RCl, son aquellos que avanzan científicamente con mucha rapidez y generan debate en torno a ellos, así como aquellos temas que crean dudas durante la consulta, tienen mayor incidencia o prevalencia y en los que su abordaje no es claro ni estandarizado para la mayoría del personal de salud. Cuando estos temas son amplios en su abordaje, como por ejemplo una enfermedad, se recomienda que se estructure en apartados, para guiar al lector en la visión general del tema. En términos generales, de acuerdo con Louro-González et al.^[447], al menos deberían ser obligatorios los apartados de:

- a. Definición, causas y fisiopatología. Esta última es opcional y se describe brevemente.
- b. Incidencia y características de la enfermedad o lesión.
- c. Diagnóstico y diagnóstico diferencial.
- d. Tratamiento y seguimiento.

Algunas RCl actuales en el área de las Ciencias del Movimiento Humano se describen en la tabla 1. Es visible la interdisciplinariedad con que se abordan dichas revisiones de literatura, lo cual permite la interacción de diferentes especialistas de la salud humana.

Tabla 1. Ejemplos revisiones de literatura clínicas en las Ciencias del Movimiento Humano.

Autores	Propósito o contexto de la revisión clínica	Conclusiones principales
Bytomski ^[133]	<ul style="list-style-type: none"> Una nutrición adecuada es crucial para que un atleta optimice su rendimiento para el entrenamiento y la competencia; sin embargo, no están claras las pautas a seguir por estos deportistas. 	<ul style="list-style-type: none"> Cada atleta tendrá diferentes objetivos de macronutrientes según el deporte, el momento del ejercicio y el estado de la temporada. No existen pautas específicas de micronutrientes para esta población, pero se deberían considerar pruebas de diagnóstico diferenciado para los atletas con deficiencias, que siguen regímenes de alimentación no tradicionales (e.g., vegetarianos) o que presentan lesiones.
Butts et al. ^[132]	<ul style="list-style-type: none"> El uso de creatina como suplemento dietético se ha vuelto cada vez más popular en las últimas décadas. A pesar de la popularidad de la creatina, quedan dudas con respecto a la dosificación, los efectos sobre el rendimiento deportivo y la seguridad. 	<ul style="list-style-type: none"> La creatina parece ser segura y efectiva para entornos particulares; sin embargo, saber si la suplementación con creatina conduce a mejoras en el rendimiento en competencia sigue siendo desconocido.
Chia et al. ^[166]	<ul style="list-style-type: none"> La hipertensión pulmonar (HP) es una condición clínica caracterizada por un aumento en la presión de la arteria pulmonar que da como resultado un aumento de poscarga ventricular derecha y disnea. Esto va acompañado de una reducción en la capacidad de ejercicio, calidad de vida y, finalmente, la muerte. El entrenamiento con ejercicio supervisado es recomendado como parte de un plan de tratamiento multifacético para la HP. Sin embargo, quedan muchas preguntas sobre cómo el entrenamiento físico mejora la capacidad de ejercitarse y la calidad de vida, así como la adecuada prescripción del ejercicio (i.e., frecuencia, tiempo, duración del programa e intensidad) en estos pacientes. 	<ul style="list-style-type: none"> Esta revisión proporciona una actualización sobre la fisiopatología del ejercicio y sus limitantes en la HP, sugiere mecanismos por medio de los cuales el ejercicio puede mejorar los síntomas y la función pulmonar, además ofrece recomendaciones basadas en evidencia con respecto a la frecuencia e intensidad de un ejercicio programa para pacientes con HP.
Continúa...		

Tabla 1. Continuación

Autores	Propósito o contexto de la revisión clínica	Conclusiones principales
Clauw ^[172]	<ul style="list-style-type: none"> • La fibromialgia está presente en un 2% a un 8% de la población, y se caracteriza por un dolor generalizado y, a menudo se acompaña de fatiga, problemas de memoria y alteraciones del sueño. • Esta revisión clínica revisó la epidemiología, fisiopatología, diagnóstico y tratamiento de la fibromialgia (NOTA: en esta revisión en particular, se discuten muchas terapias utilizadas en el tratamiento de la fibromialgia, pero las conclusiones principales, para efectos de este resumen se centrarán en el efecto de la actividad física y el ejercicio). 	<ul style="list-style-type: none"> • En general, y no solo para la fibromialgia sino para casi cada condición de dolor crónico, la actividad y el ejercicio físico son beneficiosos y no perjudiciales. Casi cualquier tipo de ejercicio es bueno para reducir o atenuar las alteraciones producidas por la fibromialgia o cualquier forma de dolor crónico. • Se debe recomendar a los pacientes que comiencen con ejercicio de intensidad leve o moderada y que aumentan su nivel de actividad lenta y progresivamente. Muchos pacientes tienden a aumentar el volumen y la carga de entrenamiento demasiado pronto, lo que provoca un empeoramiento del dolor. • Ser persistente en animar a los pacientes acerca de realizar ejercicio físico diariamente u otras terapias alternativas. • Evitar hacer el manejo terapéutico de estos pacientes pensando que cambiar a un fármaco diferente es el único enfoque disponible. • Si se promueve la utilización de terapias no farmacológicas de manera más agresiva y una menor prescripción de opioides, medicamentos antiinflamatorios no esteroideos y analgésicos de acción central, la fibromialgia es más fácil de manejar.
		Continúa...

Tabla 1. Continuación

Autores	Propósito o contexto de la revisión clínica	Conclusiones principales
Eckstein et al. ^[240]	<ul style="list-style-type: none"> • Detallar la evidencia para examinar la interacción entre la actividad física y la prescripción habitual de terapias hipoglucemiantes no insulínicas de personas con diabetes tipo 2 físicamente activas y/o aquellas que están usando farmacoterapia, pero que desean adoptar un estilo de vida físicamente activo. 	<ul style="list-style-type: none"> • Los resultados de la presente revisión abogan (con algunas salvedades) por el ejercicio físico para las personas en riesgo de desarrollar o que ya poseen diabetes tipo 2 mientras se está medicado (monoterapia). Se debe promover la actividad física y estar atentos en el seguimiento de los pacientes y hacer los ajustes prudentes tanto de medicamentos como de actividades físicas si se producen eventos adversos.
Matsel et al. ^[470]	<ul style="list-style-type: none"> • Resumir la evidencia actual acerca de la importancia de los programas de ejercicios para el cuidado del brazo en la reducción de la incidencia de lesiones en los jugadores de béisbol adolescentes. 	<ul style="list-style-type: none"> • Para los jugadores de béisbol adolescentes, realizar ejercicios que se centran en importantes deficiencias o problemas musculoesqueléticos, son eficaces para reducir las tasas de incidencia de lesiones.
Polli et al. ^[572]	<ul style="list-style-type: none"> • Se discute cómo la epigenética puede ser relevante para ampliar el campo de la clínica y las ciencias de la rehabilitación. Más concretamente, se centra en la epigenética del dolor, los factores psicológicos y la actividad física. • El campo de la rehabilitación puede potencialmente generar considerables beneficios de la combinación de evaluaciones clínicas con mediciones de marcadores epigenéticos. • La epigenética se puede utilizar para evaluar y apuntar a los mecanismos subyacentes que permiten adaptar, tanto como sea posible, los tratamientos a cada paciente y medir los efectos biológicos de esos tratamientos, y podría jugar un papel crucial en la predicción de las respuestas al tratamiento. 	<ul style="list-style-type: none"> • Los procesos epigenéticos, aunque tienen fines de diagnóstico, seguimiento y tratamiento siempre serán medidas secundarias. • Las mejoras relacionadas con la satisfacción del paciente siempre serán las principales y más importantes y no debe ser reemplazado por ninguna medida biológica. • Una investigación inicial prometedora ha demostrado que las primeras mediciones de biomarcadores pueden dar información relevante a la práctica clínica para predecir; por ejemplo, la susceptibilidad al desarrollo de complicaciones secundarias, como la epilepsia después de una lesión traumática cerebral. • Debido al potencial que la epigenética podría tener en el futuro cercano, esta revisión insta a los investigadores y médicos que trabajan en el campo de la rehabilitación para estar atentos al desarrollo de este emocionante campo en rápido crecimiento y considere combinar medidas clínicas con otras más biológicas.
Continúa...		

Tabla 1. Continuación

Autores	Propósito o contexto de la revisión clínica	Conclusiones principales
Tsiros et al. ^[704]	<ul style="list-style-type: none"> • Utilizando un estudio de caso y un formato clínico (SOAP: Subjetivo, Objetivo, Evaluación, Plan), exploran estrategias para mejorar el movimiento y aumentar la actividad física. • Esta revisión clínica aplicada pretende considerar las barreras al movimiento en el contexto de un estudio de caso y proporcionar sugerencias prácticas basadas en evidencias para facilitar la participación en diferentes actividades físicas y posteriormente realizar ejercicio más estructurado en niños que padecen de obesidad. 	<ul style="list-style-type: none"> • Esta revisión clínica aplicada destaca las deficiencias subyacentes en la fuerza, el estado físico funcional y el dolor comúnmente asociados con la obesidad infantil, junto con una disminución de la motricidad gruesa y el equilibrio. • Los beneficios a largo plazo se pueden lograr trabajando con los niños y las familias, diseñando cuidadosamente y progresando gradualmente en las intervenciones de ejercicio/actividad física a medida que los niños adquieren nueva capacidad, confianza y asegurando que los niños experimenten éxito en el movimiento. • Es imperativa una evaluación clínica exhaustiva y el manejo por personal calificado (i.e., expertos en desarrollo infantil, prescripción de ejercicio, y manejo musculoesquelético pediátrico) e incorporar enfoques centrados en la familia basados en el juego.
Hinman et al. ^[341]	<ul style="list-style-type: none"> • Los profesionales de la salud que atienden a mujeres embarazadas deben discutir sobre los posibles beneficios o contraindicaciones que el ejercicio puede tener en esta población. • Aunque la mayoría de las mujeres embarazadas no cumplen con las recomendaciones mínimas de ejercicio, hay un número creciente de mujeres físicamente activas que desean continuar entrenando durante todo el embarazo. 	<ul style="list-style-type: none"> • De acuerdo con esta revisión clínica, para mujeres sin contraindicaciones para la actividad física, el ejercicio es seguro tanto para la mujer como para el feto en desarrollo. • Aunque no hay evidencia concluyente de que el ejercicio previene eficazmente la diabetes mellitus gestacional, la preeclampsia, o depresión perinatal, sí parece tener efectos beneficiosos sobre la reducción de los niveles de glucosa, el riesgo de cesárea o partos vaginales instrumentales y aumento de peso materno. • En general, mujeres que son físicamente activas antes del embarazo se les debe aconsejar que mantengan o incrementen su nivel de actividad física con aval médico. Se debe alentar a las mujeres inactivas a que comiencen a hacer ejercicio supervisado por un profesional.
		Continúa...

Tabla 1. Continuación

Autores	Propósito o contexto de la revisión clínica	Conclusiones principales
Lepley et al. ^[424]	<ul style="list-style-type: none">Las alteraciones neuromusculares son un factor causal importante de lesiones primarias y secundarias en el contexto deportivo. Aunque algunos programas han tenido cierto éxito, las tasas de lesiones no han disminuido y, después de las lesiones, las personas a menudo regresan a la actividad deportiva con funcionalidad inferior a las recomendaciones clínicas. Considerar terapias alternativas a las concéntricas convencionales, tales como el ejercicio excéntrico, podría proporcionar más protocolos eficaces de prevención y rehabilitación de lesiones.	<ul style="list-style-type: none">El ejercicio excéntrico modifica beneficiosamente varios factores subyacentes de la morfología muscular (e.g., tipo de fibras, área de sección transversal, rango de trabajo y ángulo de penetración).La evidencia emergente indica que el ejercicio excéntrico también es beneficioso para la actividad neural periférica y central (i.e., reclutamiento/activación de neuronas motoras alfa, actividad del sarcolema, excitabilidad corticoespinal y activación cerebral).

1.2.5. Narrativas

Con el objetivo de difundir el conocimiento, anualmente se incrementa la cantidad de estudios publicados en las mejores bases de datos del mundo, en donde las revisiones narrativas contribuyen a resumir la información generada para facilitar su comprensión. De esta manera, las revisiones narrativas pretenden, mediante un paradigma cualitativo, identificar la literatura relevante en un área para discutirla e identificar fortalezas y limitaciones existentes^[260]. Las revisiones narrativas, al describir una serie de estudios sobre un problema de interés, pueden ser un primer acercamiento para un investigador que desea entender o dar a conocer la evidencia publicada sobre un tópico; no obstante, no necesariamente permitirá arribar a un entendimiento comprensivo y profundo del estado de la ciencia sobre un problema^[203].

Es así, como las revisiones narrativas permitirán alcanzar tres principales objetivos. Inicialmente, facilitarán ilustrar el conocimiento estudiando en un tema y consecuentemente, estas conclusiones podrán conducir a generar nuevas propuestas de investigación o a identificar nuevas estrategias de aplicación de los hallazgos generados por la ciencia. Luego de una breve introducción al tema, la elaboración de una revisión narrativa comienza con la selección de una serie de bases de datos vinculadas con el tópico de interés a investigar. Aunque algunas revisiones narrativas pueden describir sistemáticamente la estrategia de búsqueda para localizar los estudios^{[260][384]}, en su mayoría, este tipo de estudio suele omitir este tipo de información, concentrando su interés en recopilar la información sobre un tópico de interés, sin necesariamente seguir protocolos estandarizados ni estipularlos en la metodología empleada^{[203][371][566]}.

De esta manera, luego de recopilar la información que se integrará en el manuscrito, los autores establecerán categorías de análisis para facilitar su comprensión y posteriormente efectuarán un análisis del contenido de los artículos seleccionados. Por ejemplo, Kirk et al.^[402], con el objetivo de analizar la evidencia científica que estudiaba los requerimientos físicos de luchadores de artes marciales mixtas (MMA), realizaron una búsqueda en seis bases de datos y repositorios e incluyeron un total de 70 artículos. Posteriormente, analizaron la información agrupando los estudios en tres subapartados; los estudios que habían estudiado las características de los atletas de MMA; los efectos del entrenamiento de MMA en variables físicas o fisiológicas; o las características de rendimiento de las MMA, tales como cargas internas o externas, aspectos técnico-tácticos y predictores de desempeño. Si bien las revisiones narrativas pueden permitir comprender mejor el conocimiento desarrollado en un tema específico, entre las principales limitaciones presentes destaca la posibilidad de generar conclusiones sesgadas. Este inconveniente se puede generar por conflictos en la selección de las bases de datos y las palabras claves, la inclusión selectiva de los estudios hallados en la búsqueda de literatura y principalmente por la falta de criterios estandarizados para analizar los hallazgos y establecer las conclusiones^[203].

Debido a su valor, recientemente se han desarrollado revisiones narrativas en una diversidad de entornos vinculados a las Ciencias del Movimiento Humano, tales como la salud^{[624][626]}, el rendimiento deportivo^{[99][145][297][566][568]} y la educación física^[442] (Tabla 1).

Tabla 1. Ejemplos del propósito de algunas revisiones de literatura narrativa en las áreas de la salud, el rendimiento deportivo y la educación física.

Autores	Propósito de la revisión narrativa
Salud	
Ryom et al. ^[624]	Identificar los desafíos y motivos relacionados con la inactividad física, basándose en revisiones y complementado con entrevistas a expertos nacionales daneses en inactividad física, proporcionando una discusión general sobre el tema, generando en última instancia conocimiento utilizable en un contexto escandinavo.
Saeidi et al. ^[626]	Resumir el conocimiento actual sobre los efectos de la actividad física en los niveles de adipocinas en personas con sobrepeso y obesidad.
Rendimiento deportivo	
Botonis et al. ^[99]	Resumir la evidencia disponible sobre la influencia de la siesta en el ejercicio y el rendimiento cognitivo, así como los efectos de la siesta en las respuestas perceptivas de los atletas antes o durante el ejercicio.
Carmichael et al. ^[145]	Explorar el impacto de la fase del ciclo menstrual en el rendimiento percibido y medido objetivamente en poblaciones atléticas.
Grgic ^[297]	Resumir y evaluar críticamente la investigación sobre los efectos de la música en el rendimiento del ejercicio de resistencia.
Leow et al. ^[423]	Discutir las diferencias culturales existentes en las prácticas de hidratación en el contexto del ejercicio entre individuos físicamente activos, específicamente sus impactos potenciales en la rehidratación, el rendimiento deportivo y la recuperación.
Piedra et al. ^[566]	Presentar el estado actual del conocimiento sobre el monitoreo de cargas externas e internas en baloncesto.
Pierce et al. ^[568]	Describir la evidencia sobre el levantamiento de pesas para niños y adolescentes.
Educación Física	
Lopes et al. ^[442]	Resumir el conocimiento actual, identificando brechas clave en la investigación y presentando preguntas para futuras investigaciones acerca de la competencia motriz en niños y adolescentes.

Una revisión de literatura narrativa es una metodología útil que no requiere de conocimientos muy especializados; se puede desarrollar en cualquier entorno en el que previamente existan estudios científicos publicados y se presenta como una opción viable para la persona investigadora que requiere comprender mejor un tema y también busca adquirir experiencia. En las Ciencias del Movimiento Humano, se considera como una destreza básica poder conceptualizar y desarrollar una revisión de literatura narrativa, y paulatinamente ir evolucionando la síntesis del conocimiento hasta completar revisiones sistemáticas de literatura.

1.3. Revisiones sistemáticas de literatura

1.3.1. Vivas

Preparar, diseñar, ejecutar, completar, y, especialmente, publicar un estudio, se ha estimado que puede tardar entre 2.5 a 6.5 años^[755]. Lo mismo puede ocurrir cuando se trata de sintetizar varios estudios publicados a través de una revisión de literatura. Así, la precisión y utilidad de las conclusiones de las revisiones de literatura ya sean de naturaleza práctica, básica o incluso clínica en las Ciencias del Movimiento Humano, pueden ser erróneas o sesgadas en el momento en que el trabajo es finalmente publicado. En algunos casos, las consecuencias pueden ser triviales y no representan un riesgo; como, por ejemplo, en algún tipo novedoso de entrenamiento utilizado para mejorar el desempeño deportivo. Sin embargo, en otras ocasiones las implicaciones pueden ser graves, pues podrían significar la descalificación de un atleta o retrasar su recuperación de alguna lesión.

Por consiguiente, lo ideal sería poder contar con información constantemente actualizada (“en vivo”). Es así como surgen las revisiones sistemáticas vivas (en inglés, “Living Systematic Review, LSR”)^[248]. Esta metodología se presenta como una opción con base en el argumento de que las revisiones de literatura se desactualizan rápidamente; es decir, las conclusiones de una revisión de literatura pueden cambiar en períodos de tiempo que en algunas áreas del conocimiento oscilan entre los 2 y los 5 años^[648]. Al momento de publicación de esta obra no se habían encontrado estudios que indicaran en cuánto tiempo se desactualizan las revisiones sistemáticas de literatura en las Ciencias del Movimiento Humano.

Las LSR representan un enfoque novedoso propuesto por Elliott et al.^[249], y que consiste básicamente en realizar revisiones de literatura continuas de alta calidad, en línea^[755]. Como enfoque novedoso, éste también está en continuo desarrollo e incluso se ha creado una red de LSR (www.cochrane.org/lsr)^[248]. Las revisiones son constantemente actualizadas, incorporando evidencia relevante y novedosa a medida que ésta se encuentra disponible en revistas, repositorios de datos o en otros lugares^[248]. Este tipo de revisión y síntesis de la literatura es importante en diversas áreas de las Ciencias del Movimiento Humano en las cuales se produce conocimiento muy rápidamente, y en las que esos nuevos conocimientos pueden cambiar las políticas o la toma de decisiones. Por ejemplo, las LSR se podrían utilizar para cambiar las recomendaciones de prescripción de ejercicio para la población. Este cambio podría tener un impacto importante en la salud de la población y ayudaría a prevenir la aparición de múltiples enfermedades asociadas al sedentarismo. Por lo tanto, la comunidad no tendría que esperar 5 o 10 años para recibir recomendaciones actualizadas.

En las LSR se actualizan periódicamente aspectos metodológicos claves como la estrategia de búsqueda de las investigaciones, la evaluación del riesgo de sesgo, y, por supuesto, los datos que sustentan las conclusiones. Esta estrategia de síntesis permite que el artículo de revisión posea una mayor vida útil en comparación con una revisión de literatura más tradicional^[755]. Las LSR se pueden aplicar a cualquier tipo de revisión ya que utiliza métodos estandarizados de revisión e involucran un compromiso explícito y *a priori* con una frecuencia predeterminada de búsqueda y actualización del tema de la revisión.

Se espera que se sigan publicando LSR en temas propios de las Ciencias del Movimiento Humano en los próximos años, ya que las primeras apenas se comenzaron a publicar a partir del 2020^{[722][756]}, por lo que se les puede considerar pioneras. A la vez, se espera que vayan surgiendo avances metodológicos que permitan producir revisiones estandarizadas de mayor calidad. Tres ejemplos de LSR son los artículos publicados por un grupo de investigadores de Dinamarca^[756], Sudáfrica^[722] y Alemania^[634]. Todos los protocolos siguieron la metodología estandarizada del Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) y fueron inscritos en el International Prospective Register of Systematic Reviews (PROSPERO). Parte del protocolo de Winters et al.^[756] incluyó la LSR en el tema de estudio de dolor patelofemoral, el cual se iba a estar actualizando anualmente por un período mínimo de 5 años. De forma similar, Viljoen et al.^[722] indicaron que realizarían actualizaciones cada 6 meses durante al menos 5 años en el tema de factores de riesgo para corredores de campo traviesa. Por su parte, en el protocolo de Saueressig et al.^[634] se indicó que se realizaría una actualización anual por un período mínimo de 6 años en el tema de la eficacia del tratamiento quirúrgico en comparación con el tratamiento de rehabilitación para la ruptura del ligamento cruzado anterior.

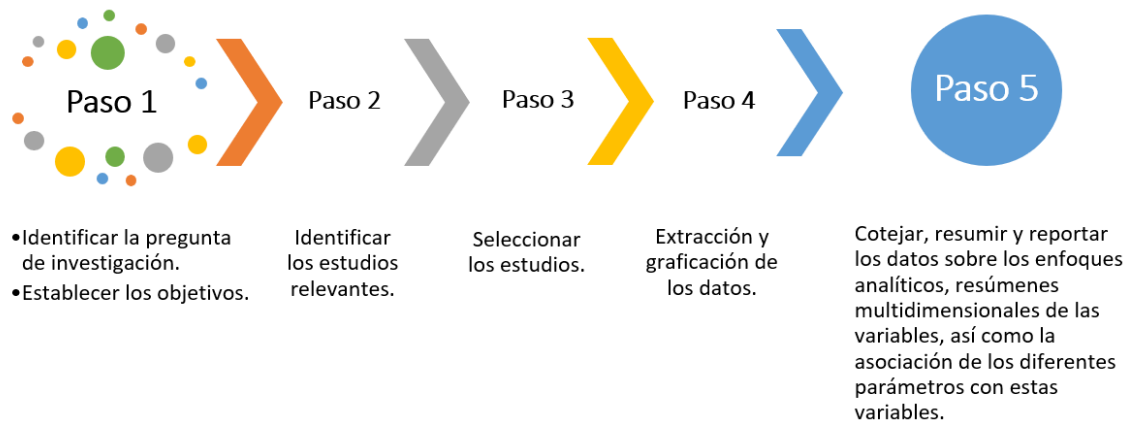
Este tipo de revisiones supone un compromiso a largo plazo, lo cual podría considerarse como una amenaza o inconveniente, ya que se necesita tiempo y recursos humanos y tecnológicos para mantener “viva” la revisión, así como el interés y compromiso de los investigadores. Sin embargo, si se cuenta con el personal y los recursos para darle continuidad a estas revisiones, se lograría tener evidencia temporal acerca de intervenciones de ejercicio sobre variables de interés en las Ciencias del Movimiento Humano.

1.3.2. De alcance

La revisión de alcance (en inglés, "scoping review") puede considerarse como una herramienta de "reconocimiento" que permite evaluar literatura científica que no ha sido revisada exhaustivamente^{[67][148]}. También se utiliza cuando el cuerpo de conocimiento específico es de naturaleza heterogénea; es decir, con hallazgos positivos, negativos o inconclusos respecto a una intervención y que por lo tanto no es posible resumirlo por medio de un enfoque más sistemático. Las revisiones de alcance brindan la flexibilidad de crear amplias narrativas dentro de una base de literatura limitada, lo que permite a los investigadores examinar la evidencia emergente donde las líneas precisas de cuestionamiento no están claramente determinadas. Además, estas revisiones ayudan a desarrollar definiciones clave y límites conceptuales, así como a identificar brechas en el conocimiento que permitirán plantear futuras investigaciones.

Una revisión de alcance no solo es una herramienta relevante para sintetizar el panorama interdisciplinario de la evidencia disponible, sino que también ofrece un enfoque sólido y transparente, que navega por temas con múltiples matices e identifica áreas para una mayor exploración donde una revisión sistemática es inapropiada. Las metodologías recomendadas para realizar una revisión de alcance son similares a las que se siguen para efectuar un meta análisis, que en síntesis representan un orden de pasos y sistematización de la información (Figura 1).

Figura 1. Ejemplo de una secuencia de pasos recomendados para la elaboración de una revisión de alcance o scoping review^[35].



Cabe mencionar que el protocolo se puede inscribir en algún repositorio, como por ejemplo, el International Prospective Register of Systematic Reviews (PROSPERO, <https://www.crd.york.ac.uk/prospero/>), el Joanna Briggs Institute (<https://jbi.global/systematic-review-register>), el Open Science Framework (OSF, <http://osf.io/>), el Research Registry (<https://www.researchregistry.com/>), el International Platform of Registered Systematic Review and Meta-analysis Protocols (INPLASY, <https://inplasy.com/>) o el protocols.io (<https://www.protocols.io/>)^[567]. La metodología puede incluir el sistema PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews)^[702], el sistema de Joanna Briggs Institute en su capítulo 11^[560], el propuesto por Arksey y O'Malley^[35], u otros^[561].

Adicionalmente, es posible que no en todas las revisiones de alcance se realicen análisis estadísticos inferenciales; sin embargo, se describen los diseños de los estudios, las poblaciones, los entornos donde se lleva a cabo la intervención, la intervención en sí misma; es decir, la prescripción del ejercicio o entrenamiento (i.e., frecuencia, intensidad, duración, modalidad, progresión, etc.). Por lo tanto, las revisiones de alcance generalmente son de naturaleza descriptiva, aunque también existen revisiones de alcance con análisis estadísticos complejos y robustos^[326]. En las Ciencias del Movimiento Humano, recientemente se han publicado revisiones de alcance, las cuales varían en cuanto a su propuesta como pregunta de investigación o como objetivo de esta (Tabla 1).

Tabla 1. Ejemplos de evidencias sintetizadas por medio de revisiones de alcance en las Ciencias del Movimiento Humano. Algunos autores proponen un objetivo o varios, mientras que otros, tratan de brindar respuesta a una pregunta concreta. El planteamiento de una u otra, depende de la metodología adoptada por los investigadores.

Autores	Objetivo o pregunta de investigación
Bell et al. ^[67]	¿Qué se sabe acerca del sobreentrenamiento en los deportes de fuerza y las poblaciones de entrenamiento contra resistencia?
Carmody et al. ^[146]	Realizar una revisión de alcance que proporcione una visión general de las condiciones de salud que ocurren en los futbolistas profesionales retirados, tanto hombres como mujeres, con énfasis en las condiciones de salud musculoesqueléticas, mentales, neurocognitivas, cardiovasculares y reproductivas.
Carrard et al. ^[148]	Revisar sistemáticamente la literatura actual y mapear los biomarcadores y las herramientas reportadas como potencialmente diagnósticas del Síndrome de Sobreentrenamiento.
Brennan et al. ^[109]	¿Cuáles son las barreras y los facilitadores de la participación en actividad física de los adultos mayores de 18 años que viven con diabetes mellitus tipo 1 independientemente del entorno de atención?
Grgic et al. ^[298]	<ul style="list-style-type: none"> a) Revisar y resumir los hallazgos de los estudios que emplearon el modelo de sustitución isotemporal en la investigación del sueño, el comportamiento sedentario y la actividad física. b) Evaluar la calidad metodológica de los estudios. c) Sugerir futuras direcciones de investigación en esta área.
Haverinen et al. ^[320]	Sintetizar las características de las intervenciones de actividad física comunitaria en los países nórdicos, centrándose en estudios que tienen como objetivo aumentar la actividad física o el acondicionamiento físico cardiorrespiratorio en un grupo grande de personas a largo plazo.
Haynes et al. ^[326]	Evaluar si la evolución temporal de la dislipidemia difiere entre sexos.
Huard Pelletier et al. ^[355]	¿Cuáles son las consecuencias de jugar videojuegos en los indicadores y comportamientos de salud física en jugadores de videojuegos saludables y cómo se investiga actualmente?
	Continúa...

Tabla 1. Continuación

Autores	Objetivo o pregunta de investigación
Ismail et al. ^[369]	a) ¿Los "exergames" influyen en la actividad física de los adultos mayores? b) ¿Qué juegos, tecnologías de juegos y ejercicios se utilizaron en los estudios? c) ¿Los "exergames" influyen en el equilibrio y la cognición en adultos mayores? d) ¿Cuáles son las experiencias de los adultos mayores con los "exergames"? e) ¿Es seguro el "exergaming" para los adultos mayores?
Luteberget & Gilgien ^[451]	Trazar y evaluar críticamente los métodos utilizados para validar diferentes sistemas de navegación por satélite global y sistemas de posicionamiento local utilizados en deportes de equipo.
Mercier et al. ^[489]	Enumerar y clasificar los factores de desempeño asociados con el rendimiento de los lanzadores de béisbol, describir los métodos utilizados para evaluar el desempeño de los lanzadores a través de todas las características del juego e identificar los factores individuales asociados con el desempeño de los lanzadores de béisbol.
Ramirez-Campillo et al. ^[591]	a) Actualizar una revisión previa sobre las principales metodologías características y deficiencias en la literatura sobre el entrenamiento de salto pliométrico. b) Recomendar, a la luz de las lagunas metodológicas identificadas, perspectivas de investigación futuras.
Rastogi et al. ^[601]	Mapear enfoques analíticos avanzados y sus variables multidimensionales utilizadas para proporcionar una imagen completa del comportamiento de la actividad física de un individuo.

En síntesis, las revisiones de alcance en las Ciencias del Movimiento Humano proporcionan información relevante acerca de las características de un cuerpo de evidencia, los conceptos que se utilizan y cómo se ha reportado un tema. A la vez pueden proporcionar una visión general de los campos de investigación y políticas específicas, así como respaldar las agendas de investigación y políticas e identificar áreas para la síntesis de pruebas posteriores^[561]. Los investigadores pueden realizar revisiones de alcance en lugar de revisiones sistemáticas cuando el propósito de la revisión sea identificar lagunas de conocimiento, analizar un cuerpo de literatura, o para aclarar conceptos. Las revisiones de alcance pueden ser útiles precursores de revisiones sistemáticas y pueden utilizarse para confirmar la relevancia de los criterios de inclusión y las posibles preguntas de investigación^[518].

1.3.3. Sombrilla

Una revisión de literatura general, de tipo “sombriilla” (“Umbrella Review”), ofrece una revisión rápida de la evidencia, y brinda al lector la posibilidad de tener un amplio panorama acerca de los diferentes enfoques que se analizan de una misma temática; además, proporcionan una evaluación general del conjunto de información disponible para un tema en particular^[36]. Se puede decir que una revisión tipo sombrilla permite:

- a. Analizar de manera continua las revisiones sistemáticas (RS) disponibles y las revisiones sistemáticas con meta análisis (RSMA) publicadas con relación a un área de conocimiento o un tema en particular.
- b. Evaluar la calidad metodológica de las RS y RSMA e identificar sus fortalezas y limitaciones.
- c. Resumir la evidencia principal presentada en las RS y RSMA, identificar posibles vacíos en la literatura y proporcionar recomendaciones para futuras investigaciones acerca del tema de interés.

Un diseño de investigación que permita cumplir con estos objetivos puede ayudar a proporcionar una amplia imagen de la investigación realizada hasta la fecha (pues utiliza como materia prima de trabajo las RS y RSMA, que de por sí incluyen a una gran cantidad de estudios que no limitan su diseño y cuya calidad está previamente evaluada), así como, definir nuevas líneas y oportunidades de investigación futuras. También ayudará a proporcionar información relevante acerca de la coherencia o no de los resultados, al resumir la evidencia del meta análisis sobre el tema.

La revisión de literatura en este tipo de estudio puede utilizar una metodología de revisión rápida para identificar, evaluar y sintetizar RS y RSMA^{[73][306]}. Los criterios de inclusión y exclusión se pueden desarrollar utilizando una versión adaptada de las guías PICO (**P**oblación, **I**ntervención, marco de **C**omparación, resultado **O**bservado), y añadir otros criterios específicos, como, por ejemplo, idioma, fecha de publicación, tipo de publicación y escenarios o contextos. Esta metodología busca identificar cierto tipo de estudios, de acuerdo con el problema de investigación específico que cada estudio se plantee. La búsqueda de literatura en una revisión sombrilla puede incluir bases de datos electrónicas y búsqueda manual de artículos impresos; una vez que esté disponible una lista de estudios potencialmente útiles para completar la revisión, ésta se puede compartir con algún grupo de referencia de expertos en el tema para comprobar la existencia de estudios adicionales. Como complemento a la búsqueda de literatura y con el objetivo de poder replicar las búsquedas, se pueden añadir al artículo, archivos o documentos con las búsquedas completas llevadas a cabo en cada base de datos.

La recolección o extracción de datos o información relevante se realizará utilizando tablas que deberán incluir, entre otra información: autores, fecha, país, elementos PICO-S y los resultados más relevantes. Los autores de cada estudio por incluir en la revisión sombrilla, podrán ser contactados por los revisores para solicitar información faltante o aclaración cuando sea necesario, y donde la información es considerada esencial. Todos los revisores extraerán los datos de un conjunto de estudios elegibles. El 10% de todos los trabajos serán seleccionados al azar y los datos de estos serán extraídos de forma independiente por un segundo revisor. Con el objetivo de evitar

desigualdades en la extracción de datos, debe conseguirse una similitud “aceptable” de al menos el 80% entre ambos revisores. La calidad final de las RS y RSMA serán evaluadas mediante la lista AMSTAR 2^[645].

Según Popay et al.^[574], debido al amplio alcance y heterogeneidad de los estudios y temas incluidos en este tipo de revisión, el análisis de datos se realiza de manera narrativa para resumir y explicar los hallazgos. Los resultados obtenidos, se evalúan de acuerdo con el enfoque CERQual (Confidence in Evidence from Reviews of Qualitative research), que valora^[426]:

- a. Las limitaciones metodológicas de los estudios que componen la evidencia.
- b. La relevancia de los hallazgos a la pregunta de revisión.
- c. La coherencia de los hallazgos.
- d. La idoneidad de los datos que respaldan los hallazgos.

Una vez que se describen en la sección de metodología los procesos relacionados con la selección y características de las RS y RSMA incluidas, se procede a describir detalladamente los resultados obtenidos; en algunos estudios también se le llama “sección de evidencias”. En todos los casos, la interpretación de las evidencias es interpretativa, considerando el resumen de la información de todos los artículos en tablas. A continuación, y a manera de ejemplo, se resume la información de cinco RS hipotéticas (Tabla 1).

Tabla 1. Resumen de RS y RSMA incluidas con datos hipotéticos.

Autores	Objetivo	Tipo de participantes y edad (años)	Número y tipo de estudios incluidos	Intervención de AF	Medición de adiposidad	Contexto o escenario de acción
Moncada et al. (2021)	Medir el efecto de la AF en la adiposidad de personas adultas mayores.	Adultos mayores (65-87)	16 SR + 8 RSMA	Caminata 10000 pasos (12 semanas)	DXA	Universitario
Araya et al. (2019)	Revisar efectos de un programa de AF estructurada en la adiposidad de escolares.	Niños y adolescentes (6-14)	163 cuasi experimentales y longitudinales	Clases de EF + incrementos de AF extracurricular	Bioimpedancia	Escolar
Carazo et al. (2020)	Identificar la evidencia sobre efectos de la AF en grasa corporal de adultos que practican taekwondo.	Adultos (18-50)	17 RCT, cuasi experimentales	Programa estructurado de Taekwondo (16 semanas)	Bioimpedancia	NR
Castillo et al. (2021)	Cuantificar la cantidad de AF que realizan adultos asmáticos con obesidad y sobrepeso.	Adultos (18-35)	34 RCT+ 2 RS	Programa de caminata no estructurado	DEXA	NR
Hernández et al. (2018)	Cuantificar el tipo de AF por acelerometría en niños con sobrepeso y obesidad.	Niños y adolescentes (4-10)	59 experimentales, cuasi experimentales y transversales	Acelerometría 8 días (METs, AFMV)	Índice peso/altura	Escolar

MVPA: DXA: absorciometría de rayos X de energía dual; Actividad Física Moderada-Vigorosa; NR: No Reportado; AF: Actividad Física; PE: Educación Física; RCT: Ensayos controlados aleatorizados; RS: Revisión Sistemática; RSMA: Revisión sistemática con Meta análisis.

Cuando adicionalmente se cuente con RSMA, y éstos puedan proveer información de tamaños de efecto, el resumen de la información puede también incluirse en tablas (Tabla 2) o en gráficos de tipo “forest plot”.

Tabla 2. Resumen de los tamaños de efectos recuperados de los metanálisis (datos ficticios).

Autores	Resultados			
	TE globales		TE específicos o individuales	
Moncada et al. (2021)	TE global g = -0.30	••	TE Mujeres g = -0.20 TE hombres g = -0.17	• •
Araya et al. (2019)	NR		TE (6-10años) g = -0.44, IC95% = 0.27, 0.61 TE (11-14años) g = 0.15, IC95% = -0.04, 0.34	•• ∅
Carazo et al. (2020)	TE global d = 0.367	••	NR	
Castillo et al. (2021)	TE global g = 0.09, IC95% = -0.05, 0.22	∅	TE (sobrepeso) g = -0.13, IC95% = 0.02, 0.24 TE (obesidad) g = -0.25, IC95% = -0.03, 0.48	•••
Hernández et al. (2018)	TE global d = 0.14	•	NR	

IC95%: Intervalos de confianza al 95%; d: tamaño de efecto *d* de Cohen; g: tamaño de efecto *g* de Hedges; TE: Tamaño del efecto; NR: No reportado. Efectos: ∅: Tamaño del efecto nulo; •: Tamaño del efecto pequeño; ••: Tamaño del efecto moderado; •••: Tamaño del efecto importante.

Una vez resumidos los principales hallazgos de las RS y RSMA analizados, es importante valorar la calidad metodológica de esos estudios que se van a incluir en la revisión sombrilla con el objetivo de brindar recomendaciones para futuras investigaciones en el área de interés. Esta sección generalmente sintetiza también este tipo de información en una tabla que evalúa los ítemes que componen la herramienta AMSTAR 2, y que se puede consultar en *Assessing the Methodological Quality of Systematic Reviews (AMSTAR)*^[645].

Cuando el objetivo del estudio sea realizar un análisis global de calidad e integrar los resultados obtenidos en estudios robustos de tipo RS y RSMA, se puede acudir a una revisión sombrilla. En las Ciencias del Movimiento Humano, actualmente se pueden encontrar algunas revisiones de sombrilla muy interesantes que brindan una guía basada en evidencia de los estudios realizados en un tema específico, así como la calidad y las limitaciones de estos estudios (Tabla 3). Un panorama de este estilo permite tomar mejores decisiones y utilizar de manera eficiente los recursos disponibles.

Tabla 3. Preguntas de investigación relacionadas con las Ciencias del Movimiento Humano, que han sido respondidas mediante estudios de tipo “sombriila” o Umbrella reviews.

Autores	Objetivo o pregunta de investigación
Lam et al. ^[415]	<ul style="list-style-type: none"> a. Abordar la pregunta de investigación de ¿qué tan efectivas son las diferentes intervenciones de comportamiento sedentario para reducir el comportamiento sedentario en diferentes poblaciones y entornos? b. Investigar la efectividad, los costos y la rentabilidad de las intervenciones de comportamiento sedentario para diferentes combinaciones de grupos y entornos meta para generar información sobre el impacto económico de la reducción del comportamiento sedentario para los tomadores de decisiones en el sistema de atención médica y los empleadores.
Clemente et al. ^[173]	Proporcionar información sobre la coherencia o no de los Juegos en Espacio Reducido (Small-Sided Games) para el desarrollo de cualidades físicas o habilidades técnico/tácticas, al resumir la evidencia de RS y RSMA sobre el tema.
Barbosa et al. ^[52]	<ul style="list-style-type: none"> a. Resumir la evidencia disponible presentada en RS y RSMA sobre el efecto de la Actividad Física (AF) en el Rendimiento Académico (RA) de niños en edad escolar y adolescentes, y explorar el efecto de los programas o modalidades de AF sobre RA en temas específicos.
Beynon et al. ^{[73]*}	Identificar los factores de riesgo asociados con el juego de azar y el problema de juego de azar, mediante dos preguntas: <ul style="list-style-type: none"> a. ¿Qué factores de riesgo están asociados con el juego de azar? b. ¿Qué factores de riesgo están asociados con los diferentes niveles de intensidad de juego de azar?
Cillekens et al. ^[171]	Evaluar la relación dosis-respuesta entre la Actividad Física Ocupacional (OPA) y la salud, y si estas relaciones difieren de la Actividad Física en el Tiempo Libre (LTPA).
Fiedler et al. ^[264]	Determinar los aspectos metodológicos de diferentes estrategias (e.g., fundamentos teóricos, técnicas de cambio de comportamiento, contextos sociales, e intervenciones en tiempo real) y su efecto en las intervenciones electrónicas (eHealth) y móviles (mHealth) realizadas recientemente en atención preventiva de salud, y con qué magnitud contribuyeron a la efectividad de la intervención.

* Protocolo de investigación.

1.3.4. De métodos mixtos

Las revisiones sistemáticas de métodos mixtos (RSMM) buscan sintetizar, en un mismo trabajo, evidencia proveniente de estudios individuales que implementan metodologías cuantitativas, cualitativas y mixtas. La inclusión de literatura mixta en una misma revisión sistemática permite una comprensión integral de los efectos o asociaciones entre las variables de interés en las Ciencias del Movimiento Humano^[174] que, como ciencia social, no siempre se limita a medidas estrictamente cuantitativas. En la categoría cuantitativa, una misma RSMM puede incluir estudios aleatorizados controlados, estudios no aleatorizados controlados y estudios transversales, mientras que, en la categoría cualitativa, podrían considerarse aquellos estudios prospectivos, transversales u observacionales que dentro de su metodología de recolección de información utilizan discusiones abiertas, grupos focales y entrevistas semiestructuradas^{[174][254][531]}.

En general, los pasos metodológicos recomendados para realizar una RSMM son similares a los expuestos en la Figura 1 de la sección de Revisiones de alcance, y se pueden seguir las guías PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses), las cuales han sido descritas anteriormente en este libro. Los criterios de inclusión de los estudios cuantitativos y cualitativos de las RSMM pueden establecerse utilizando los criterios PICO-S. De acuerdo con estos criterios, los lineamientos de inclusión con base en los grupos se utilizarían para la elección de estudios que, en efecto, incluyan grupos de comparación; por ejemplo, los estudios controlados aleatorizados o experimentales, por lo que el indicador de grupo podría no tomarse en consideración para la elección de estudios cualitativos.

La credibilidad de la información incluida en las RSMM puede optimizarse al determinar como criterios de exclusión aquellos estudios que no se encuentren publicados en texto completo o que sean considerados literatura gris (e.g., publicaciones que no hayan llevado un proceso de revisión por pares, tesis y disertaciones, resúmenes de conferencias, carteles o posters, memorias de congresos). Igualmente, la calidad metodológica y el riesgo de sesgo de los estudios de corte cuantitativo podrían valorarse utilizando las escalas RoB2^[671], TESTEX^[654] o la National Heart, Lung, and Blood Institute Quality Assessment Tool^[526]. Por otra parte, los estudios de corte cualitativo podrían juzgarse con las herramientas **Critical Appraisal Skills Program** (CASP, <https://casp-uk.net/casp-tools-checklists/>)^[186] o Grading of Recommendations Assessment, Development and Evaluation-Confidence in the Evidence from Reviews of Qualitative Research (GRADE-CerQual)^[521]. Existe una opción específica para valorar la calidad de los estudios de métodos mixtos y que puede utilizarse para estas revisiones sistemáticas, llamada Mixed Methods Appraisal Tool (MMAT, <https://www.mcgill.ca/familymed/research/projects/mmat>), que cuenta con varias versiones^[543].

Puede realizarse una RSMM con una cantidad variada de artículos. Esto va a depender de la popularidad y especificidad del tema, las palabras clave seleccionadas, la cantidad de bases de datos consultadas, y otros criterios de inclusión y estrategias de búsqueda. Es así como en las Ciencias del Movimiento Humano se han podido identificar RSMM realizadas con 9^[531] y 33 estudios^[193]. Al igual que otros tipos de revisiones sistemáticas, las de métodos mixtos también pueden complementarse con técnicas meta analíticas, de manera que el meta análisis puede incluir todas las variables dependientes consideradas en la revisión sistemática, o solamente una selección de variables de interés. Por ejemplo, en el estudio de Ng et al.^[531] se examinó el efecto que tenía el equipo de los parques para ejercitarse al aire libre (también conocidos como gimnasios al aire libre) en los niveles de actividad física, indicadores de función física, variables

psicosociales y de calidad de vida de personas adultas mayores que viven en la comunidad. Los autores evaluaron también la evidencia existente sobre el uso que se le daba a dichas máquinas por parte de esta población. Sin embargo, el meta análisis que incluyen en su estudio estuvo delimitado solamente a las variables obtenidas por medio de dos pruebas de función física, la prueba de 30 s de levantarse de la silla y la prueba de postura sobre una pierna, por ser indicadores relevantes de la capacidad funcional en las personas adultas mayores.

Extracción e integración de los datos

Un elemento particular de las RSMM es lo concerniente a la extracción e integración de los datos. Un primer nivel de integración de datos puede ser consistente entre publicaciones de corte cualitativo y cuantitativo. Por ejemplo, la sistematización de la información demográfica de las personas participantes y los diseños de estudio pueden codificarse en una hoja de cálculo sencilla incluyendo todos los estudios de la revisión. Posteriormente, para los estudios cualitativos, las variables de interés pueden categorizarse como temas comunes. Para este paso, se acude a métodos de síntesis temáticas para revisiones sistemáticas como el propuesto por Thomas y Harden^[693]. Para los estudios cuantitativos, se puede realizar un conteo de la frecuencia de las variables de interés, para finalmente calcular una suma ponderada que permita una síntesis de las variables por categorías de todos los estudios de la RSMM. Una vez construidas estas categorías temáticas, se puede identificar el nivel de importancia de cada una de ellas con base en la frecuencia en que son reportadas, o en el grado de pertinencia teórica establecida por los autores de los estudios individuales.

Un ejemplo de estos procedimientos es el reportado por Clifford et al.^[174]. En esa RSMM se buscaba reportar sistemáticamente la literatura relacionada con barreras, facilitadores y preferencias para ejercitarse en personas sobrevivientes de cáncer. Para ponderar la categorización de barreras, facilitadores y preferencias – llamados factores – identificadas de acuerdo con su relevancia, los autores desarrollaron una escala de tres puntos. Para los estudios cualitativos, cuando los factores que se reportaban con mayor frecuencia o eran identificados como factores más pertinentes por los autores del estudio analizado, se les asignaba un puntaje de 3. Aquellos factores con una frecuencia o relevancia moderada, se les asignaba un puntaje de 2, y aquellos factores a los que se les daba menor énfasis o que se reportaban con menor frecuencia se les asignaba un puntaje de 1.

En los casos en que la barrera, el facilitador o la preferencia no se reportara del todo, se les asignaba un 0, mientras que, por defecto, se asignaba un puntaje de 2 en los casos en que el factor fuese reportado, pero no pudiese ser calificado^[174]. Para los estudios cuantitativos, si una barrera, un facilitador o una preferencia era reportada por más del 50% de los participantes se le asignaba un puntaje de 3, si era reportada por una pequeña proporción de participantes (<25%) se le asignaba un 1. De esta manera, se evidencia la forma en que los hallazgos obtenidos por técnicas cualitativas y cuantitativas se llevaron a una métrica común, facilitando una comparación mixta, que posteriormente lleva a una conclusión integradora.

En los casos en que se revisan estudios individualmente asignando un puntaje para sintetizar las variables de interés, se recomienda que los puntajes sean asignados por más de un revisor/evaluador de manera independiente, y que posteriormente se calcule y reporte algún coeficiente de confiabilidad o concordancia inter-evaluador, como el coeficiente kappa^{[174][477]}. Esto es clave, considerando que existe evidencia que reporta una baja confiabilidad inter-evaluador cuando se juzga particularmente literatura cualitativa^[543]. Además, se recomienda contar con una tercera persona evaluadora que permita resolver cualquier discrepancia en la valoración/codificación de los estudios.

Algunas de las RSMM que se han identificado en las Ciencias del Movimiento Humano se describen en la Tabla 1. Estas han incluido búsquedas en bases de datos electrónicas como SPORTDiscus, Medline, CINAHL Plus de texto completo, Scopus, EMBASE, Cochrane Library, y PsycINFO^{[174][531]}. Además, una persona profesional en bibliotecología y ciencias de la información especializada en publicaciones científicas podría apoyar durante el proceso de desarrollo de términos de búsqueda estandarizados o normalizados e identificación de evidencia cuantitativa y cualitativa relevante para estas revisiones.

Tabla 1. Ejemplos de Revisiones Sistemáticas de Métodos Mixtos afines a las Ciencias del Movimiento Humano.

Referencia	Tema y Área	Objetivo	Características de los participantes en los estudios analizados	Número y tipo de estudios incluidos	Hallazgos relevantes que podrían informar futuros estudios
Aitchison et al. ^[7]	Experiencias de las personas con discapacidad que realizan deporte. Actividad Física y Deporte Adaptado.	Indagar las experiencias y los beneficios para la salud percibidos de la participación en el deporte en cuatro poblaciones con discapacidad: niños y adolescentes, adultos, atletas élite y veteranos con discapacidad.	Personas con discapacidad física, visual o intelectual (Niños y Adolescentes: 3-16 Adultos: 18-67 Atletas: 15-60 Veteranos: 20-50)	2 Cuantitativos 34 Cualitativos 3 Mixtos	Algunos aspectos positivos de la participación deportiva que se destacan en las cuatro poblaciones son: oportunidades de socialización, disfrute, sentido de libertad y de propósito, y que represente una esfera para desafiar estereotipos de la discapacidad. Existe una gran escasez de estudios en personas veteranas con alguna discapacidad, mientras que la población de atletas élite ansía más reconocimiento de sus logros y apoyo por parte del gobierno y de otras personas deportistas.
					Continúa...

Tabla 1. Continuación.

Referencia	Tema y Área	Objetivo	Características de los participantes en los estudios analizados	Número y tipo de estudios incluidos	Hallazgos relevantes que podrían informar futuros estudios
Dakic et al. ^[193]	Padecimientos del suelo pélvico femenino y participación en ejercicio físico. Ejercicio y Salud.	Revisar el efecto de los síntomas de suelo pélvico en la participación en ejercicio físico en mujeres y explorar dichos síntomas como una barrera para ejercitarse.	Mujeres jóvenes, adultas, de mediana edad y adultas mayores (16-83)	24 Cuantitativos 5 Cualitativos 4 Mixtos	Aproximadamente la mitad de las mujeres con síntomas (e.g., incontinencia urinaria) reportaban que dichas molestias afectaban su participación en el ejercicio físico, y que detuvieron o modificaron su participación en actividades de alta intensidad, ejercicios en grupo y deporte organizado. Es muy limitada la evidencia sobre los cambios de conducta de ejercicio, por lo que futuros estudios deberían reportar los cambios que se realizan en respuesta a los síntomas (i.e., interrumpir o modificar las actividades) y qué modificaciones específicas se hacen (tipo, intensidad, duración), el tipo de ejercicio o deporte que se deja de hacer y el grado de molestia cuando se interrumpe la participación en el ejercicio preferido.
Farrance et al. ^[254]	Adherencia y participación en programas grupales de ejercicio físico en la comunidad. Ejercicio y Salud en entornos comunitarios.	Indagar los puntos de vista y la adherencia de las personas mayores que asisten a programas de ejercicio a nivel comunitario de ≥ 6 meses de duración.	Adultos mayores (Cuantitativos: >65 y Cualitativos: >60)	5 Cuantitativos 3 Cualitativos 2 Mixtos	La incorporación de los puntos de vista de las personas adultas mayores en el diseño de programas de ejercicio podría llevar a intervenciones innovadoras que permitan una adherencia prolongada en las comunidades.
					Continuación...

Tabla 1. Continuación.

Referencia	Tema y Área	Objetivo	Características de los participantes en los estudios analizados	Número y tipo de estudios incluidos	Hallazgos relevantes que podrían informar futuros estudios
Fritz et al. ^[271]	Terapia física y ejercicio en la enfermedad de Huntington. Ejercicio y Salud en entornos comunitarios y clínicos.	Indagar la efectividad de la terapia física y las intervenciones de ejercicio en personas con enfermedad de Huntington y examinar las percepciones de los pacientes, las familias, y las personas cuidadoras respecto a dichas intervenciones.	Adultos en distintas etapas, manifiestas o confirmadas, de la enfermedad de Huntington (Cuantitativos: promedios entre 28–57 y Cualitativos: NR)	18 Cuantitativos 2 Cualitativos	Tanto las personas con el padecimiento como las cuidadoras perciben que el ejercicio es beneficioso, destacando beneficios físicos (e.g., caminata, balance, control motor, reducción de caídas) y sociales (e.g., auto confianza, independencia, socialización con la familia y amigos, estado de ánimo, bienestar). En general, las intervenciones de ejercicio son seguras y factibles. Futuros estudios deberían determinar la dosificación óptima y los efectos a largo plazo del ejercicio aeróbico y contra resistencia en la capacidad funcional y otras variables relacionadas con la salud como el sueño, la función intestinal y la presión arterial en esta población.
Voelker et al. ^[726]	El papel del entrenador en la conducta alimentaria del atleta. Rendimiento y Psicología del Deporte.	Sintetizar el conocimiento actual acerca de las conductas y prácticas específicas de los entrenadores asociadas con la patología alimentaria de los atletas en múltiples esferas de influencia e interacción.	Atletas activos y retirados (Cuantitativos: 12.35-25.03 y Cualitativos: 14.6-28)	21 Cuantitativos 13 Cualitativos 2 Mixtos	Los atletas perciben que los entrenadores influyen sus conductas alimentarias patológicas por medio de factores relacionados con la ética deportiva imperante, la calidad y las características de la relación entrenador-atleta, comentarios sobre el cuerpo de los atletas, asesoría y prescripción dietética, mediciones corporales y pesaje, y monitoreo de la alimentación por parte del entrenador.

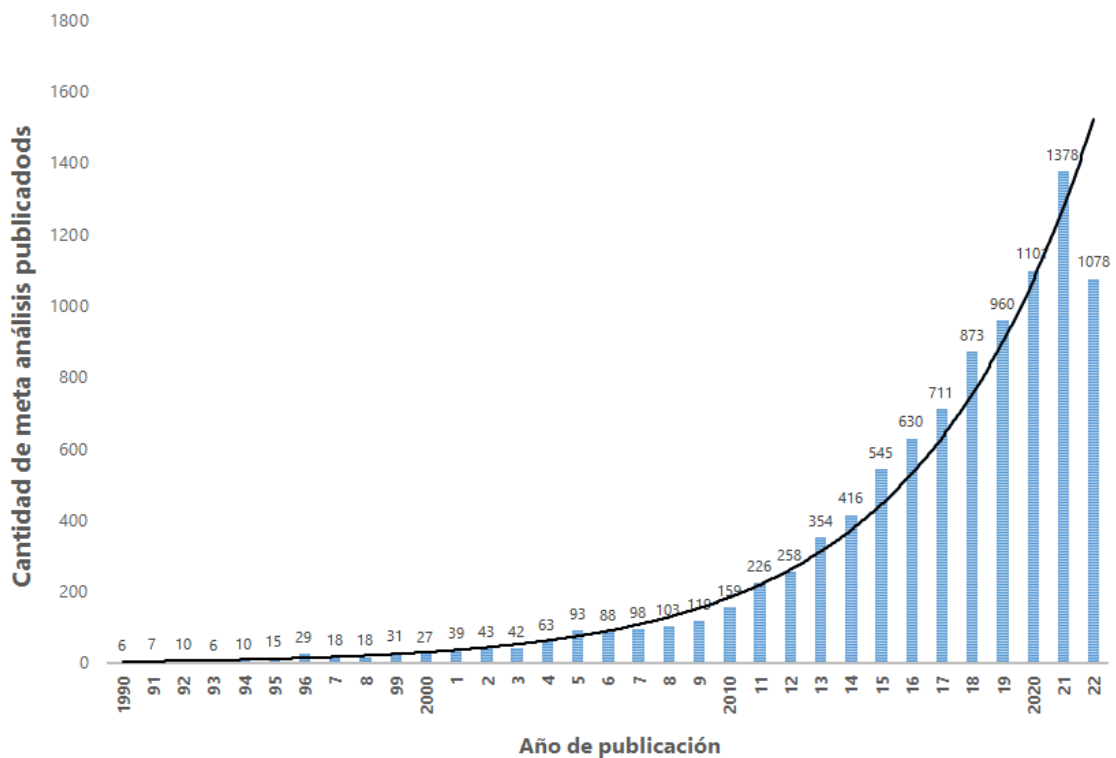
NR: No Reportado.

1.4. Revisiones meta analíticas

1.4.1. Meta análisis

El meta análisis es una técnica estadística que permite combinar los hallazgos de diferentes estudios bajo una métrica común llamada tamaño del efecto^[502]. La técnica meta analítica se encuentra en la cúspide de la pirámide de la calidad de la evidencia científica^{[125][126]}, y en la base de datos PubMed se encuentran indexados 9554 revisiones meta analíticas relacionadas con la palabra ejercicio ("exercise"), lo cual muestra un crecimiento exponencial (Figura 1). De hecho, del total de meta análisis indexados, 8300 (83%) han sido publicados del año 2012 al 2022.

Figura 1. Meta análisis sobre ejercicio publicados en la base de datos PubMed hasta noviembre de 2022.



Existen autores que describen detalladamente la forma en que se puede resumir dicha evidencia^{[95][393][497][502][503][643]}. También se reportan diferentes tipos de meta análisis, los cuales se pueden clasificar en los que producen tamaños de efecto de la familia de las relaciones y los que producen tamaños de efecto de la familia de las diferencias. Entre los meta análisis de la familia de las relaciones o correlaciones se encuentran los de tipo epidemiológico y correlacional, cuyo objetivo es establecer una asociación no causal entre las variables. Por otra parte, los meta análisis de la familia de las diferencias incluyen el de datos agregados (DA), en el cual se agrupan y sintetizan los datos de estudios individuales^{[112][141][142][147][167][282][304][378][513][689]}; con este tipo de meta análisis se pueden establecer asociaciones de causa-efecto ya que por lo general comparan diferencias entre grupos y condiciones experimentales con algún tipo de grupo o condición de control.

Otro miembro de esa familia es el meta análisis de datos individuales de los participantes (IPD)^{[116][128][321][366]}, en el cual se obtienen, agrupan y sintetizan los datos de cada persona participante en el estudio. Se ha descrito que un meta análisis basado en IPD requiere de mayor cantidad de tiempo y puede generar un mayor costo que un meta análisis de DA, y además existen limitaciones para compartir los datos que pueden ser explicadas por cláusulas legales en algunos contratos para realizar investigación, el egoísmo de los autores, o el temor a ser descubiertos por publicar datos falsos, entre otros^{[180][716]}.

Sin embargo, el meta análisis IPD es más confiable y puede responder detalladamente más preguntas que el de DA a través de análisis de subgrupos y el uso de métodos de análisis estadísticos diferentes, evitando, por ejemplo, la Paradoja de Simpson^[217], que establece que la dirección de una asociación entre dos variables se invierte después de agrupar utilizando una covariable. Por ejemplo, se podría encontrar que un tipo de entrenamiento físico puede ser efectivo independientemente para hombres y mujeres, pero sería ineficaz cuando se combinan los datos de hombres y mujeres. Por tales motivos, el meta análisis IPD se le considera superior al de DA^[656].

1.4.2. Meta análisis armonizado

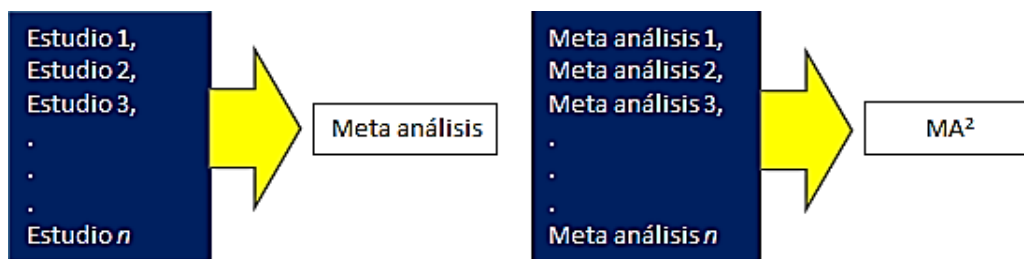
A pesar de los beneficios de los meta análisis de DA e IPD, existe un reto importante cuando se realizan meta análisis IPD, y es cuando las variables de interés fueron medidas con métricas distintas en cada estudio individual. Es por ello que se introdujo el concepto de armonización, con el cual se describen los procedimientos para poner las variables en una sola métrica o escala que permita agrupar los datos de múltiples estudios. La armonización de datos aumenta la validez de los hallazgos ya que en cada estudio publicado se utilizaron métodos de recolección de datos distintos, y el proceso de armonización involucra precisamente reprocesar todos los datos de acuerdo con un protocolo estandarizado^[244].

En el año 2020 se publicó un estudio en el que se utilizó la armonización de datos acerca de las asociaciones conjuntas de actividad física medida por medio de acelerómetros y el tiempo sedentario y la mortalidad por todas las causas^[244]. En ese meta análisis armonizado, se analizaron más de 44000 personas, y se tuvo que solicitar los datos individuales de acelerometría de los participantes de siete estudios. El grupo de investigadores recopiló los datos provenientes de acelerómetros marca Actigraph® y Actical®, y entre otros procedimientos, armonizaron los datos medidos en el eje vertical, en epochs, durante 60 s. Esta armonización les permitió llegar a la conclusión de que un mayor tiempo sedentario se asocia con una mayor mortalidad en las personas menos activas cuando se mide el tiempo con acelerometría. También pudieron concluir que entre 30 y 40 min diarios de actividad física atenúan esa maligna asociación^[244]. Estas son conclusiones más precisas y robustas que las que se habrían obtenido con un meta análisis de DA.

1.4.3. Meta análisis²

Como se describió anteriormente, en las Ciencias del Movimiento Humano, la técnica del meta análisis permite resumir cuantitativamente una amplia cantidad de estudios individuales para generar conclusiones acerca, por ejemplo, de la efectividad de algún tipo de entrenamiento físico o de alguna intervención cognitiva^[502]. Para ello, recurre a la selección, identificación, extracción de información y análisis estadístico de cada estudio particular. Posteriormente, a través del cálculo del tamaño de efecto, se crea una métrica que permite conocer la dirección y magnitud del efecto de la intervención sobre una variable de interés, por ejemplo, la fuerza muscular. Como se mencionó anteriormente, en la actualidad, esta técnica de síntesis de información se ubica en la cúspide de la pirámide de la calidad de la evidencia^[125], y existen áreas del conocimiento en donde se han publicado al menos dos meta análisis acerca del mismo tema. Ya que dos meta análisis acerca de un mismo tema podrían llegar a conclusiones distintas, se ha propuesto la elaboración de meta análisis de meta análisis, lo que se podría conocer como meta meta análisis o meta análisis al cuadrado (MA²) (Figura 1).

Figura 1. Síntesis de estudios individuales en meta análisis. El siguiente paso representa la síntesis de evidencia a partir de meta análisis individuales (MA²).



El MA² consiste, nuevamente, en la selección, identificación, extracción de información y análisis estadístico de cada meta análisis particular, con el propósito de informar con mayor precisión, acerca del efecto de una intervención sobre una variable de interés. Por ejemplo, durante muchos años se ha estudiado el efecto del ejercicio en el sueño en las personas adultas, lo que ha permitido generar suficientes publicaciones que posteriormente fueron meta analizadas. Sin embargo, se realizó un MA² para conocer el efecto del ejercicio en variables de sueño en adultos a partir de meta análisis previos^[391]. Esto también lo han hecho otros investigadores en temas como el efecto del ejercicio en la depresión y la ansiedad^[602].

La metodología para realizar un MA² sigue los mismos pasos que se utilizan para realizar un meta análisis. Así, se recomienda inscribir el protocolo en un sitio reconocido (e.g., PROSPERO, Joanna Briggs Institute, **O**pen **S**cience **F**ramework [OSF]), y seguir las recomendaciones PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis). Entre estas recomendaciones, se describen los criterios de inclusión y exclusión de los meta análisis previos, el diseño de los estudios meta analizados, el sexo de los participantes, la edad y la métrica de tamaño de efecto calculado en esos meta análisis. También se mencionan las bases de datos utilizadas para realizar la búsqueda de meta análisis. En el MA² de ejercicio y sueño^[391] se buscaron meta análisis en 8 bases de datos, mientras que en el MA² de ejercicio y depresión y ansiedad^[602] solamente en tres bases; sin embargo, no existe una guía estandarizada para definir la cantidad

de bases de datos que se deban revisar. Se ha mencionado que se deben consultar al menos tres bases de datos especializadas en la temática para completar un meta análisis para garantizarse una apropiada cobertura de información y de diversidad de fuentes de información.

Posteriormente se siguen procedimientos estandarizados para seleccionar los meta análisis, codificar las variables de interés, evaluar la calidad metodológica de cada meta análisis y finalmente calcular el tamaño del efecto y los sesgos de publicación.

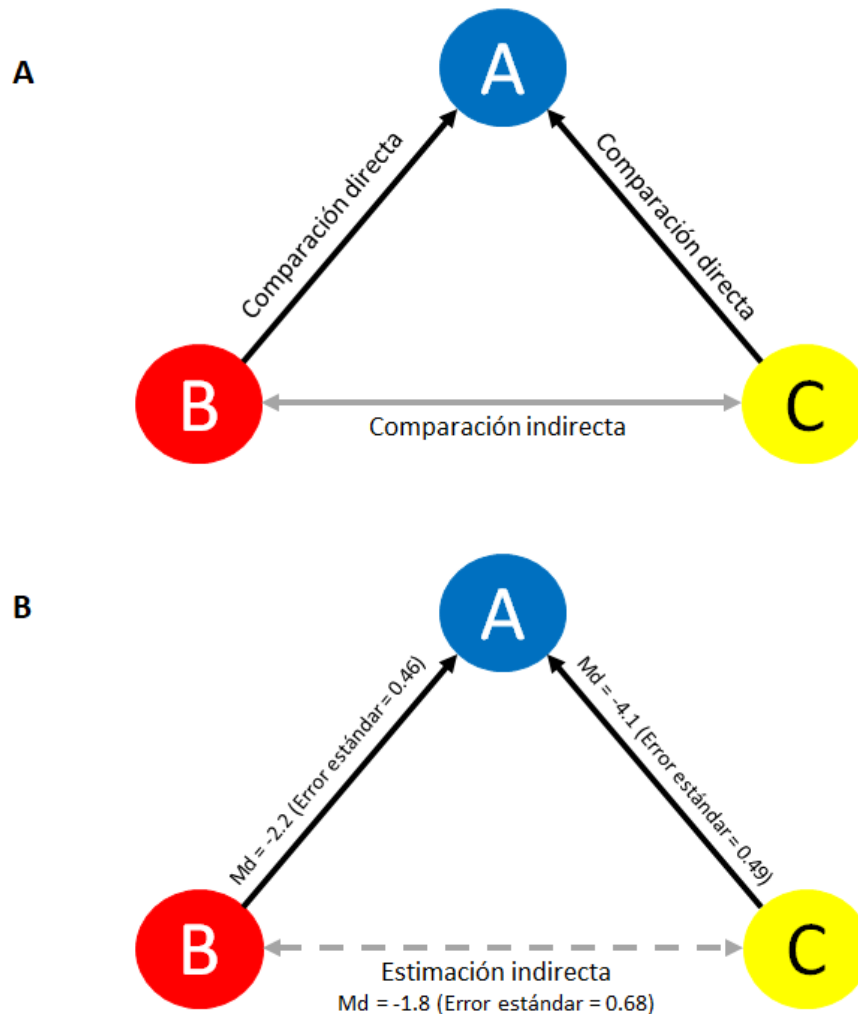
Cabe mencionar que un MA² se podría realizar únicamente con los datos provenientes de los meta análisis^[391] o se podrían tomar todos los estudios individuales reportados en cada meta análisis y hacer un nuevo meta análisis. Esto podría ser una opción válida para no contar con estudios duplicados en el nuevo MA².

En síntesis, cuando la evidencia científica acerca de un tema particular ha producido una amplia cantidad de publicaciones y puede ser resumida, primero se debe meta analizar. Sin embargo, cuando ya existen meta análisis acerca del mismo tema y las conclusiones derivadas de los mismos son diferentes o heterogéneas, es posible que un MA² permita llegar a un consenso acerca del efecto de la intervención sobre la variable dependiente. Es probable que en los próximos años se observe un aumento paulatino en la cantidad de publicaciones de MA² en las Ciencias del Movimiento Humano debido a la gran cantidad de meta análisis publicados recientemente.

1.4.4. Meta análisis en red

El meta análisis en red (en inglés "Network meta-analysis"), es uno de los desarrollos recientes de la familia de técnicas de investigación meta analíticas. Se le puede definir como una técnica estadística que permite obtener estimaciones de los efectos relativos de múltiples intervenciones, combinando comparaciones directas e indirectas que consideran la red completa de los estudios disponibles, debiendo existir al menos una intervención o condición que sea común para cada cadena de comparación^{[47][218][290][496][676][765]}. Existen diferentes comparaciones que se realizan en un meta análisis en red (Figura 1).

Figura 1. Comparaciones típicas de un meta análisis en red. En la comparación directa, se compara el tratamiento **A** con el tratamiento **B**. En otro estudio se compara el tratamiento **A** con el tratamiento **C** (Panel A). En la comparación indirecta, la comparación del tratamiento **B** con el tratamiento **C** no existe en los estudios disponibles, pero se puede estimar indirectamente (Panel B)^{[208][496]}.



El meta análisis tradicional permite comparar dos grupos (e.g., experimental vs. control) o dos mediciones de un mismo grupo (e.g., pre-test vs. post-test), a través del tamaño del efecto (TE), en una muestra de estudios con el mismo diseño estadístico. El meta análisis en red va más allá; es como una meta revisión^[559], permitiendo comparar múltiples intervenciones de estudios incluso con distinto diseño, pero respetando ciertos supuestos:

- a. Homogeneidad y ausencia de sesgo en los estudios incluidos (similar al meta análisis tradicional).
- b. Exhaustividad y consideración (y análisis) de covariables confusoras o variables moderadoras (similar al meta análisis tradicional).
- c. Transitividad para la evidencia indirecta: si **B** es mejor que **A**, y **A** es mejor que **C**, se asume que **B** es mejor que **C**.
- d. Consistencia o coherencia de las evidencias directas e indirectas (resultados consistentes, en la misma dirección, no divergentes).

En la figura 2 se ilustran las diferencias generales entre el meta análisis tradicional y el meta análisis en red^[676]. Ambas técnicas son robustas; sin embargo, proporcionan información diferente, pues el meta análisis tradicional proporciona estimaciones puntuales basadas en las comparaciones directas. El meta análisis en red proporciona, además, estimaciones basadas en comparaciones indirectas, lo cual amplía el conocimiento que se puede generar con el meta análisis tradicional.

Existen elementos claves para realizar un meta análisis en red^{[53][155][163][542][564][756][765]}:

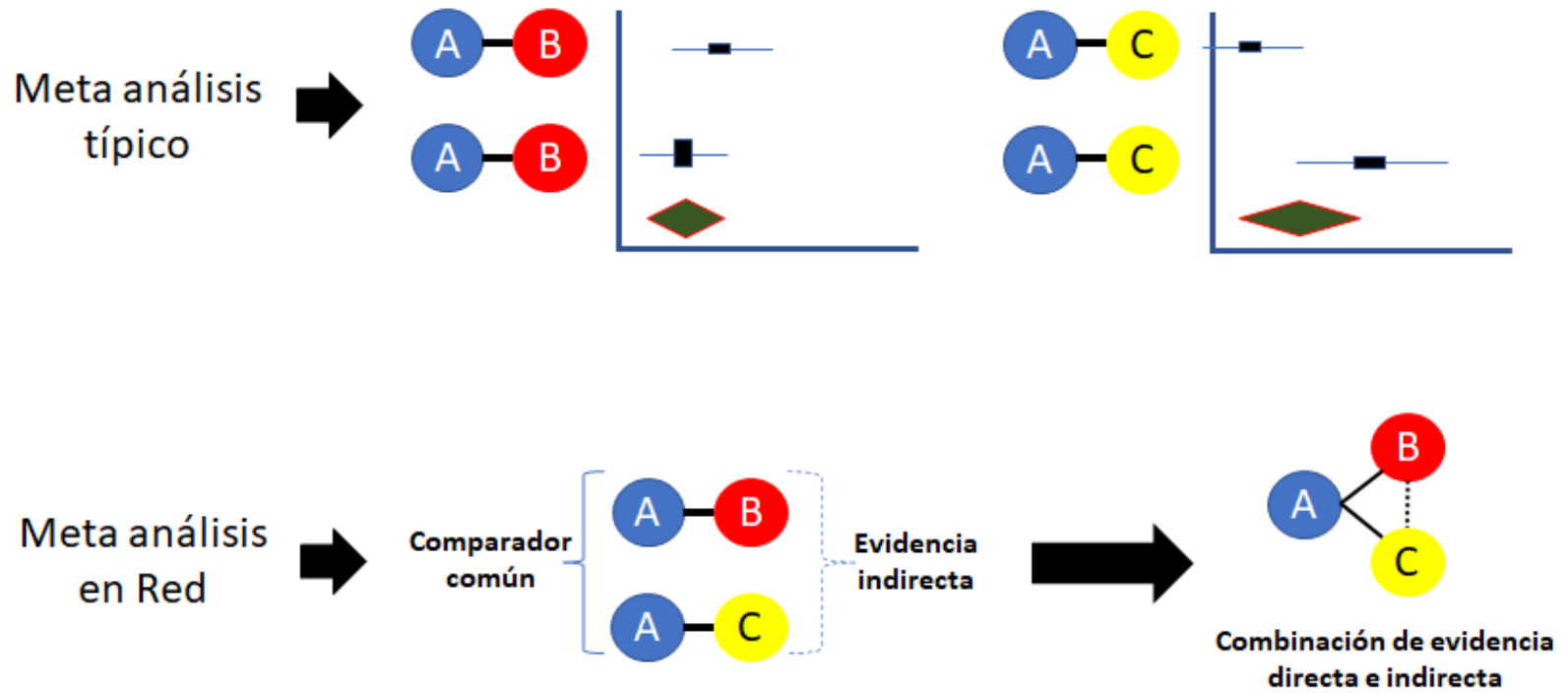
- a. Establecer la estructura metodológica del meta análisis en red: se recomienda el uso de la guía Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Network Meta-Analyses (PRISMA-NMA), que son recomendaciones específicas para esta técnica.
- b. Evaluar el riesgo de sesgo de cada estudio individual: existen herramientas como la Cochrane Collaboration Risk of Bias Tool, que es ideal para los meta análisis en red.
- c. Evaluar la calidad de la evidencia y ranquearla: se recomienda consultar el capítulo 11 del manual de la Colaboración Cochrane, el cual se encuentra disponible en la dirección: <https://training.cochrane.org/handbook/current/chapter-11>.
- d. Las intervenciones de ejercicio se pueden ranquear utilizando la herramienta llamada **S**urface **U**nder the **C**umulative **R**Anking curve (SUCRA, por sus siglas en inglés) y la clasificación media. El SUCRA es una estimación precisa de las probabilidades de clasificación acumuladas para un tratamiento superior en la clasificación.
- e. Aplicar los criterios de evaluación de la calidad de la evidencia tras el ranking de los tratamientos, mediante la herramienta llamada **G**rating of **R**ecommendations **A**ssessment, **D**evelopment and **E**valuation (GRADE).

También se puede usar la herramienta **Confidence In Network Meta-Analysis** (CINeMA). Este método evalúa la confianza en la evidencia mediante los siguientes dominios:

- a. Sesgo dentro del estudio.
- b. Sesgo de notificación.
- c. Evidencia indirecta.
- d. Imprecisión.
- e. Heterogeneidad.
- f. Incoherencia.

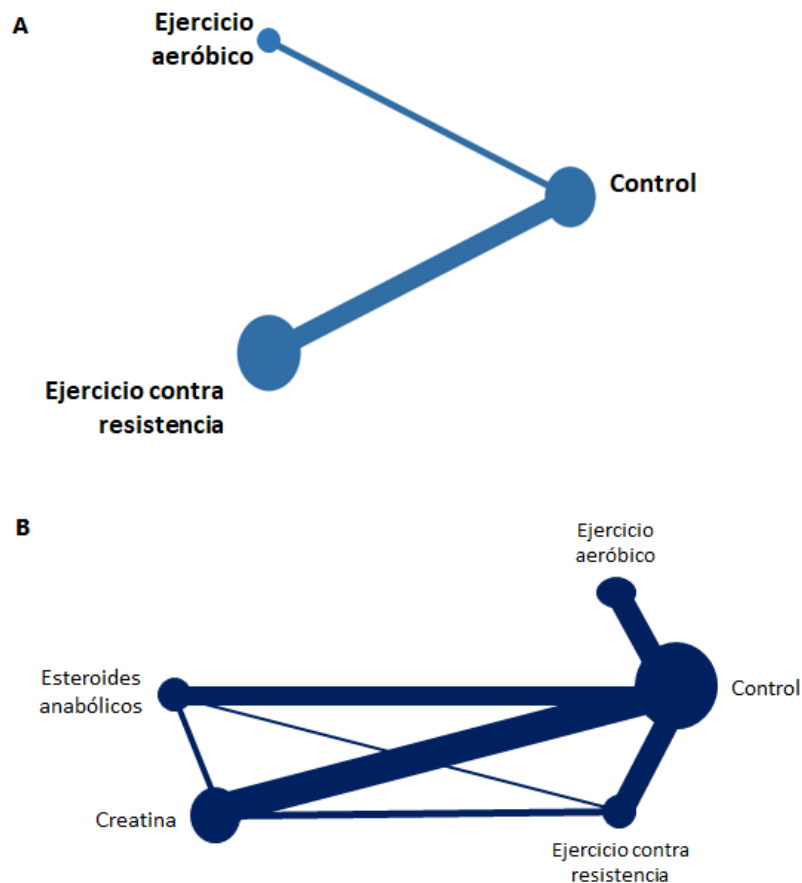
El efecto estimado para cada comparación dentro de cada resultado se clasifica luego como confianza muy baja, baja, moderada o alta^[548].

Figura 2. Esquema que muestra las diferencias principales entre el meta análisis típico y el meta análisis en red. El meta análisis tradicional compara los estudios de forma puntual, mientras que el meta análisis en red tiene la posibilidad de hacer comparaciones directas e indirectas, aportando estimaciones adicionales.



La representación gráfica geométrica de los meta análisis en red es particular y se puede realizar con el software especializado *WinBUGS*^[206]. Los gráficos poseen dos componentes principales: a) el tamaño del círculo y b) la línea que une los círculos. El grosor de las líneas y el tamaño de los círculos son proporcionales al número de comparaciones de estudios y al número de participantes en los brazos de tratamiento, respectivamente (Figura 3).

Figura 3. Ejemplo hipotético de la presentación gráfica de los resultados en un meta análisis en red acerca del entrenamiento físico y la fuerza muscular. El tamaño de los círculos y el grosor de las líneas representan la magnitud de la comparación entre los tratamientos; así como las comparaciones directas (Panel A) e indirectas (Panel B).



La interpretación de la figura 3 es muy simple e intuitiva. En el panel A, la evidencia correspondiente al efecto del ejercicio contra resistencia es mayor y más fuerte que la del ejercicio aeróbico, en comparación con sujetos control que no se ejercitan. Mientras que en el panel B se podría concluir que la evidencia del efecto de la suplementación con creatina, en comparación con sujetos control, es superior y más fuerte que el efecto del consumo de esteroides anabólicos y que los efectos del ejercicio aeróbico y contra resistencia sin suplementación. Además, se observa cómo la diferencia entre el efecto del ejercicio contra resistencia y los esteroides es débil.

En las Ciencias del Movimiento Humano se han publicado meta análisis en red (Tabla 1). Se espera que en los próximos años esta técnica se utilice aún más para resumir la evidencia acumulada proveniente de los estudios experimentales. Sin duda alguna, las conclusiones provenientes de los meta análisis en red serán de extraordinario provecho para los profesionales del área. Para profundizar más acerca del meta análisis en red, se recomienda consultar los trabajos de Dias et al.^[206], Dias y Caldwell^[207], Doosti-Irani et al.^[218] y Molina Arias^[496], en donde se podrán encontrar definiciones conceptuales, procedimientos metodológicos, y formas de graficación e interpretación de los hallazgos.

Tabla 1. Ejemplos de meta análisis en red en las Ciencias del Movimiento Humano.

Referencia	Propósito	Comparaciones
Challoumas et al. ^[155]	Proporcionar información sobre la efectividad de las modalidades de tratamiento disponibles en la tendinopatía rotuliana.	Una intervención en comparación con cualquier otra intervención, placebo o ningún tratamiento de la tendinopatía rotuliana.
Chen et al. ^[163]	Analizar y comparar la efectividad de diferentes tipos de ejercicio para reducir la fatiga relacionada con la esclerosis múltiple.	Control comparado con varias formas de ejercicio (e.g., acuático, aeróbico, contra resistencia, concurrente, danza, yoga, escalada).
Gianola et al. ^[285]	Evaluar la efectividad de las intervenciones para el dolor lumbar inespecífico agudo y subagudo.	Terapia inerte comparada con ejercicio, envoltura térmica, opiáceos, terapia manual, fármacos antiinflamatorios no esteroideos y relajantes musculares.
Hall et al. ^[314]	Comparar la efectividad de diferentes tipos de ejercicio sobre el bienestar psicológico en personas con artrosis de rodilla.	Control comparado con varias formas de ejercicio (e.g., aeróbico, contra resistencia, estiramiento) y de terapias mente-cuerpo y la combinación de las anteriores.
Han et al. ^[317]	Evaluar las terapias de ejercicios que tienen como objetivo mejorar la propiocepción en pacientes con inestabilidad crónica del tobillo.	Control comparado con ejercicio contra resistencia, solo ejercicio de equilibrio estático, ejercicio correctivo, ejercicio propioceptivo, solo ejercicio de equilibrio dinámico, ejercicio acuático, ejercicio de rehabilitación con aparatos ortopédicos, y ejercicio mixto de equilibrio estático/dinámico.
Ho et al. ^[342]	Comparar la efectividad de los videojuegos activos versus ningún comparador en varios niveles de actividad física y su resultado sobre el control de peso en niños y adolescentes.	Grupos con intervenciones de videojuegos activos y grupos que no tenían videojuegos activos.
Huang et al. ^[354]	Comparar y clasificar la efectividad de varias intervenciones de ejercicio sobre la función cognitiva en pacientes con deterioro cognitivo leve o demencia.	Todas las formas de ejercicio, incluido el aeróbico (e.g., caminar, correr, andar en bicicleta), el entrenamiento de resistencia (e.g., bandas elásticas, máquinas de pesas), el ejercicio multicomponente (e.g., aeróbico + entrenamiento de resistencia, entrenamiento del equilibrio), el ejercicio de la mente y el cuerpo (e.g., Tai Chi, yoga, baile), vibración de todo el cuerpo y ejercicio con los dedos.
		Continúa...

Tabla 1. Continuación.

Referencia	Propósito	Comparaciones
Lopez et al. ^[444]	Analizar el efecto del entrenamiento de fuerza realizado hasta el fallo voluntario con cargas bajas, moderadas y altas sobre la hipertrofia muscular y la fuerza muscular en adultos sanos.	Comparaciones entre carga alta, carga moderada, y carga baja.
Naci et al. ^[523]	Comparar el efecto de los regímenes de ejercicio y los medicamentos sobre la presión arterial sistólica.	Medicinas inhibitoras de la enzima convertidora de angiotensina, bloqueadores de los receptores de angiotensina-2, β -bloqueadores, bloqueadores de los canales de Ca^{++} comparados con intervenciones de resistencia, resistencia dinámica, resistencia isométrica y resistencia combinada y ejercicios de resistencia.
O'Donoghue et al. ^[538]	Evaluar y clasificar la eficacia comparativa de diferentes intervenciones de ejercicio sobre la antropometría, la aptitud cardiorrespiratoria y otros factores de riesgo metabólicos en personas obesas.	Control comparado con varias formas de ejercicio (e.g., aeróbico, contra resistencia o combinado de intensidad baja, moderada y alta).
Owen et al. ^[542]	Examinar la efectividad de modos específicos de entrenamiento físico en el dolor lumbar crónico no específico.	Controles que no hacían ejercicio comparados con diferentes tipos de entrenamiento físico (e.g., contra resistencia, estabilización, control motor, pilates, yoga, McKenzie, flexión, aeróbico, ejercicios en agua, estiramiento, entre otros).
Pescatello et al. ^[559]	Comparar los efectos del ejercicio, medicinas y la combinación de ejercicio+medicinas sobre la presión arterial en adultos con hipertensión.	Se compararon entre sí los estudios de ejercicio, los estudios de medicina y los estudios de ejercicio+medicinas.
Winters et al. ^[756]	Investigar la efectividad comparativa de todos los tratamientos para el dolor patelofemoral.	Todos los tipos de tratamiento (e.g., educación, ejercicio, órtesis, vendaje patelar/movilizaciones, la combinación de todos los anteriores) contra el control.

2.1. Evaluación de calidad y riesgo de sesgos

Se ha definido que el sesgo de publicación es un tipo de error sistemático cuando se resume evidencia científica que no representa la verdad subyacente^[646]. Por lo tanto, es mucho más probable que los estudios con resultados favorables se publiquen y, por lo tanto, exageren la evidencia que posteriormente es sintetizada en las revisiones sistemáticas con meta análisis. El sesgo atenta directamente contra la validez de las conclusiones que se puedan derivar de un estudio. De hecho, en una dura y franca reflexión se indica que:

... "actualmente, existe una producción masiva de revisiones sistemáticas y meta análisis innecesarios, engañosos y conflictivos. En lugar de promover la medicina y la atención de la salud basadas en la evidencia, estos instrumentos a menudo sirven principalmente como unidades publicables o herramientas de marketing de fácil producción. Las revisiones sistemáticas y los meta análisis subóptimos pueden ser perjudiciales dado el gran prestigio e influencia que han adquirido este tipo de estudios. La publicación de revisiones sistemáticas y meta análisis debe realinearse para eliminar sesgos e intereses creados e integrarlos mejor con la producción primaria de evidencia" (p. 485)^[364]

De hecho, de la producción de meta análisis reportados en el año 2014 en la base de datos PubMed, se considera que el 3% se podían considerar como decentes y con utilidad clínica, el 20% tenían errores que no se podían corregir, el 13% era engañosos o sesgados, el 17% eran decentes, pero poco útiles, el 27% eran redundantes e innecesarios, y el 20% no eran publicables^[364]. Estas cifras permiten reflexionar acerca de la pertinencia y calidad metodológica de las revisiones sistemáticas de literatura que se producen en las Ciencias del Movimiento Humano.

Se ha definido que "las prácticas de investigación cuestionables son prácticas intencionales y no intencionales que pueden ocurrir al diseñar, realizar, analizar y reportar investigaciones, produciendo resultados de estudios sesgados" (p. 1)^[129]. Así, cuando se realiza una investigación, tanto los autores de éstas como el público que las lee, deben tener presente que pueden existir sesgos metodológicos que podrían invalidar o degradar la calidad de dicha investigación, como por ejemplo, el reporte de falsos positivos (i.e., decir que el tratamiento funciona cuando no es cierto) y tamaños de efecto inflados (i.e., un efecto exagerado de una intervención cuyo verdadero efecto es pequeño, nulo o negativo)^{[131][685]}. Aunque no existe investigación perfecta, es importante identificar algunos elementos que atentan contra la calidad de las investigaciones para que quienes toman decisiones puedan hacerlo de forma más informada.

Existen al menos tres tipos de prácticas de investigación cuestionables:

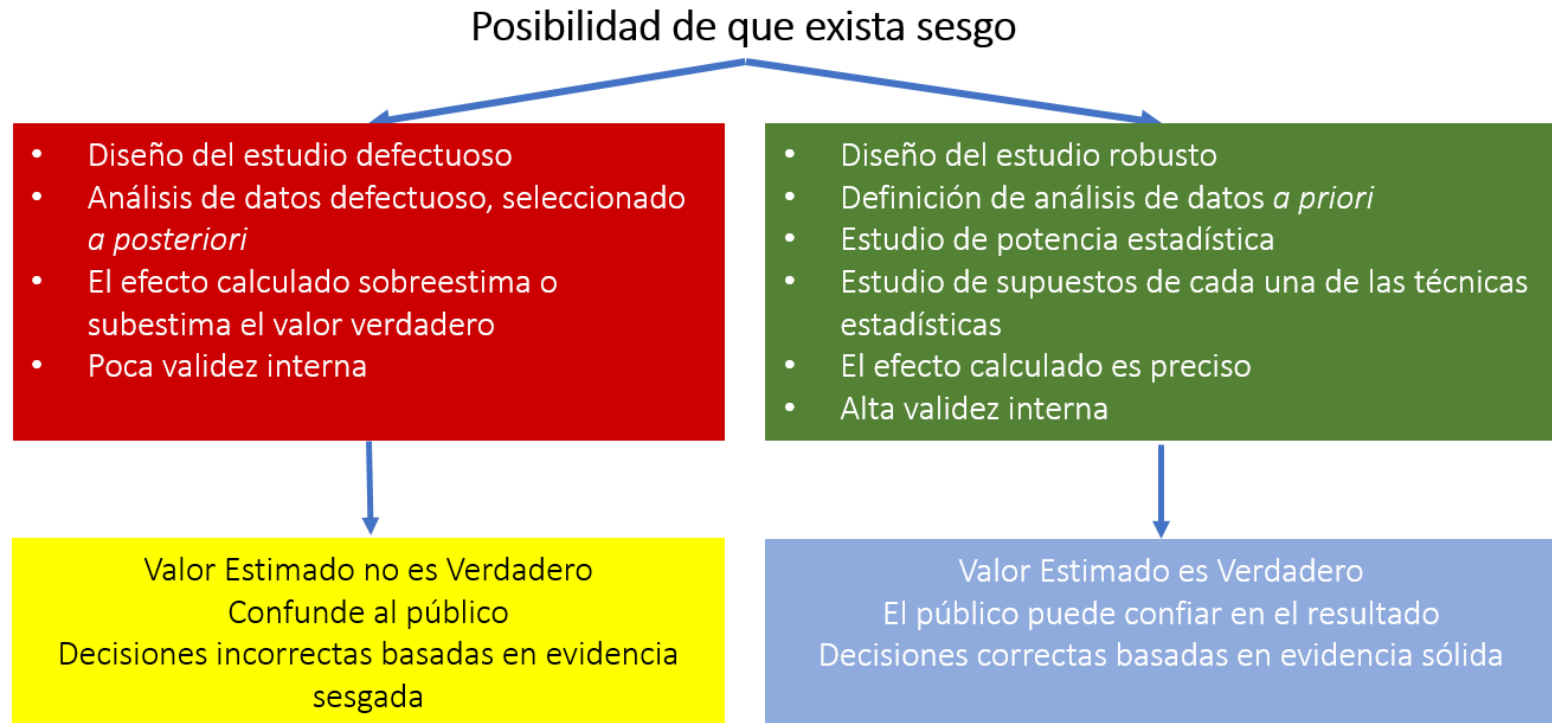
- a) Piratería de *P*.
- b) Selección selectiva.
- c) Hipotetizar después de que se conocen los resultados^[129].

Con el pirateo o "hacked" de P (i.e., significancia estadística), los investigadores exploran una serie de análisis estadísticos y escogen los análisis de datos más novedosos, eliminan valores extremos ("outliers") y covariables, y, a menudo, solamente escogen los datos estadísticamente significativos ($p \leq 0.05$) para reportar sus resultados^[149].

Con la selección selectiva de resultados o el cambio de resultado, los investigadores utilizan múltiples resultados en un estudio de investigación, pero solamente informan el más conveniente.

Finalmente, "Hipotetizar después de que se conocen los resultados" (en inglés, 'HARKing'), es el comportamiento de generar una hipótesis de estudio para explicar el (los) resultado (s) de un estudio de investigación, ya sea estadísticamente significativo o no significativo, solo después de que los datos han sido analizados y cuando ya se conocen los resultados. Aunque existen otros tipos de prácticas de investigación cuestionables, es evidente que pueden existir sesgos en las publicaciones que se revisan día a día para informar al público y tomar decisiones importantes para el rendimiento físico de los atletas, la elaboración de un programa de enseñanza de la Educación Física o para la salud de la población general (Figura 1).

Figura 1. El sesgo en la investigación puede producir conclusiones erróneas que perjudican la labor de los consumidores del conocimiento publicado ya que se basan en estimaciones falsas o antojadizas de acuerdo con los intereses de los investigadores o de quienes financian sus estudios^[130].



Actualmente existen herramientas que permiten evaluar el riesgo de sesgos; por ejemplo, la escala RoB2, que es la segunda edición revisada de la herramienta “Cochrane Risk of Bias” para estudios clínicos aleatorizados^[338]. RoB2 es una herramienta centrada en los resultados y basada en el dominio que evalúa el riesgo de sesgo en los resultados en ensayos de grupos paralelos aleatorizados individualmente, ensayos cruzados aleatorios y grupales^[671]. RoB2 consiste de cinco dominios de riesgo de sesgo para ensayos de grupos paralelos aleatorizados individualmente: a) Sesgo que surge del proceso de asignación al azar, b) Sesgo debido a desviaciones de las intervenciones previstas, c) Sesgo debido a datos de resultados faltantes, d) Sesgo en la medición del resultado, y e) Sesgo en la selección de los resultados reportados.

También existen otras escalas que permiten evaluar el riesgo de sesgo, entre las que se destacan la escala de PEDro^[457], la de Downs y Black^[222] y la escala TESTEX (en inglés, **T**ool for the **a**ssessment of **S**tudy **q**uali**T**y and reporting in **EX**ercise)^[654]. Esta última es específica para medir la calidad y el reporte de los estudios que utilizan ejercicio físico, por lo que actualmente se recomienda para resumir sistemáticamente las investigaciones en las que se utilizan programas de entrenamiento físico como variable independiente para evaluar su efecto en variables dependientes físicas (e.g., fuerza, potencia, capacidad aeróbica), perceptuales/cognitivas (e.g., tensión, depresión, ansiedad, fortaleza mental), o respuestas y adaptaciones motrices.

Cuando se elaboran revisiones sistemáticas y meta análisis, se deben identificar posibles sesgos, ya que estas metodologías pretenden resumir y sintetizar numerosos hallazgos provenientes de una amplia cantidad de estudios. Por ejemplo, para la realización de meta análisis, tradicionalmente se utilizaron los modelos de efectos fijos o de efecto común (“fixed-effects” “common-effect”) y el de efectos aleatorios (“random-effects”)^[95]. El modelo de efectos fijos asume que hay un tamaño de efecto real que subyace a todos los estudios en el análisis, y que todas las diferencias en los efectos observados se deben al error de muestreo o error aleatorio^{[95][213]}. Este modelo se asegura de que los estudios más grandes (i.e., aquellos con la menor probabilidad de error aleatorio) tengan una mayor influencia en la estimación combinada. Sin embargo, el modelo muestra un aumento de la sobredispersión (i.e., un estimador que tiene una mayor varianza observada que la esperada teóricamente) a medida que aumenta la heterogeneidad, lo cual se interpreta como una debilidad^[213].

El modelo de efectos aleatorios permite que los verdaderos tamaños del efecto difieran; es decir, puede ser posible que todos los estudios compartan un tamaño del efecto común, pero también es posible que el tamaño del efecto varíe de un estudio a otro. El modelo supone que las diferencias en los tamaños de efecto siguen una distribución normal con una varianza común. Sin embargo, este supuesto no se justifica porque los efectos subyacentes incluidos en el meta análisis no constituyen una muestra aleatoria de la población^[213]. A pesar de ello, el modelo ignora la necesidad de aleatorización en la inferencia estadística, y la varianza de estos efectos subyacentes generalmente se aproxima mediante un estimador basado en momentos. La aplicación de esta varianza común al modelo tiene el efecto involuntario de redistribuir las ponderaciones del estudio en una sola dirección: de estudios más grandes a más pequeños.

Por lo tanto, los estudios con la menor probabilidad de error aleatorio se penalizan y no influyen tan fuertemente en las estimaciones combinadas^[213]. La inclusión de esta varianza común entre estudios estimada también parece ser un mecanismo que intenta abordar la sobredispersión con una heterogeneidad creciente, produciendo intervalos de confianza más amplios y menor significancia estadística que la que se obtendría mediante el modelo de efectos fijos^[214]. Sin embargo, debido a los supuestos erróneos, el modelo no funciona como se esperaría y, a medida que aumenta la heterogeneidad, la cobertura del intervalo de confianza cae muy por debajo del nivel nominal.

Adicionalmente, al modelo de efectos aleatorios, también conocido como el enfoque de DerSimonian y Laird, se le ha reconocido como un método subóptimo que puede generar muchos resultados estadísticamente significativos cuando el número de estudios es pequeño y además existe una heterogeneidad moderada o sustancial^[361]. Esto quiere decir que si por ejemplo, un tratamiento de ejercicio físico es ineficaz y la prueba se realiza a un nivel de significancia $p = 0.05$, la tasa de error debe ser del 5%, es decir, solo una de cada 20 pruebas debe dar un resultado estadísticamente significativo. Sin embargo, al usar el modelo de efectos aleatorios, la tasa de error puede ser mucho mayor, a menos que el número de estudios sea grande (> 20) y no haya heterogeneidad o solo sea mínima. Por esos motivos se recomienda realizar ajustes (e.g., método de Hartung-Knapp-Sidik-Jonkman)^[361] o cambiar el análisis por un modelo mucho más robusto.

Así, ya que se requiere de un mejor modelo libre de supuestos en la distribución de los tamaños de efecto, actualmente se recomienda usar el modelo de la heterogeneidad de la varianza inversa (en inglés, IVhet)^{[213][214]}. El modelo IVhet difiere del de efectos aleatorios en que las estimaciones agrupadas de IVhet favorecen los estudios más grandes; es decir, no penaliza los estudios más grandes como lo hace el modelo de efectos aleatorios. El modelo IVhet produce un intervalo de confianza más conservador con una probabilidad de cobertura correcta y exhibe una menor probabilidad observada (i.e., verdadera) de la varianza independientemente del grado de heterogeneidad^[213]. Dependiendo del meta análisis, el modelo IVhet no siempre produce una estimación puntual conservadora. Cuando se trata del intervalo de confianza, en muchos casos el intervalo de confianza del modelo de efectos aleatorios es demasiado estrecho (i.e., la sobredispersión no se aborda adecuadamente) lo cual puede producir una significancia falsa, en contraposición al modelo IVhet, que no produce significancia falsa porque la cobertura del intervalo de confianza derivado del modelo permanece en el nivel nominal^[213].

En síntesis, cuando existe heterogeneidad en el estudio, las covariables clínicas o metodológicas (i.e., variables moderadoras) son importantes para explicar las diferencias entre las magnitudes de los tamaños de efecto, y, en esa situación, el análisis de IVhet solo debe verse como un intento de mejorar la interpretación de los resultados, ya que la estimación combinada de IVhet podría ser errónea; sin embargo, ese margen de error está claramente especificado en el intervalo de confianza alrededor de la estimación de IVhet. A pesar de eso, el modelo de meta análisis IVhet es una mejora significativa sobre los modelos de efectos fijos y aleatorios para manejar la heterogeneidad y además funciona mejor que éstos^{[213][214]}.

Entonces, desde la selección del modelo se podría estar añadiendo sesgo a las conclusiones que se deriven del análisis estadístico de los datos. La selección de un modelo debe basarse únicamente en qué modelo se ajusta a la distribución de los tamaños del efecto y, por lo tanto, tiene en cuenta las fuentes de error pertinentes (Tabla 1). Cuando los estudios se obtienen de la literatura publicada, el modelo de efectos aleatorios es generalmente más plausible. Pero atención, la estrategia de comenzar con un modelo de efectos fijos y luego pasar a un modelo de efectos aleatorios si la prueba de heterogeneidad es significativa se basa en una lógica defectuosa y debe desaconsejarse enérgicamente. Si bien el modelo de efectos aleatorios es a menudo el modelo apropiado, hay casos en los que no se puede implementar correctamente porque hay muy pocos estudios para obtener una estimación precisa de la varianza entre estudios^[95].

Por lo tanto, meta analizar una cantidad reducida de estudios podría generar sesgo en las conclusiones y se recomendaría únicamente realizar una descripción cualitativa y no cuantitativa inferencial de los datos analizados. También, se ha indicado que en los meta análisis de psicología, fisiología del ejercicio y nutrición humana, los lectores no deberían esperar llegar a conclusiones definitivas, sino que se piensa que los resultados más productivos se generarán mediante comprobaciones del rendimiento de los métodos, análisis de sensibilidad^[410] y la capacidad para diseñar y realizar cuidadosos protocolos inscritos previamente^[149].

Otro aspecto que aumenta el sesgo de los meta análisis es la falta de datos que no se pueden procesar, como, por ejemplo, la información que no se reporta en el texto del artículo, en las tablas o cuadros, sino en los gráficos. Sin embargo, recientemente se han diseñado algoritmos para "aproximar" o estimar los valores graficados en las publicaciones cuando los autores originales del estudio no han atendido la solicitud de proporcionar los datos originales^[121]. Una de las formas en que se realiza esta operación es a través de programas informáticos basados en inteligencia artificial como el *WebPlotDigitizer* (<https://automeris.io/WebPlotDigitizer/>), que es gratuito y que ha demostrado una pequeña diferencia absoluta de los valores reales, excelente consistencia, satisfactoria, precisa y eficiente^[121].

Sin embargo, los datos estimados en una revisión sistemática pueden introducir sesgos y dar lugar a conclusiones incorrectas acerca de los efectos de la intervención de un estudio individual y del efecto agrupado general al influir en la significancia estadística (i.e., intervalos de confianza cambiantes) o acerca del efecto clínicamente significativo. Por lo tanto, se considera evaluar la factibilidad de utilizar estos programas informáticos con cautela, especialmente cuando los autores originales de los estudios no han sido capaces de compartir los datos brutos para realizar un análisis más apropiado.

Unido al tema de la falta de datos para meta analizar, se encuentra el posible sesgo introducido por la falta de consistencia en los hallazgos encontrados al procesar los estudios dentro del análisis estadístico. Es decir, cuando se realiza un meta análisis, su valor es especialmente claro cuando los resultados de los estudios que se incluyen muestran efectos clínicamente importantes y de una magnitud similar, pero las conclusiones son menos claras cuando los estudios incluidos tienen resultados diferentes^[339]. Este posible sesgo se puede estudiar mediante pruebas estadísticas de heterogeneidad como la *Q* de Cochran, cuya hipótesis nula establece que todos los estudios evalúan el mismo efecto. La prueba de *Q* de Cochran determina si existen diferencias genuinas subyacentes a los resultados de los estudios (i.e.,

heterogeneidad), o si la variación en los hallazgos es compatible solo con el azar (i.e., homogeneidad).

Cuando la cantidad de estudios incluidos en un meta análisis es reducida, se ha encontrado que la capacidad de la prueba de Q de Cochran es pobre; por lo que actualmente se recomienda utilizar la prueba I^2 , la cual proporciona una mejor medida de la consistencia entre los estudios incluidos en un meta análisis^[339]. La prueba I^2 describe el porcentaje de variación total entre los estudios debida a la heterogeneidad en lugar del azar o el error de muestreo^[774], ofreciendo un valor entre 0% y 100%, en donde 0% indica que no hay heterogeneidad observada, y valores más cercanos al 100% significan que sí hay una mayor heterogeneidad. Además, se ha establecido arbitrariamente, que una baja, mediana y alta heterogeneidad se describen por I^2 de $\leq 25\%$, 50% y $\geq 75\%$, respectivamente^[339].

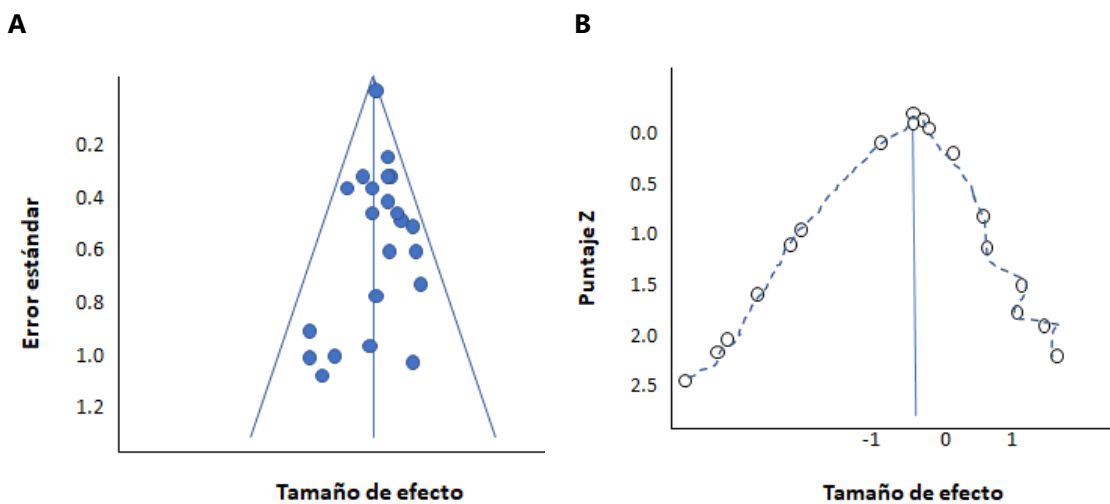
Tabla 1. Comparación de las propiedades de los modelos tradicionales de efectos fijos y de efectos aleatorios utilizados para la elaboración de meta análisis^[195]. Actualmente se recomienda utilizar el modelo de la heterogeneidad de la varianza inversa (IVhet), el cual supera ambos modelos al no depender de los supuestos en la distribución de los tamaños de efecto^{[213][214]}.

Característica	Efectos fijos	Efectos aleatorios
Supuesto	Todos los estudios del meta análisis comparten un tamaño de efecto verdadero común.	El tamaño real del efecto puede variar de un estudio a otro.
Tamaño de efecto global	Es una estimación del efecto que es común a todos los estudios del análisis.	Es una estimación de la media de una distribución de efectos reales.
Ponderaciones de los estudios	Menos uniformes o similares entre sí. A los estudios grandes se les asigna más peso relativo y a los estudios pequeños se les asigna menos peso relativo.	Más uniformes o similares entre sí. A los estudios grandes se les asigna menos peso relativo y a los estudios pequeños se les asigna más peso relativo.
Intervalo de confianza	El error estándar del tamaño de efecto global y los intervalos de confianza son menos amplios. Se podría decir que la estimación tiene mayor precisión.	El error estándar del tamaño de efecto global y los intervalos de confianza son más amplios. Se podría decir que la estimación tiene menor precisión.
Fuentes de varianza	La única fuente es el error de estimación dentro del estudio. Con un tamaño de muestra suficientemente grande, acumulado en los estudios, esta fuente de variación disminuirá y el tamaño del efecto común se estimará con precisión.	Hay dos fuentes: a) la varianza del error de estimación dentro del estudio, y b) la varianza entre los estudios. Con un tamaño de muestra suficientemente grande, acumulado en los estudios, el efecto de la primera fuente de variación disminuirá. Sin embargo, si la varianza entre estudios es sustancial, la única forma de obtener una buena precisión es aumentar el número de estudios. Si se aumenta el tamaño de la muestra en algunos estudios, es posible conocer el tamaño del efecto en esos estudios con mucha precisión, pero aún no se tiene una estimación precisa de la media en todos los estudios.

Cabe mencionar que también existen métodos gráficos para determinar la posibilidad de existencia de riesgo de sesgo. Entre estos métodos se encuentran el de “recorte y relleno” (en inglés, “trim-and-fill”)^[646]. El método de recorte y relleno tiene como objetivo estimar los estudios potencialmente faltantes debido al sesgo de publicación en el gráfico de embudo (en inglés, “funnel plot”)^[670] y ajustar la estimación del efecto general. En el gráfico de embudo, el eje X representa la magnitud del efecto, el eje Y es una medida de precisión, que puede ser el tamaño muestral o el inverso de la varianza (Figura 2A). Cada estudio representa un punto y se analiza la nube de puntos. Así, la concentración de puntos debe ser simétrica, de lo contrario, se indica que hay sesgo de publicación; es decir, los puntos que faltan en la otra mitad del gráfico serían los estudios que no se pudieron incluir en el meta análisis.

El supuesto fundamental del método de recorte y relleno es que se suprimen los estudios con los tamaños de efecto más extremos, ya sea en el lado izquierdo o en el derecho del gráfico de embudo. La dirección de los estudios que faltan depende de las expectativas de las partes interesadas en general (incluidos participantes, investigadores, tomadores de decisiones y patrocinadores) caso por caso. Como se mencionó, se supone que el gráfico de embudo es simétrico si no aparece ningún sesgo de publicación. Los estudios ausentes suprimidos por sesgo de publicación en un meta análisis generalmente conducen a un gráfico de embudo asimétrico bastante notable. A diferencia de otros métodos populares para detectar el sesgo de publicación, como por ejemplo la regresión de Egger – que además es susceptible al tamaño de la muestra –, el método de recorte y relleno no solo indica la importancia del sesgo de publicación, sino que también proporciona resultados ajustados al sesgo^[646]. Otra alternativa reciente, es utilizar el gráfico Doi (Figura 2B) y el índice LFK^[274].

Figura 2. Ejemplos con datos ficticios de: A) Funnel Plot y B) Gráfico Doi.



El gráfico de Doi disminuye la subjetividad del gráfico Funnel Plot, y aunque el gráfico de Doi se interpreta también subjetivamente, tiene la ventaja de que mejora su apariencia para que sea menos subjetiva su interpretación por medio la elaboración de centiles. Para crear el centil normal, se considera el tamaño de la muestra y el error estándar de cada estudio incluido en el meta análisis. Los percentiles se convierten a un centil normal tipificado (i.e., puntaje z) para cada estudio

y se grafican los valores. La representación se interpreta de acuerdo con la simetría mostrada, y si ésta es asimétrica, se dice que hay sesgo de publicación^[274]. Además, los valores del índice LFK van unidos al gráfico Doi y sirven para detectar asimetría. Se considera que los valores LFK fuera del intervalo -1 y 1 son considerados como asimétricos, por lo que se concluye que hay sesgo de publicación^[274].

Existen implicaciones relevantes para conducir un meta análisis después de verificar la heterogeneidad de los hallazgos. Primero, la inconsistencia de los resultados de los estudios reduce la confianza de las recomendaciones sobre el tratamiento. Segundo, se ha indicado que cuando existe una alta heterogeneidad, se puede recurrir al análisis de subgrupos; es decir, subdividir los estudios en categorías lógicas (e.g., estudios de ejercicio aeróbico, estudios de ejercicio contra resistencia). Sin embargo, se advierte que es posible que se pierda potencia y que si se encuentra una falta de heterogeneidad al hacer el análisis de subgrupos se deba a la pérdida de estudios (i.e., reducido n) y no necesariamente a la homogeneidad de los estudios^[339]. Tercero, se ha sugerido que, en presencia de una alta heterogeneidad, no se debería proseguir con el meta análisis^[497], sino alternativamente realizar una descripción cualitativa, no inferencial, de los resultados. Finalmente, existen herramientas para evaluar la confianza y solidez de los hallazgos de un meta análisis, lo que a su vez puede colaborar en reducir el sesgo; por ejemplo, la herramienta “Grading of Recommendations Assessment, Development and Evaluation” (GRADE)^[643]. Esta es una herramienta subjetiva dividida en cinco áreas que evalúa la fuerza de la evidencia encontrada para un resultado específico^[305]: a) Limitaciones del estudio, b) Imprecisión, c) Inconsistencia, d) Evidencia indirecta, y e) Sesgo de publicación. La solidez general de los resultados se clasifica de acuerdo a la figura 3^[305].

Figura 3. Clasificación de la solidez de los hallazgos al utilizar la escala GRADE.



Sin embargo, se debe considerar que GRADE define la certeza de la evidencia como la certeza de que un efecto verdadero se encuentra en un lado de un umbral específico o dentro de un rango elegido. Por ello, al utilizar GRADE, es deseable especificar los umbrales y aclarar el objetivo de certeza de la calificación de la evidencia. GRADE proporciona tres niveles de contextualización, enfoque mínimo, parcial y totalmente contextualizado, y posibles umbrales bajo cada nivel de contextualización para facilitar las calificaciones de certeza^[772].

Un ejemplo de la utilización de GRADE se muestra en estudio de Pinheiro et al.^[570]. El propósito del estudio fue investigar la prevalencia, incidencia y perfil de lesiones musculoesqueléticas en atletas. Durante el meta análisis se evaluó la calidad general de la evidencia, la cual fue baja; es decir, usando los criterios GRADE se observó que la calidad disminuyó debido a la imprecisión y la falta de direccionalidad de los hallazgos. Por ejemplo, el análisis de subgrupos del meta análisis mostró una tasa de incidencia de lesiones significativamente menor en los estudios de mayor calidad metodológica en comparación con los estudios de menor calidad metodológica. El análisis de subgrupos no mostró una diferencia clara con respecto al tamaño de la muestra, ya que mientras los estudios con un tamaño de muestra grande estimaron una tasa de incidencia de lesiones de 14.4 por cada 1000 días-atleta, los estudios con un tamaño de muestra pequeño mostraron una tasa de incidencia de 14.7 por cada 1000 días-atleta^[570]. Este tipo de información complementa las estadísticas objetivamente calculadas y permite llegar a conclusiones más integrales relacionadas con los datos y la evidencia analizada.

En síntesis, debido al dramático aumento en la cantidad de revisiones de literatura con y sin meta análisis que se publican actualmente en diversas áreas de las Ciencias del Movimiento Humano, se hace cada vez más necesario poder evaluar la calidad de la evidencia, de manera que las conclusiones no sean sesgadas. Existen indicadores que permiten conocer el posible sesgo y se estimula a los investigadores a utilizarlos en el reporte de sus trabajos. Esto permitirá llegar a conclusiones sólidas con respecto al artículo publicado y mejorará la calidad y rigurosidad metodológica de las investigaciones futuras.

3.1. La variabilidad en la investigación

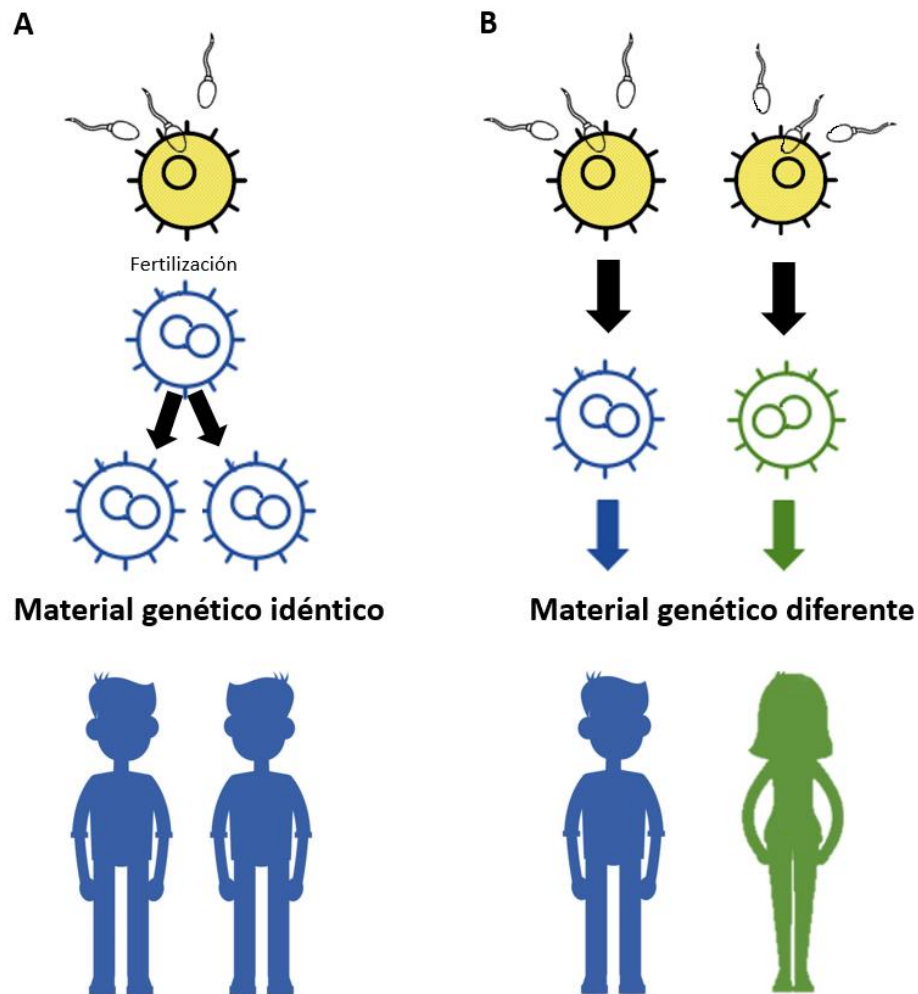
Uno de los objetivos de la ciencia es generar conocimiento novedoso y generalizable para beneficio de la población. Cuando se diseñan investigaciones y se realizan análisis de datos, los hallazgos frecuentemente se generalizan a cierto grupo de personas. Por ejemplo, los efectos beneficiosos de un programa de entrenamiento para personas con cáncer de mama, solamente se pueden generalizar para personas con esa condición; es decir, el entrenamiento no es aplicable para personas con otra patología hasta que no se investigue con ese grupo específico de personas. Sin embargo, en un estudio en el que solamente exista un grupo al que se le aplica un programa de ejercicio, solamente puede generar inferencias individuales porque no se comparó con un grupo control que no recibiera el programa de ejercicio. Por lo tanto, no se puede llegar a conocer el verdadero efecto del ejercicio porque únicamente se tendría la desviación estándar de los cambios individuales observados debidos al ejercicio^[89].

Consecuentemente, las recomendaciones de entrenamiento o de actividad física generalmente se diseñan para la mayoría de las personas; es decir, se asume o se supone que el entrenamiento físico tiene un efecto beneficioso similar para toda la población. Esas recomendaciones generalmente se basan en la efectividad promedio de las variables de interés, lo cual es positivo porque se quiere que la sociedad, como un todo, pueda alcanzar una mejor calidad de vida. Sin embargo, ese enfoque generalista invisibiliza los efectos individuales, que son heterogéneos^[6]; es decir, “no reconoce que existen considerables diferencias interindividuales en las respuestas a cualquier programa de ejercicio” (p. 2).^[103]

3.1.1. Estudios en gemelos

En el área del entrenamiento físico en las Ciencias del Movimiento Humano, desde inicios de los años 80, el Dr. Claude Bouchard, en su laboratorio en la Universidad Laval en Canadá, fue uno de los pioneros en estudiar las razones que explican la variabilidad humana en respuesta al entrenamiento físico^{[100][101]}. Sus primeras observaciones fueron precursoras para hipotetizar que algún componente genético podría explicar dichas variaciones; y a partir de entonces, se diseñaron estudios en seres con el mismo material genético (i.e., gemelos); tanto en modelos animales como en humanos. La biología básica indica que los gemelos idénticos o monocigotos (MZ), a diferencia de los dicigóticos (DZ), son el resultado de la fecundación de un único óvulo y un espermatozoide. Se sabe que esas células se dividen y se multiplican, y en algún momento durante las primeras fases de su desarrollo se escinden en dos, dando lugar a dos individuos que comparten el mismo material genético (Figura 1).

Figura 1. Esquema de la fertilización cuyo producto son los gemelos idénticos o monocigotos, MZ (Panel A) y los gemelos fraternos o dicigóticos, DZ (Panel B).



El fundamento de las hipótesis de los investigadores era que, en un diseño experimental, se podía contar con que uno de los gemelos realizara el entrenamiento físico (i.e., experimental) y el otro gemelo serviría como sujeto control, es decir, no haría el entrenamiento físico. Así, se pensaba que los cambios observados en uno de los gemelos se debían exclusivamente al entrenamiento físico, y que cuando se replicara el entrenamiento en el gemelo control que no había realizado ejercicio (i.e., en un diseño cruzado), se obtendrían los mismos resultados. Sin embargo, esos estudios iniciales, no consideraban la influencia del ambiente; es decir, la epigenética, y hoy se sabe cada vez más que las influencias ambientales cambian las características de los gemelos idénticos ^[475].

Desde mediados de los años 80, se comenzaron a realizar experimentos con pares de gemelos MZ y se encontró que las diferencias individuales en la capacidad de respuestas y adaptaciones al entrenamiento no estaban distribuidas al azar ^[582]. Eso significa que hubo una consistente y mayor variación en las respuestas de entrenamiento entre gemelos MZ que la observada entre hermanos o hermanas que no eran gemelos. Esos experimentos de entrenamiento con gemelos MZ

mostraron un fuerte efecto de la interacción entre el genotipo y el entrenamiento que contribuía a la variabilidad en las características de interés, generalmente relacionadas con la capacidad aeróbica máxima ($VO_2\text{máx}$)^[100]. Posteriormente, Bouchard estableció una línea de investigación en conjunto con varias universidades de EE.UU., y conformó el estudio familiar llamado, por sus siglas en inglés "HERITAGE" (**HE**alth, **R**isk factors, exercise **T**raining **A**nd **GE**netics)^[105], en donde se estudiaron los efectos del entrenamiento y la heredabilidad de los cambios encontrados, principalmente aquellos relacionados con la capacidad aeróbica. En esos estudios se encontró que, por ejemplo, para el efecto del entrenamiento en el $VO_2\text{máx}$, para personas con una mayor variación genómica compartida; es decir, con una mayor semejanza genética, se esperaba que existiera una respuesta parecida al entrenamiento, y se encontró que después de un programa de entrenamiento de 20 semanas la heredabilidad alcanzaba el 40%; y en otras variables la respuesta al entrenamiento mostró una alta heredabilidad (~35%)^[102]. Hoy en día la heredabilidad de ciertas características se puede estudiar con esos modelos de gemelos MZ; por ejemplo, en un estudio realizado en 88414 gemelos en Suecia, se encontró que la contribución genética a la ruptura del ligamento cruzado anterior tenía una heredabilidad aproximada de ~69%^[455].

A partir de entonces, y posteriormente con la secuenciación del genoma humano, se han realizado múltiples estudios para determinar genes, alelos, proteínas y sitios (i.e., loci) específicos que puedan servir para apoyar las explicaciones acerca de la influencia genética (i.e., genes candidatos) y ambiental en las respuestas y adaptaciones de los distintos tejidos y sistemas humanos al entrenamiento físico^{[2][632]}. Por ejemplo, consistentemente se ha encontrado que la proteína de unión a actina del músculo esquelético llamada α -actinina-3 (ACTN3) está relacionada con el rendimiento deportivo de élite, y sus diferentes tipos (i.e., polimorfismos) se asocian a la capacidad aeróbica o a la capacidad anaeróbica^[541]. Lo anterior se confirma con un estudio reciente^[408], en el que también se encontró que corredores de larga distancia y ciclistas de ruta tienen una mayor probabilidad de portar variantes genéticas específicas en comparación con controles, especialmente para los polimorfismos ACE II (I/D), ACTN3 TT (rs1815739), PPARGC1A GG (rs8192678), AMPD1 CC (rs17602729) y HFE GG + CG (rs1799945). Sin lugar a dudas, en los próximos años la detección de talentos deportivos incluirá un perfil genético que pueda ser potencializado a través de un ambiente adecuado que ofrezca condiciones de salud y desarrollo positivos (i.e., epigenética).

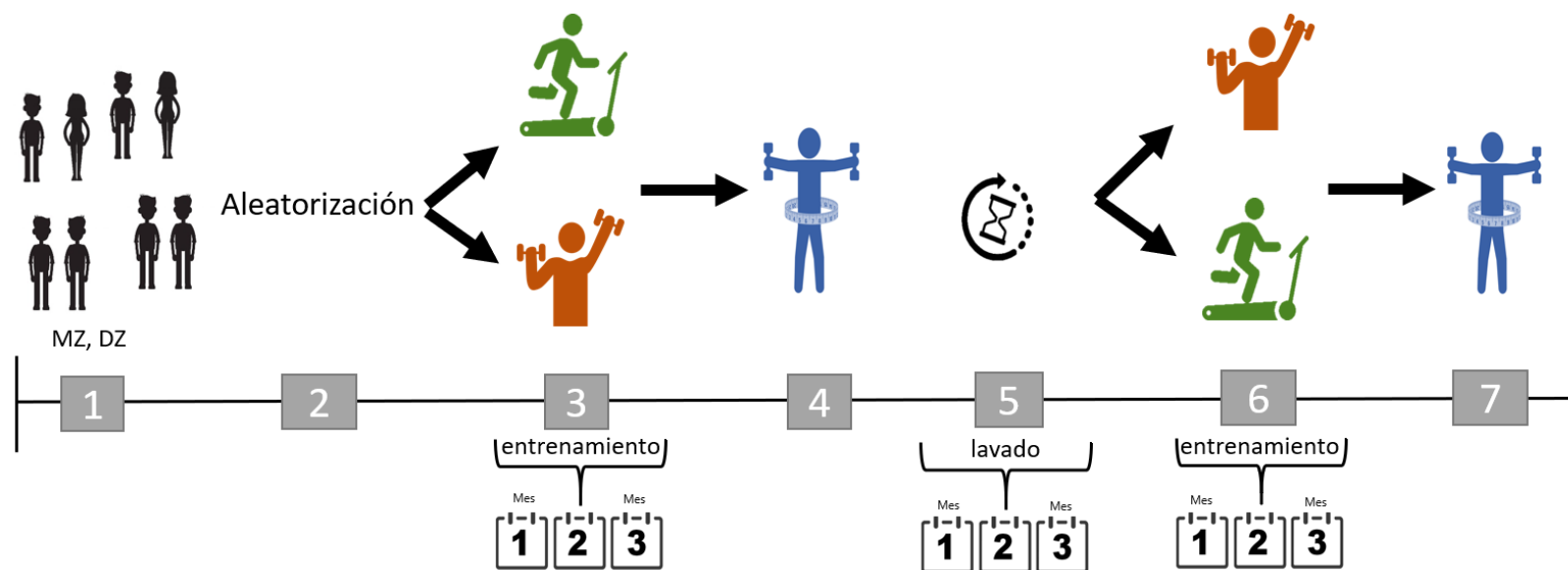
En el área de la salud, se realizan estudios de aleatorización mendeliana en los que se determinan, por ejemplo, los beneficios de la actividad física en la composición corporal^[590]. La aleatorización mendeliana se define como un tipo de análisis de variables que aprovecha las variantes genéticas de los estudios de asociación del genoma completo para inferir la causalidad entre una exposición (e.g., ejercicio físico) y un resultado determinado (e.g., cambios en la composición corporal). La asignación de variantes genéticas durante la formación de gametos concede una exposición de por vida para individuos aleatorizados a una variante alélica determinada. Es por ello, que en comparación a las exposiciones medidas a nivel de fenotipo, las variantes genéticas no están asociadas con una cantidad de factores conductuales, sociales o fisiológicos que pueden confundir una asociación epidemiológica. Por lo tanto, la aleatorización mendeliana evita muchos de los errores de los estudios observacionales (i.e., sesgo, causalidad inversa).

Para quienes deseen realizar una revisión histórica más exhaustiva, se recomienda revisar el trabajo pionero de Rankinen et al.^[596] acerca del mapa de genes humanos para fenotipos de aptitud física relacionados con el rendimiento y la salud humana y sus subsecuentes actualizaciones^{[107][557][558][595][597][598][619][633][758][759]}. Adicionalmente, se recomienda revisar los avances científicos más recientes que sintetizan posibles genes relacionados con la respuesta del músculo al entrenamiento físico, encontrados a partir de meta análisis realizados con métodos computacionales de vanguardia^{[18][569]}.

La alimentación y el ejercicio físico se consideran como conductas que forman parte del ambiente en el que viven las personas; y esos factores epigenéticos regulan la expresión genética de forma específica (e.g., músculo esquelético, músculo cardíaco), por lo que constituyen el vínculo entre el genotipo y el medio ambiente^[370]. Por ejemplo, en un estudio con gemelos diseñado para determinar los factores motivacionales que influyen en la cantidad de actividad física que realizaban, se encontró que entre el 5% y el 42% de la variación en los motivos fue explicado por factores genéticos y entre el 58% y el 95% por factores ambientales^[2]. La propia modalidad de entrenamiento es un factor epigenético que ha explicado la adaptabilidad, respuesta o variabilidad entre gemelos^[466]. En dos estudios de diseño cruzado y aleatorizado^{[466][692]}, 84 pares de gemelos MZ (n = 30) y DZ (n = 12) entrenaron aeróbicamente (AE) y contra resistencia (CR) durante tres meses (Figura 2). El programa de ejercicio estuvo separado por un período de lavado (“wash-out”) o de desentrenamiento de tres meses. Posteriormente, se evaluaron las respuestas al entrenamiento en variables de capacidad aeróbica y fuerza^[466], así como de composición corporal^[692]. Se encontró que la capacidad de respuesta individual difiere entre las modalidades de ejercicio; incluso, los participantes que no respondieron a una modalidad respondieron a la otra modalidad y las respuestas de los gemelos fueron mejores para los MZ que para los DZ en las variables de composición corporal (i.e., mejor perfil de composición corporal). Las estimaciones de heredabilidad transversales fueron mayores que las estimaciones de respuesta al entrenamiento, y el entorno compartido tuvo la mayor influencia en los cambios en la composición corporal^[692].

De esta forma, se concluye que los genes no desempeñan un papel tan relevante en la adaptación al entrenamiento como anteriormente se había pensado. Y en conjunto, estos hallazgos indican que los factores ambientales (i.e., epigenética) contribuyen en gran medida con los motivos que tienen las personas para mantenerse físicamente activas (i.e., cogniciones) y con los cambios observados en variables de la salud y el rendimiento físico (i.e., VO₂máx, fuerza, composición corporal), lo cual les permite a los profesionales de las Ciencias del Movimiento Humano diseñar intervenciones más personalizadas tomando en consideración, además, el sexo y otras variables moderadoras de esos efectos^[2].

Figura 2. Esquema de un diseño de estudio aleatorizado y cruzado con gemelos monocigóticos (MZ) y dicigóticos (DZ). Los pares de gemelos (1) son aleatorizados (2) a una modalidad de entrenamiento de tres meses (3) y luego son medidos en las variables de interés (4). Posterior a un período de lavado o desentrenamiento de tres meses (5), completan la otra modalidad de entrenamiento durante tres meses (6) y luego son medidos en las variables de interés (7). Entre el paso 5 y 6 ya no hay aleatorización, cada grupo de parejas completa el entrenamiento que no había realizado anteriormente^{[466][692]}.



3.1.2. ¿Por qué se estudia la variabilidad?

La forma más reconocida y aceptada por la comunidad científica para reportar la estadística descriptiva de una variable cuantitativa es la medida de tendencia central llamada promedio o media aritmética. Sin embargo, el promedio no muestra la variación en las respuestas individuales de las personas al entrenamiento^[460]. Así, cuando se diseña y ejecuta un protocolo de ejercicio o de entrenamiento físico, no todas las personas responden de la misma forma, ya sea en magnitud o en dirección. Es decir, algunos participantes responden en la dirección esperada (i.e., mejoran), otros no responden (i.e., se quedan igual, como si nada hubiera ocurrido) y otros responden en la dirección no deseada (i.e., empeoran). También, la magnitud de la respuesta puede ser gigantesca para unas personas, pero para otras, apenas podría ser un cambio casi imperceptible. Estas respuestas individuales producen lo que se conoce como variabilidad o varianza, y a los investigadores de las Ciencias del Movimiento Humano les corresponde reconocer, identificar y explicar por qué ocurre esa variabilidad en la respuesta a una intervención.

Así, en un experimento en el cual se manipula una variable independiente; por ejemplo, tipos de entrenamiento, se espera que exista variabilidad en la respuesta de los participantes en una variable de interés; por ejemplo, en la distancia recorrida en una carrera. De esta forma, se esperaría encontrar una variación en las distancias en la carrera que se explican por el tipo de entrenamiento, conocida como varianza verdadera^[508], así como también una variabilidad aleatoria, también llamada variación residual, variación inexplicable o varianza no explicada^[368]. Esta última expresión se usa porque cuando se hace la prueba, parte de la variación en la distancia de la carrera se atribuye al factor "tipo de entrenamiento". El resto de la variación no se atribuye al factor "tipo de entrenamiento", por lo que esta variación residual se denomina "inexplicable" o aleatoria^{[229][368]}.

La variación debido a cambios normales que presenta una persona de un día a otro (i.e., fluctuación diaria, variabilidad dentro del mismo sujeto, en inglés "within-subject"), a la variabilidad en las respuestas entre diferentes personas (i.e., variabilidad entre sujetos, en inglés "between-subjects") y a la variación introducida por el error del instrumento utilizando durante la medición de las variables^[618]. La variabilidad aleatoria dentro del mismo sujeto hace que parezca que las personas que fueron sometidas a una intervención con ejercicio hayan respondido de manera diferente a esa intervención, pero en realidad hay diferencias dentro del mismo sujeto que no son importantes en la verdadera respuesta a una intervención^[39].

El desafío estadístico que se presenta al evaluar la respuesta individual es que, a diferencia de los promedios grupales, el error aleatorio en las respuestas individuales observadas no disminuye con el aumento en el tamaño del grupo^[618]. De esta forma, la varianza de la respuesta observada, que algunas veces se le llama "variabilidad de la respuesta bruta", siempre sobrestimaré la varianza verdadera de la respuesta interindividual porque incluye la varianza debida al error aleatorio, que a su vez puede explicar la mayor parte o la totalidad de la variabilidad interindividual observada. La respuesta individual a un programa de ejercicio físico es el cambio sin errores que ocurren en una persona y que es causado por la intervención del ejercicio; por lo que el cambio observado es la suma del cambio causado por la intervención, más el cambio que habría ocurrido en ausencia de la intervención (e.g., control), así como las diversas fuentes de error (e.g., el error de medición [que además se acumula de pre-test a post-test], variabilidad biológica diaria, entre otras).

Actualmente, existe una creciente conciencia acerca del estudio de la variabilidad de la respuesta al entrenamiento para poder identificar los predictores de las adaptaciones inducidas por el ejercicio que sirvan para realizar una prescripción mucho más personalizada^[368]. En la fisiología del ejercicio, así como en otras ramas de la medicina, se ha recurrido a biomarcadores moleculares de la respuesta adaptativa al ejercicio. A pesar de que estos biomarcadores moleculares podrían esclarecer los pasos regulatorios claves que contribuyen a las adaptaciones crónicas o a largo plazo inducidas por el ejercicio, también hay que considerar la influencia confusora del error aleatorio o ruido atribuible a fuentes técnicas o biológicas, las cuales pueden limitar la capacidad para dilucidar con precisión los predictores de la respuesta individual^[368].

En congruencia con el dogma central de la biología molecular (e.g., ADN → ARN → proteína), la adaptación al ejercicio está ampliamente aceptada como iniciada a nivel transcripcional. De esta forma, se explica que los aumentos agudos en los ARNm musculares que codifican proteínas transcripcionales y metabólicas preceden a los aumentos crónicos en la expresión de estas proteínas. Consecuentemente, la fisiología del ejercicio ha aceptado el paradigma de que aumentos mayores en la expresión de ARNm después del ejercicio agudo indican una mayor activación de las vías moleculares que sustentan la adaptación biológica o el efecto crónico a largo plazo^[556]. Esta afirmación teóricamente ayudaría a comprender cómo se puede reducir la variabilidad en la respuesta de las personas al ejercicio físico; sin embargo, los estudios existentes no han logrado establecer relaciones directas entre la expresión genética aguda y los cambios fenotípicos crónicos mediados por el entrenamiento.

De hecho, se ha encontrado que los cambios en la expresión del ARNm muscular no son repetibles; es decir, las personas entrenadas parecen responder de manera diferente al mismo estímulo^[367]. De esta forma, la variabilidad intraindividual en la expresión del ARNm no podría explicarse por un error técnico derivado de los principales pasos analíticos involucrados en el análisis de genes, que apuntan a fuentes de error aleatorio que se originan dentro del propio músculo (e.g, diferencias morfológicas entre muestras, cambios aleatorios en la expresión génica), fluctuaciones diurnas en el metabolismo o la transcripción, entre otras. En síntesis, las respuestas al ejercicio difieren en condiciones experimentales idénticas, lo que cuestiona la utilidad del ARNm como un biomarcador del potencial adaptativo para la prescripción personalizada del ejercicio^[368].

Aparte de la obvia variabilidad genética (i.e., biológica) y epigenética (i.e., ambiental) que existe entre cada persona y que se ha encontrado que puede explicar hasta el 50% de la varianza en variables físicas, cognitivas y en la estructura cerebral^[335], también se han identificado algunos elementos que intervienen en la variabilidad de la respuesta a una intervención, especialmente cuando se comparan diferentes estudios. Por ejemplo, se sabe que la validez, la confiabilidad y la sensibilidad de la prueba afectan la variabilidad en las respuestas de las personas que participan en los estudios de entrenamiento^[190]. Existen pruebas para determinar el rendimiento físico que varían según el protocolo utilizado; por ejemplo, aquellas en las que se busca llevar hasta el cansancio físico a la persona para posteriormente medir los tiempos alcanzados, hasta protocolos que miden un tiempo constante o protocolos contra reloj. Así, tanto la modalidad de la prueba como la duración del protocolo van a ser considerados como posibles fuentes de variabilidad.

También existen otros factores o niveles^[229] que pueden afectar la respuesta del individuo, como, por ejemplo, la familiarización con la prueba, si hay o no hay refuerzo o apoyo verbal, si el protocolo utiliza música, si hay algún tipo de retroalimentación para el participante, o incluso la

cantidad de mediciones de la variable dependiente que le realicen a la persona^[190]. Por lo tanto, para efectuar un análisis de la varianza (ANOVA), se debe comprender que hay que analizar los factores o categorías por separado y su interrelación o interacción^{[229][508]}.

La confiabilidad es una medida importante ya que brinda una indicación de la variación biológica y técnica de los protocolos, y permite estudiar la variabilidad. También se utilizan otros indicadores de la confiabilidad, como por ejemplo el coeficiente de variación (CV), que expresa la desviación estándar (DE) de la medida como un porcentaje de la media (M) o promedio, lo que facilita la comparación de la cantidad de variación entre diferentes datos:

CV (%) = (M/DE) x 100	El CV representa la magnitud de las diferencias diarias cuando se obtienen múltiples mediciones de los participantes ^[190] . Es útil para comparar métodos distintos de entrenamiento; sin embargo, solamente explica el 68% de la variabilidad (porque representa la variabilidad del día a día) ^[618] , por lo que se debería utilizar junto con otros indicadores de variabilidad para tomar decisiones más acertadas.
------------------------------	---

Por ejemplo, en la tabla 1, se muestra el CV (%) para dos grupos de datos ficticios, en los que se presentan los puntajes de una variable llamada cambio en el estado de ánimo de un grupo control que no realizó ejercicio y un grupo experimental en que los participantes realizaron ejercicio físico.

Tabla 1. Valores de cambio en el estado de ánimo de participantes en un experimento ficticio. Los valores representan un % de cambio.

Participantes	Δ_{Control}	Participantes	$\Delta_{\text{Experimental}}$
Ana	-3.5	Paula	-53.7
Beatriz	-60.4	Rosa	-23.9
Carolina	-11.7	Roxana	12.4
Paulina	25.3	Zaida	-7.2
Melinna	65.3	Yamileth	26.5
Isaura	-43.4	Fiorella	123.3
Jessenia	-44.2	Valeria	66.3
Andrea	-5.8	Gabriela	143.7
Martha	-6.7	Cielo	700.0
Alvaro	-18.2	Rodrigo	109.4
Bruno	29.4	Pedro	43.7
Iván	27.7	Alejandro	93.9
Javier	185.6	Gerardo	104.6
Eduardo	64.6	Luis	279.0
Franco	64.1	Juan	107.5
Guillermo	51.6	José	177.3
M=	20.0	M=	118.9
DE=	59.8	DE=	175.6
CV (%)	33.4	CV (%)	67.7
p=	0.041		

M=Media aritmética o promedio; DE = Desviación estándar; p = significancia estadística; CV (%) = coeficiente de variación.

Como se puede observar claramente en la tabla 1, existe una diferencia estadísticamente significativa entre los promedios, denotado por un valor de significancia $p = 0.041$. Esto representa una diferencia en la varianza entre grupos o "between-groups". También se puede observar que esa variabilidad (DE) es mucho mayor en el grupo de puntajes del grupo experimental. Este mismo comportamiento se observa en el CV, lo que indica que existe una mayor variación en los puntajes que conforman los participantes del grupo experimental. Eso significa que en los participantes del grupo experimental hay un indicio de que no todos los participantes responden de la misma manera al ejercicio físico a pesar de que como grupo, sí son diferentes al compararse con los participantes del grupo control que no realizaron ejercicio. Este hallazgo se desarrollará más adelante en el apartado de variabilidad dentro de los mismos sujetos, "within-subject" o "intrasujeto".

Otra medida de confiabilidad utilizada es la correlación de Pearson (r) o correlación interclase, en donde una correlación significativa alta puede llevar a la conclusión de que un protocolo o grupo de datos es confiable y consistente en el tiempo. Aunque r puede brindar una significancia estadística y así se puede conocer una probabilidad, tiene la desventaja de que no puede detectar cambios en la media influenciados por la variabilidad dentro del mismo sujeto^[190]. El coeficiente de correlación intraclase (CCI) o coeficiente intraclase, es otra forma de correlación que puede usarse para evaluar la confiabilidad, siendo considerado como bajo o cuestionable un valor entre 0.70 a 0.80 y alto si > 0.90 ^[190].

Dependiendo de la literatura que consulte, esos valores pueden ser interpretados de forma distinta; por ejemplo, el CCI puede ser interpretado como bajo si < 0.50 , moderados si están entre 0.50 y 0.70, buenos si se ubican entre 0.75 y 0.90, y excelentes si se encuentran > 0.90 ^[409]. Por tal motivo, es importante contextualizar el tipo de estudio (e.g., salud, rendimiento deportivo, entrenamiento físico, cognitivo, educativo) y buscar la interpretación cuantitativa y cualitativa que mejor se adapte para emitir un juicio de valor respecto a la variabilidad encontrada. Cabe anotar, que el CCI es sensible al sesgo sistemático y a la heterogeneidad o diversidad de la muestra, por lo que se debe usar con cautela solamente si el diseño del estudio lo justifica.

También se pueden utilizar los límites de acuerdo al 95% ("Limits of Agreement", LoA95%), donde la variación se expresa en términos absolutos en forma de intervalo, y se pueden considerar como un error o intervalo de tolerancia que permiten la comparación subjetiva de los límites de concordancia y el valor medio de la medida. Sin embargo, los LOA95% se ven seriamente afectados cuando el error de medición aumenta en conjunto con el puntaje obtenido en la prueba, fenómeno conocido como heterocedasticidad^[190].

Se ha indicado que estadísticos como el R^2 múltiple o el valor absoluto de la diferencia entre los promedios de los grupos (i.e., experimental vs. control) presentan un sesgo hacia arriba (i.e., mayores valores, sobrestimación), aunque ese sesgo disminuye con un aumento en el tamaño de la muestra. Hay que considerar que generalmente las investigaciones diseñadas para comparar los promedios de los grupos generalmente no tienen la potencia estadística suficiente para identificar variables modificadoras o moderadoras de los efectos.

Por ejemplo, para una variable modificadora binaria (e.g., sexo), la cual en el mejor de los casos está equilibrada dentro de los grupos (i.e., igual cantidad de hombres y mujeres en cada grupo de intervención y control), se sabe que el tamaño de muestra total requerido para detectar una

diferencia en la respuesta al tratamiento de δ (i.e., $\Delta\delta = M_{\text{Experimental}} - M_{\text{Control}}$, mayor para las mujeres que para los hombres) requeriría cuatro veces el tamaño total de la muestra en comparación con un estudio diseñado para detectar una diferencia general entre el control y el experimental^[413]. La acumulación de medidas como R^2 sobre múltiples variables puede resultar en una sobrestimación sustancial de la cantidad de variabilidad de la respuesta interindividual atribuida a estas variables, por lo que el uso de medidas de corrección de sesgo, como R^2_{ajustada} , puede reducir, pero no eliminar el problema^[618].

Algunas estrategias recomendadas para reducir la variabilidad en la respuesta al ejercicio incluyen estandarizar el ejercicio (i.e., duración, intensidad, gasto calórico, tiempo de sesión, orden de ejercicios), registrar o llevar un control estricto de la adherencia al programa de entrenamiento (i.e., para conocer con precisión la cantidad de tiempo que los participantes estuvieron expuestos al tratamiento), asignar los tratamientos de forma apareada (o pareada) a los participantes con base en alguna característica que los identifique (e.g., niveles de condición física, edad, escolaridad, tipo de enfermedad o condición de salud), e identificar a aquellas personas que responden y a las que no responden a la intervención (en inglés, "responders" vs. "non-responders")^{[336][618]}; es decir, aquellas personas que sobrepasan el umbral de respuesta en la variable dependiente de interés (e.g., $VO_{2\text{máx}}$, fuerza, velocidad, ansiedad, fatiga) y aquellas que no lo sobrepasan o que por el contrario, muestran un respuesta en la dirección opuesta a la esperada.

También se han explicado las fortalezas y las debilidades de los diseños de investigación que se han utilizado para estudiar las respuestas al entrenamiento, en donde los diseños de medidas repetidas con múltiples mediciones de la variable dependiente tienden a ser más robustos para explicar la variabilidad porque la varianza total se reduce debido a que no hay varianza intergrupala entre un grupo de control y un grupo experimental, y al reducirse además el error estadístico, se dice que estos diseños suelen ser más económicos que los diseños de grupos de control^[618]. Para una mayor profundización en el tema, se recomienda consultar los excelentes trabajos de Hecksteden et al.^{[330][331]}

A pesar de que se pueden controlar posibles fuentes de varianza extraña que impidan registrar una señal clara o una varianza verdadera, existe el dilema de si el control excesivo puede reducir la validez externa o generalización de las conclusiones^[387]. Es decir, a pesar de que se realicen intentos para reducir las fuentes de variabilidad no deseada (e.g., teniendo participantes muy homogéneos en la variable de interés), cuando los hallazgos de las investigaciones se generalizan a la población, es muy posible que la intervención no funcione para la gran mayoría de individuos; entonces, "¿Por qué esperamos que un efecto observado en esta situación estandarizada, con una variación artificialmente baja, se traduzca en un efecto significativo una vez que esté presente la variación biológica normal?" (p. 2).^[387]

Para tratar de responder esa pregunta, se ha propuesto un cambio de paradigma en la forma en que se plantean los experimentos para así mejorar la reproducibilidad. Se acepta como un hecho que existirá la variación, pero no al azar, sino por el contrario, se recomienda la heterogeneización sistemática mediante la utilización de diseños de bloques aleatorios. Con un diseño de bloque, las condiciones basales (i.e., de descanso, alimentación, actividad física habitual) se estandarizan dentro de un bloque; sin embargo, las condiciones entre bloques pueden variar (i.e., prescripciones de ejercicio o entrenamiento). Con este tipo de diseño, se puede estimar el efecto promedio en los bloques, pero también explorar la consistencia de esta respuesta. Esta es

sin lugar a duda, una nueva perspectiva metodológica que vale la pena explorar en investigaciones en diversos temas de las Ciencias del Movimiento Humano.

3.1.3. Variabilidad entre diferentes individuos

Como se describió anteriormente, en la tabla 1, se muestran los cambios (Δ) ocurridos de la medición inicial a la medición final observados en un estudio ficticio en el que se comparó un grupo control (i.e., que no realizó ejercicio) y un grupo experimental (i.e., que realizó ejercicio). La diferencia entre los grupos (i.e., "between-subjects", "between-groups", inter-sujetos) se observa fácilmente por la significancia estadística (p). Esto quiere decir, que la variación, variabilidad o varianza es diferente entre los puntajes grupales de quienes no realizaron ejercicio y quienes sí lo hicieron. Esa es una conclusión basada en la evidencia, apoyada con la probabilidad estadística; sin embargo, no es una conclusión absoluta ni completa para cada persona que participó en el estudio.

La variabilidad significa que algunas personas respondieron de una forma, mientras que otras lo hicieron de otra manera. Por lo tanto, hasta el momento, el análisis tradicional de las diferencias grupales representa apenas un enfoque incompleto del análisis de los datos, y consecuentemente, solamente permite conocer una parte de la verdad que busca la ciencia. Es por ello, que se debe complementar el análisis considerando los puntajes y la variabilidad de cada persona para poder llegar a conclusiones más robustas. Esos análisis se describirán en la sección de "Variabilidad dentro del mismo individuo".

Para conocer la verdadera respuesta a la intervención (i.e., el efecto del entrenamiento físico) de los participantes, primero se debe "eliminar" la influencia del error técnico de la medición (ETM) o en inglés, el "technical error of measurement". Esto permite conocer las respuestas entre individuos (i.e., inter-individuos); es decir, conocer quienes responden a la intervención y quienes no lo hacen. Para ello, se deben cuantificar los dos componentes del cambio: a) aleatorio; es decir, el cambio producido por el error técnico y la variabilidad biológica de las personas; y b) sistemático; es decir, el cambio provocado por la intervención^[660].

Así, primero se calcula la variación de la persona de una medición a la otra, lo cual se conoce como la desviación estándar intra-sujeto (i.e., variabilidad dentro de la misma persona) o ETM. La fórmula es conceptualmente sencilla. ETM es la desviación estándar de las diferencias entre los puntajes medidos (i.e., post – pre) dividida entre la raíz cuadrada de 2. El ETM conserva las unidades de medición originales (e.g., m, kg, cm, s, mol/l, ml·kg⁻¹·min⁻¹). Debido a que en una distribución normal el 95% de los puntajes se aglomeran entre 1.96 desviaciones estándar de la media, se puede entonces conocer cuáles personas obtuvieron puntajes de diferencia superiores o inferiores a un valor "umbral" o puntaje de corte calculado a partir de 1.96 x ETM.

De esta manera, se podría clasificar a cada persona como respondedora o como no respondedora a la intervención (i.e., programa de entrenamiento) si ésta alcanza o no ese umbral calculado (Tabla 2).

Tabla 2. Valores iniciales y finales de fuerza muscular (kg) de un grupo de participantes en un estudio ficticio.

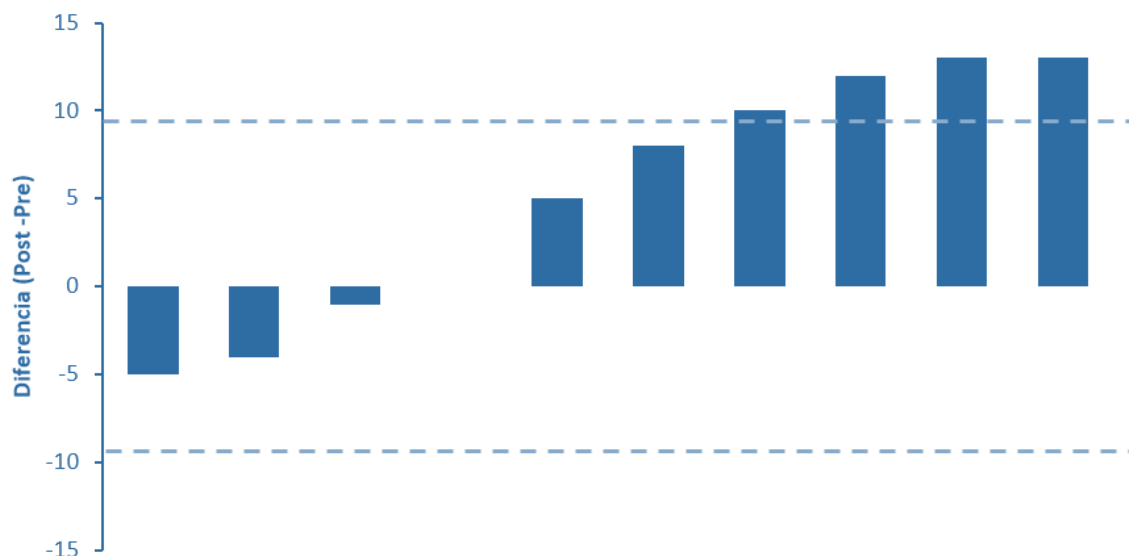
Participante	Pre-Test	Post-Test	Diferencia (Post-Pre)	¿Responde al Ejercicio?
Kyle	15	10	-5	No
Peter	15	11	-4	No
Esteban	10	9	-1	No
Samaria	12	12	0	No
Andy	10	15	5	No
Leslie	3	11	8	No
Roberto	6	16	10	Si
Alberto	4	16	12	Si
Alma	5	18	13	Si
Nicko	6	19	13	SI
M₁ =	8.6	13.7		
M_{global} =	11.15			
p =	0.049			
		DE_{dif} =	7.1	Fórmulas:
		ETM =	5.0	← DE_{dif} / √2
		“Umbral” =	± 9.8	← 1.96 x ETM

M₁=Media aritmética o promedio de la medición pre-test y post-test; M_{global}=Media aritmética o promedio de las mediciones pre- y post-test combinadas; DE_{dif} = Desviación estándar de las diferencias; p = significancia estadística; ETM = Error técnico de la medición.

En la tabla 2 se puede observar que el ETM es de 5.0 kg. Esto quiere decir que se “elimina” la variación biológica de la persona y el error técnico del instrumento de medición que se utilizó para obtener el puntaje de la variable de interés. Todo eso representa el cambio aleatorio. Entonces, al conocer la magnitud del cambio aleatorio, se puede determinar con un 95% de confianza, el valor a partir del que se observa un verdadero efecto del entrenamiento físico; es decir, el cambio sistemático. En la tabla 2 se observa que ese puntaje umbral es de ± 9.8 kg. Por lo tanto, se puede conocer la cantidad de personas que respondieron y las que no respondieron. En el ejemplo, 60% de las personas no sobrepasaron 9.8 kg, y 40% sí lo hicieron.

Entonces, a pesar de que existe una diferencia estadísticamente significativa entre los promedios iniciales y finales ($p = 0.049$), se observó que 6 de las 10 personas no respondieron a la intervención, lo cual refleja una alta variación entre los individuos (Figura 1). Este tipo de análisis permite juzgar con mayor integralidad los resultados más allá de lo que el análisis de la prueba de hipótesis indica, lo cual es beneficioso para evaluar objetivamente el efecto de una intervención, así como para identificar posibles variables mediadoras o moderadoras de los efectos de dicha intervención. Estos insumos son relevantes para el replanteamiento de investigaciones que sigan una misma línea, o para proponer nuevas variables independientes en futuros estudios.

Figura 1. Gráfico de respuestas individuales en donde se muestra a quienes respondieron a la intervención y quienes no lo hicieron. El criterio o “umbral” para clasificar a los participantes es el umbral de ± 9.8 kg obtenido después de haber considerado el error técnico de la medición y el 95% de probabilidad en el cambio sistemático producido por el entrenamiento físico (líneas discontinuas). Un valor igual o menor a -9.8 kg reflejaría un efecto perjudicial del ejercicio, y un valor igual o superior a $+9.8$ kg reflejaría un efecto beneficioso del ejercicio. Los valores ubicados entre las líneas discontinuas representan cambios no relevantes o triviales.



Una observación clave que se puede inferir de este tipo de análisis, es que se debe aumentar lo más posible el cambio sistemático (i.e., **MAX**imizar varianza verdadera) y reducir lo más posible el cambio aleatorio (i.e., **MIN**imizar varianza de error); mientras se controlan todos los factores posibles que puedan afectar la intervención (i.e., **CON**trolar varianza de error). A esto se le conoce como MAXMINCON^{[398][508]}. Esto permitiría tener un panorama más real acerca de la efectividad de una intervención de ejercicio físico sobre alguna variable de interés. También es evidente que, a pesar de haberse encontrado diferencias estadísticamente significativas entre los promedios, se puede determinar que muchos participantes no alcanzan el umbral de mejoría como resultado del programa de ejercicio, lo que permite reflexionar a los investigadores acerca de estrategias para mejorar la prescripción del ejercicio en una mayor cantidad de personas. Finalmente, el estudio de quienes responden y quienes tienen baja o nula respuesta al ejercicio se vislumbra como una oportunidad para diseñar programas de ejercicio mucho más individualizados que generalizados.

3.1.4. Variabilidad dentro del mismo individuo

Estudiar la variabilidad en la respuesta a un tratamiento o entrenamiento que ocurre dentro de una misma persona, permite conocer la eficacia de ese tratamiento después de apartar otras posibles razones que podrían estar ocultando el verdadero efecto de esa intervención, fenómeno que se conoce también como “ruido”. Las respuestas individuales reproducibles o consistentes son aquellas que podrían explicarse por diferencias entre sujetos en características o rasgos estables heredados y adquiridos, mientras que las respuestas aparentemente aleatorias podrían deberse a

cambios en las características o estados de los sujetos entre administraciones del tratamiento^[345]; es decir, en los cambios que pudieron ocurrir mientras los sujetos participaban en un estudio o eran sometidos a un programa de entrenamiento. Esos cambios la mayoría de las veces no los puede controlar el investigador.

Los métodos estadísticos necesarios para estudiar la variabilidad dentro del mismo individuo pueden ser avanzados (e.g., modelo a escala de ubicación de efectos mixtos)^[742] o relativamente sencillos para la mayoría de los investigadores. Así, cuando se diseña un estudio, las respuestas individuales se presentan como una desviación estándar (DE) mayor de los puntajes de cambio en el grupo experimental que en el grupo control. Por ello, es esencial que los investigadores reporten las DE de los puntajes de cambio junto con la media (M) o promedio^[345]. Las respuestas individuales (RI) se calculan mediante una DE, a partir de la raíz cuadrada de la diferencia entre los cuadrados de las DE de los cambios entre el grupo experimental (DE_{Exp}) y el grupo control (DE_{Con}):

$$DE_{RI} = \sqrt{(DE_{Exp})^2 - (DE_{Con})^2}$$

La DE_{RI} debe interpretarse como la cantidad en la que el efecto promedio neto del tratamiento difiere típicamente entre los participantes; o lo que se conocería como la verdadera respuesta individual al tratamiento^[39].

Una vez que se obtiene la DE_{RI} , se calculan los intervalos de confianza y los tamaños de efecto estandarizados. Luego, se puede interpretar la magnitud de las DE_{RI} en relación a una diferencia mínima clínicamente importante (DMCI) (en inglés, "Minimal Clinically Important Difference", MCID) o al cambio mínimamente significativo (CMS) (en inglés, "Smallest Worthwhile Change", **SWC**), que generalmente se establece en $0.2 \times DE$ de la línea base^[89]. El valor 0.20 proviene de la recomendación de Cohen para representar un tamaño de efecto pequeño, pero no trivial^[677].

El error técnico (ET) o error típico de la medición (ETM) se ha utilizado también para conocer la variabilidad dentro de la misma persona, ya que representa una combinación entre el error de la medición (derivado de cualquier instrumento que se utilice para medir una variable) y la variabilidad diaria o biológica que se espera de la misma persona^{[343][349][618]}. Se debe recordar que cada vez que se mide una variable, el valor observado está influenciado tanto por el valor real del individuo como por el error de medición aleatorio^[89].

El ETM es el resultado de una combinación del error técnico introducido por el equipo o por la confiabilidad del experimentador y la variabilidad aleatoria diaria de los factores biológicos capaces de alterar la variable de interés medida. Los factores biológicos que contribuyen a la variabilidad diaria aleatoria incluyen elementos que pueden afectar el estado mental o físico de un individuo cuando ejecuta la prueba (e.g., factores ambientales, ritmo circadiano, sueño, dieta, ejercicio previo, estado de hidratación).

Se espera que el ETM introduzca ruido aleatorio en las mediciones pre y post de un estudio; por lo tanto, si bien este ruido aleatorio probablemente ejerce una influencia mínima en la capacidad de detectar diferencias grupales en una intervención de entrenamiento, puede influir en el cambio observado de un individuo después del entrenamiento^[89]. Así, los datos de la persona son más consistentes (o menos variables) cuanto menor sea el ETM^[343].

ETM_{absoluto} = $\sqrt{\sum D^2 / 2N}$	En la fórmula, $\sum D^2$ es la diferencia entre las mediciones hechas en cada persona elevada al cuadrado y N es la cantidad de mediciones realizadas a esa persona (e.g, pre-test a post-test, N= 2). También podría ser la diferencia entre puntajes obtenidos a una misma persona por dos investigadores distintos ^{[343][427]} .
--	--

El valor obtenido es absoluto, pero también se puede calcular de forma relativa para una mayor facilidad de interpretación cuando se desean comparar los valores de variabilidad o de error entre distintas personas. Para obtener el valor relativo del ETM (ETM_{relativo}), se utiliza el valor promedio de la variable (VPV) que fue medida en ambas ocasiones; es decir, el promedio de los promedios:

ETM_{relativo} = (ETM/VPV) x 100	Se obtiene un error o una variabilidad que se puede expresar en términos relativos, en este caso, en términos porcentuales al promedio total de la variable de interés. Entre más pequeño el valor, se considera que la precisión es mejor ^{[343][427]} .
---	--

Ya que la verdadera respuesta individual debe estimarse a partir de una respuesta observada (que además contiene un error aleatorio), las respuestas observadas más bajas tienden a subestimarse y las respuestas observadas más altas tienden a sobrestimarse^[618]. Este problema se puede reducir utilizando predictores lineales no sesgados para estimar la verdadera respuesta de un individuo, pero siempre habrá influencia cuando el error dentro de la misma persona sea considerable. Aunque no existen criterios matemáticamente fundamentados, se podría considerar como un cambio importante si el valor del ETM es de 1.5 a 2 veces de magnitud, que ya que el ETM representa la variabilidad medida dentro de la persona^[343].

Sin embargo, ese es un criterio subjetivo y no necesariamente podría generalizarse a todas las situaciones en donde exista una intervención de ejercicio físico; es decir, no sería lo mismo esperar un cambio en el ETM en un estudio con personas sedentarias que en un estudio con atletas de élite, ya que para unos un ETM= 1.5 puede ser suficiente para mejorar su salud, pero para otros, un ETM <1.5 puede representar la diferencia entre ganar una medalla o no. Es por ello que se recomienda contextualizar el estudio, y definir, *a priori*, cuál es el ETM umbral o aceptable para el tipo de estudio que se desea realizar, aunque ese enfoque también posee críticas importantes.

El ETM puede mostrar heterocedasticidad (i.e., aumenta a medida que aumentan las unidades de medición) y para corregir ese problema, se recomienda normalizarlo, ya sea usando una transformación logarítmica o expresando el valor como un porcentaje de la media respectiva; es decir, como el coeficiente de variación del ETM^[660]. Otra característica del ETM es que no considera las variaciones dentro de los sujetos ("within-subject variation")^[352], por lo que se debe complementar con otro tipo de información para una toma de decisiones más integral.

Para complementar la información, la DE_{RI}, se puede interpretar en razón de valores umbrales establecidos para el cambio promedio estandarizado, usando, por ejemplo, 0.1 (efecto pequeño), 0.3 (efecto moderado) y 0.6 (efecto grande)^[345], o en función del CMS^{[343][347]}, tal y como se mencionó anteriormente.

$$\text{CMS} = \text{ETM} \times \sqrt{2} \times 0.2$$

Representa el efecto o cambio más pequeño en la variable de interés que se considera de significancia práctica o importante en un contexto determinado (e.g., rendimiento deportivo, ámbito clínico o de la salud)^{[343][678]}.

La determinación del CMS depende del área de investigación, y de los datos que se puedan ir obteniendo de la teoría junto con la práctica. Por ejemplo, en algunas áreas, como el rendimiento deportivo, un cambio de 1% en una variable como la velocidad puede significar ganar una medalla de oro; mientras que, en otros ámbitos, un cambio de 5 mm Hg en la presión arterial sistólica puede ser muy importante para reducir el consumo de medicinas. Por lo tanto, todavía queda por explorar, para muy diversas áreas, esos CMS para ponerlos en un contexto específico y no generalizado.

En la tabla 3, se ilustran los conceptos descritos anteriormente. Como puede observarse, se parte de que existen diferencias entre grupos (i.e., el análisis "between-groups"), argumento que se apoya en una probabilidad de $p = 0.041$; es decir, $p < 0.05$ considerado mundialmente como el umbral para definir que hubo un hallazgo estadísticamente significativo^[776]. No hay duda de que, a nivel grupal, quienes realizan ejercicio obtuvieron un mayor cambio en la variable de interés en comparación con el cambio del grupo control. Por su parte, el análisis de la variabilidad individual muestra otra información valiosa que hoy en día se estimula reportar en los manuscritos enviados a revistas científicas.

Por ejemplo, la DE_{RI} obtenida de 165.2 se interpreta como la cantidad en la que el efecto promedio neto del ejercicio difiere típicamente entre los participantes; es decir, es la verdadera respuesta individual al tratamiento^[39]. El ETM_{absoluto} fue de 42.3 unidades, y representa el error técnico introducido por el equipo y/ o por la confiabilidad del experimentador y la variabilidad aleatoria diaria de los factores biológicos que afectan la variable de interés. En términos relativos, representa el 15.2% (ETM_{relativo}). Por su parte, el CMS o SWC fue de 12.0 (calculado como $0.2 \times DE$ de la línea base)^[89].

Tabla 3. Datos ficticios para ilustrar los posibles tipos de análisis grupales e individuales, por medio de los cuales se comprende la variabilidad entre participantes (“between-subjects”, inter-sujeto) y la variabilidad dentro de cada participante por separado (“within-subjects”, intra-sujeto).

Participantes	$\Delta_{Control}$	Umbral >2.5%	Tipo de respuesta	Decisión Final	Participantes	$\Delta_{Experimental}$	Umbral >2.5%	Tipo de respuesta	Decisión Final
Ana	-3.5	No responde	Incierta	Podría funcionar	Paola	-53.7	No responde	Adversa	No funciona
Beatriz	-60.4	No responde	Adversa	No funciona	Rosa	-23.9	No responde	Incierta	Podría funcionar
Carolina	-11.7	No responde	Incierta	Podría funcionar	Roxana	12.4	Responde	Incierta	Podría funcionar
Paulina	25.3	Responde	Incierta	Podría funcionar	Zaida	-7.2	No responde	Incierta	Podría funcionar
Melinna	65.3	Responde	Positiva	Funciona	Yamileth	26.5	Responde	Incierta	Podría funcionar
Isaura	-43.4	No responde	Adversa	No funciona	Fiorella	123.3	Responde	Positiva	Funciona
Jessenia	-44.2	No responde	Adversa	No funciona	Valeria	66.3	Responde	Positiva	Funciona
Andrea	-5.8	No responde	Incierta	Podría funcionar	Gabriela	143.7	Responde	Positiva	Funciona
Martha	-6.7	No responde	Incierta	Podría funcionar	Cielo	700.0	Responde	Positiva	Funciona
Alvaro	-18.2	No responde	Incierta	Podría funcionar	Rodrigo	109.4	Responde	Positiva	Funciona
Bruno	29.4	Responde	Incierta	Podría funcionar	Pedro	43.7	Responde	Positiva	Funciona
Iván	27.7	Responde	Incierta	Podría funcionar	Alejandro	93.9	Responde	Positiva	Funciona
Javier	185.6	Responde	Positiva	Funciona	Gerardo	104.6	Responde	Positiva	Funciona
Eduardo	64.6	Responde	Positiva	Funciona	Luis	279.0	Responde	Positiva	Funciona
Franco	64.1	Responde	Positiva	Funciona	Juan	107.5	Responde	Positiva	Funciona
Guillermo	51.6	Responde	Positiva	Funciona	José	177.3	Responde	Positiva	Funciona
M=	20.0				M=	118.9			
DE=	59.8				DE=	175.6			
p=	0.041								
DE_{IR}=	165.2								
ETM_{absoluto}=	42.3								
ETM_{relativo} (%) =	60.5								
CMS=	12.0								

Notas: M= Media o promedio, DE= Desviación estándar, p = significancia estadística, DE_{IR} = Desviación estándar de las respuestas individuales, ETM_{absoluto} = Error típico de la medición absoluto; ETM_{relativo} = Error típico de la medición relativo en %; CMS = cambio mínimamente significativo.

El **Tipo de respuesta** se basa en el ETM_{absoluto}, que de acuerdo con otros autores se puede multiplicar por 1.5 o 2.0 dependiendo de la variabilidad. En el ejemplo de la tabla, el ETM_{absoluto} se podría multiplicar por 2.0 para reflejar más claramente el tipo de respuesta.

En la tabla 3 también se puede observar el umbral $>2.5\%$, que, en el contexto imaginario del ejemplo, representaría un cambio definido arbitrariamente como el valor sobre el cual se consideraría que existe un efecto más que biológico producto de la intervención o no intervención. El paso siguiente, sería analizar a cada persona del grupo control y del grupo experimental, para determinar si alcanzaron ese umbral y definir si respondieron o no a la intervención. La hipótesis es que quienes no recibieron el tratamiento (i.e., cada persona del grupo control) no debería cambiar o tal vez podrían empeorar, pero no se esperaría que mejoraran o alcanzaran el umbral. Por su parte, la hipótesis para cada persona del grupo experimental es que deberían mejorar; es decir, alcanzar un cambio que sea igual o superior al umbral de 2.5% , y que ese cambio se explicaría por la intervención del programa de ejercicios.

Como se puede observar, en el grupo control, se encontró que 8 personas no respondieron a su condición de no realizar ejercicio, lo cual apoya la hipótesis. Sin embargo, hubo 8 personas que superaron el umbral de 2.5% , lo cual rechaza la hipótesis de que al no hacer ejercicio empeorarían, o al menos no mejorarían. En síntesis, 50% de los participantes respondió en la dirección esperada y 50% no lo hizo. Por su parte, en el grupo experimental, se encontró que 3 personas no respondieron al ejercicio físico, mientras que 13 si lo hicieron, lo cual apoya la hipótesis de que el ejercicio mejora la variable de interés. En síntesis, 19% de los participantes no respondió al ejercicio, y el 81% sí lo hizo. En conjunto, esta interpretación muestra la variabilidad individual y permite reflexionar acerca de las razones por las cuales no todas las personas responden de igual manera aun reduciendo el error asociado a la medición de la variable de interés.

En la tabla 3 también se puede observar la columna "tipo de respuesta". Este "tipo de respuesta" es similar al umbral mínimo; sin embargo, el valor proviene del $ETM_{absoluto}$ de 42.3. Su interpretación sigue las recomendaciones de Walsh et al.^[733]:

- a) Positiva si > 42.3
- b) Incierta entre -42.3 y 42.3
- c) Adversa si < -42.3 .

Con base en esta clasificación, se encontró que en el grupo control hubo 5 personas (31%) que respondieron positivamente a no hacer ejercicio, 8 personas (50%) con respuesta incierta y 3 personas (19%) con respuesta adversa. Como se mencionó anteriormente, la hipótesis era que quienes no hacían ejercicio se mantendrían igual o empeorarían, pero no mejorarían. En síntesis, 69% de los participantes no mejoraron o empeoraron, mientras que 31% respondió en la dirección equivocada o teóricamente opuesta al mejorar sin haber realizado ejercicio. Por su parte, en el grupo experimental, se encontró que 11 personas (69%) respondieron positivamente, 4 de forma incierta (25%), y 1 (6%) de forma adversa. Este patrón concuerda con la hipótesis, al menos en su mayoría, en donde se observa que el 69% respondió en la dirección y magnitud esperada debido al ejercicio; aunque 31% no lo hicieran.

Lo valioso de este tipo de análisis es que se puede llegar a una "decisión final" (Tabla 3). Se puede concluir que para 5 personas (31%) del grupo que no realizó ejercicio, hubo un cambio positivo en la variable de interés más allá de la variabilidad biológica y del error técnico de la medición. Las razones por las que esas personas mejoraron pueden ser múltiples; por ejemplo,

hicieron ejercicio a pesar de que se les dijo que no lo hicieran, mejoraron su alimentación, cambiaron su nivel de actividad física, entre otras. También, para 8 personas (50%) no realizar ejercicio podría funcionar para mejorar la variable de interés; mientras que para 3 personas (19%) no realizar ejercicio les empeoraría los valores de la variable de interés. Esto indica una alta variabilidad en las respuestas de los participantes en el grupo control, y permite reflexionar acerca de los cuidados y las instrucciones que se les debe proporcionar a las personas que colaboran participando como controles en las investigaciones.

Por último, las decisiones finales acerca del efecto del programa de ejercicios en los participantes del grupo experimental indican que para 11 personas (69%), completar el programa de ejercicios les funciona para mejorar la variable de interés, y adicionalmente para 4 personas (25%) podría beneficiarles. Solamente para 1 persona (6%), el efecto del ejercicio fue inverso; es decir, en lugar de beneficiarlo, lo perjudicó. Estas conclusiones se basan integralmente en los análisis individuales, y permiten introducir el tema del problema de las personas que responden (i.e., "responders") y las que no responden ("non responders") a una intervención. Este tema se abordará a continuación.

3.1.5. Respondedores vs. No respondedores

La capacidad de adaptarse a un entrenamiento, la "entrenabilidad", o la respuesta de entrenamiento individual, se puede definir como la respuesta consistente de un individuo a una intervención específica^[728]. Es decir, también se podría definir como "cuánto mejor (o peor) respondió un individuo a un programa de entrenamiento de ejercicios específico, en comparación con la respuesta típica/promedio que se observa en el grupo" (p. 38)^[728]. Desde el punto de vista estadístico, sería lo que se conoce como la interacción "sujeto por entrenamiento" (i.e., efecto aleatorio) en un modelo lineal mixto.

Así, cada vez que a un grupo de personas se les aplica algún tipo de intervención, como, por ejemplo, un programa de entrenamiento físico cuyo propósito sea determinar el cambio en alguna variable de interés (e.g., felicidad, fuerza, salto vertical, potencia, VO_2 máx), siempre se va a encontrar que no todos respondieron de la misma forma a la intervención; es decir, su grado de entrenabilidad fue distinto^[91]. Por lo tanto, existirá variabilidad o heterogeneidad en las respuestas y adaptaciones a ese tipo de entrenamiento. Por ejemplo, en la respuesta del VO_2 máx, se ha observado que después de realizar un programa de entrenamiento, el VO_2 máx de algunas personas se redujo en $100 \text{ ml}\cdot\text{min}^{-1}$ mientras que en otras aumentó en más de $1100 \text{ ml}\cdot\text{min}^{-1}$ ^[102]. En otros estudios se ha encontrado que los participantes pueden mejorar su fuerza en una repetición máxima (1RM) desde un 0% hasta un 250%, pero que entre 7% y 10% de los participantes más bien empeoran o muestran respuestas no deseadas al entrenamiento físico en otras variables^[104].

Existen diferentes fuentes de variabilidad que afectan las respuestas individuales al entrenamiento físico. Se han descrito la variabilidad técnica, la variabilidad biológica del día a día, la variabilidad debida al entrenamiento físico independientemente del individuo, la variabilidad debida al individuo independientemente del entrenamiento físico, la variabilidad en las respuestas al mismo ejercicio de entrenamiento entre individuos, y la variabilidad intrasujeto^[728]. Es por ello, que aparte de las comparaciones grupales, es sumamente relevante conocer las respuestas individuales a una intervención, para lo cual comúnmente se han considerado dos categorías de

personas en función de su respuesta al tratamiento o entrenamiento físico: a) las que responden o respondedoras (en inglés, "responders") y b) las que no responden o no respondedoras (en inglés, "non-responders").

El análisis consiste en identificar ambos tipos de personas, así como las características que puedan explicar la respuesta al ejercicio, y se ha convertido en un tema metodológico que requiere de una estricta revisión, debido a las implicaciones que tiene en el planteamiento de nuevas investigaciones y a la calidad y pertinencia de las conclusiones derivadas de un estudio, las cuales eventualmente se podrían resumir por medio de un meta análisis. También se debe considerar que algunos investigadores^{[45][51][565]} afirman que no existen los no respondedores, sino que lo que posiblemente ocurre es que el estímulo de entrenamiento o incluso la combinación de modalidades de ejercicio^[565] no sean lo suficientemente importantes para esas personas, lo que crea "artefacto" o "ruido" que opaca el estímulo proporcionado por el entrenamiento. Ese argumento requiere de mayor corroboración; incluso, se critica fuertemente que no existan definiciones claramente establecidas; por ejemplo, un no respondedor puede referirse a la falta de un cambio clínicamente significativo, la falta de un cambio medible, un valor por encima del error técnico de la prueba o al porcentaje más bajo de sujetos en términos de respuesta^[565].

La clasificación dicotómica o binaria en función de la respuesta es estrictamente subjetiva; pues no posee un fundamento matemático, sino más bien uno práctico y lógico. Algunos consideran que la clasificación dicotómica es peyorativa y que tiende a confundir a los lectores acerca de los probados efectos positivos del ejercicio y el entrenamiento físico para la salud y la prevención de enfermedades^[565]. De esa forma, también se podrían utilizar clasificaciones politómicas, en las que, por ejemplo, las personas podrían ser categorizadas como "súper respondedores", "respondedores promedio", "respondedores bajos", "no respondedores" o "respondedores adversos" a la misma dosis, programa de ejercicio o de entrenamiento físico. Incluso se podría clasificar a las personas en categorías de respondedores "positivos", "negativos" y "triviales", por lo que actualmente no existe una guía o recomendación clara para generar categorías más allá del criterio experto de los investigadores.

Por otra parte, los parámetros o umbrales para definir a partir de qué valor se pueden hacer estas clasificaciones tampoco están claramente descritos en la literatura, y posiblemente no puedan estandarizarse; sin embargo, en algunas áreas ya se han reportado algunos valores en variables de interés considerados como los mínimos para clasificar a las personas con base en su respuesta a un estímulo o intervención. Por ejemplo, un aumento en $1.75 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ en el $\text{VO}_2\text{máx}$ es considerado como importante, ya sea de relevancia clínica o de relevancia para el rendimiento deportivo^[491], y un cambio de 1 g en la masa de hemoglobina debido al entrenamiento se puede asociar con un cambio de $4 \text{ ml}\cdot\text{min}^{-1}$ en el $\text{VO}_2\text{máx}$ ^{[536][638]}.

A este método de clasificación se le conoce como "valores umbrales" o "puntajes de corte", y se sabe que tiene limitaciones porque la variabilidad en la medición diaria de una variable de interés medida durante varios días no considera la variabilidad biológica que existe durante la duración del período de tratamiento^[618]. Sin embargo, para otros investigadores, dicotomizar las categorías no es apropiado, porque la clasificación basada en umbrales o puntajes de corte tiende a clasificar erróneamente a las personas que tienen una alta probabilidad de mostrar una respuesta positiva similar a la de los que no responden; es decir, a pesar de basarse en un enfoque robusto, se sobreestima la prevalencia de los no respondedores^[90]. Por lo tanto, esta selección arbitraria y

subjetiva no permite una clasificación consistente de las personas, lo que, a su vez, evita poder realizar comparaciones entre diferentes estudios.

Algunos investigadores han propuesto un método por medio del cual los respondedores diferenciales se pueden clasificar después de tener en cuenta la presencia del error aleatorio que se cuantifica a partir de un grupo de control apareado en el tiempo^[195]. Con el método, quienes exceden el error aleatorio de la respuesta promedio del grupo de intervención pueden etiquetarse con seguridad como respondedores altos y bajos. La cantidad de respondedores diferenciales será proporcional a la relación de varianza en los grupos de ejercicio y control. Este método podría ayudar a evitar la clasificación errónea de individuos en función de un error aleatorio y, a su vez, aumentar la replicabilidad de los estudios de respuesta diferencial.

Otros investigadores han utilizado el ETM para indicar que cuando éste se duplica (i.e., 2 x ETM), se puede establecer ese valor como el umbral a partir del cual se pueden clasificar a las personas como respondedoras o no respondedoras^{[90][343]}. También se han utilizado diferentes umbrales basados en el ETM, el CV o en el CMS para el cambio, por ejemplo, en el VO₂máx (Tabla 1).

Tabla 1. Ejemplos de la variedad de umbrales para definir la respuesta al entrenamiento*.

Referencia	Fórmula	Valor real del umbral
Astorino & Schubert ^[38]	< 1 CV	< ~1.3 ml·kg ⁻¹ ·min ⁻¹
Gurd et al. ^[302]	< 2 x ETM	< 1.74 ml·kg ⁻¹ ·min ⁻¹
Lee et al. ^[421]	< CMS	< 1.75 ml·kg ⁻¹ ·min ⁻¹
Phillips et al. ^[563]	< 1 x ETM	< 0.86 ml·kg ⁻¹ ·min ⁻¹
Williams et al. ^[753]	< CMS - 1 x ETM	< 1.80 ml·kg ⁻¹ ·min ⁻¹

ETM: error típico de la medición; CV: coeficiente de variabilidad; CMS: cambio mínimamente significativo.

* La lista no es exhaustiva.

Utilizando otro umbral, se podría clasificar a una persona como respondedora, cuando la masa de hemoglobina total aumenta más de 1.7% después de una intervención de entrenamiento en altitud, mientras quienes no lo hacen, se les clasificaría como no respondedores^[536]. Ese valor sobre 1.7% excede el ETM de la variable, y varía de laboratorio a laboratorio; por lo que un umbral se puede establecer sabiendo el ETM o variabilidad de la variable de interés que ha sido medida sistemáticamente en un ambiente determinado, como, por ejemplo, en un laboratorio de investigación, siguiendo procedimientos estándares de medición y utilizando siempre el mismo equipo calibrado siguiendo las recomendaciones del fabricante.

Este tipo de umbrales se sustentan en que el cambio real para las respuestas observadas y que logren duplicar el ETM por encima de cero tienen una probabilidad >90% de ser positivas^[343]. Para evitar confusiones e interpretaciones erróneas al utilizar únicamente el ETM, se ha recomendado utilizar intervalos de confianza de la estimación calculada (IC, generalmente al 90% o 95% de precisión) y el CMS, ya que complementan el ETM y proporcionan más información que el enfoque basado solamente en una clasificación dicotómica basada en umbrales^{[343][347][678]}. Este enfoque parece ser superior cuando se realizan mediciones repetidas en la variable dependiente, ya que las medidas repetidas, y tratamientos de mayor duración y de mayor volumen, intensidad y duración del ejercicio, proporcionan una estimación más precisa de la verdadera respuesta de los participantes, y además ayudan a reducir la tasa de no respondedores^{[331][565]}.

IC95% = $\pm 1.96 \times \text{ETM}$	Cuantifica con una incertidumbre del 95% las estimaciones que no se pueden medir directamente; proporciona un rango de valores posibles con base en los datos observados ^[678] .
--	---

Por ejemplo, si como lo sugieren Swinton et al.^[678], si se calcula el IC50%, habría que utilizar el ETM para medidas repetidas en el grupo control (i.e., para un estudio típico de grupo experimental y grupo control, con mediciones pre-test y post-test). De esa manera se podría clasificar a los participantes en no respondedores, inciertos y respondedores positivos.

ETM = $\text{DE}_{\text{Control}} / \sqrt{2}$	Si el valor encontrado en la persona sobrepasa todo el IC50% del valor umbral, entonces a la persona se le considera como respondedora. Usar el IC50% significa que hay un 75% de probabilidad de una respuesta positiva para esta persona, es decir, "probable". Si el IC50% cruza el umbral de respuesta, a la persona se le clasifica como de "respuesta incierta", mientras que, si todo el IC50% se encuentra por debajo del umbral de respuesta, se define que la persona es "no respondedora" ^{[90][345]} .
---	---

Se ha encontrado que las diferencias interindividuales en los cambios inducidos por el entrenamiento superan la variabilidad técnica, biológica diaria y aleatoria dentro de los sujetos que fueron sometidos a un entrenamiento para mejorar el $\text{VO}_2\text{máx}$ ^[491], en donde se reportó una tasa de falta de respuesta del 18% y un 49% de las personas eran clasificadas como respondedoras pues mostraron un aumento en el $\text{VO}_2\text{máx}$ más alto que el CMS.

Existen otros enfoques en donde, por ejemplo, en lugar de clasificar a las personas, se clasifican los puntajes de cambio^[618]. Así, los puntajes de cambio individuales se pueden clasificar como "improbable", "probable", "muy probable" de que representen un cambio real en la variable dependiente de interés. A pesar de que este tipo de enunciado es de naturaleza probabilística; es decir, la probabilidad de que represente un cambio verdadero también requiere de varias decisiones arbitrarias, como, por ejemplo, asignar ciertas etiquetas de rango de probabilidad como "muy probable", así como asignar un valor para el CMS^[618].

Otro aspecto metodológico que surge al comparar a las personas que responden y a las que no responden, es la errónea división en categorías que consideren los niveles iniciales de la variable de interés. Es decir, algunas veces existe la tentación, en ausencia de significancia estadística (i.e., $p < 0.05$), de escudriñar los datos e hipotetizar si haciendo una división de participantes, se podría encontrar ese deseado $p < 0.05$ ^[776], y posiblemente, que el manuscrito tenga mayor posibilidad de ser aceptado en una revista. Este escenario es muy claro y sus errores estadísticos se han descrito recientemente^[544], en donde ese análisis secundario de datos implica determinar si aquellos individuos con niveles bajos de la variable de interés antes de la intervención son los mismos que tienen el mayor cambio en esa variable después de la intervención (e.g., programa de entrenamiento físico). Conceptualmente, este argumento tiene sentido porque es consistente con la noción de que las personas que se benefician de un programa de entrenamiento son aquellas que tienen un mayor margen de mejora (i.e., aquellas que están menos entrenadas o que son sedentarias).

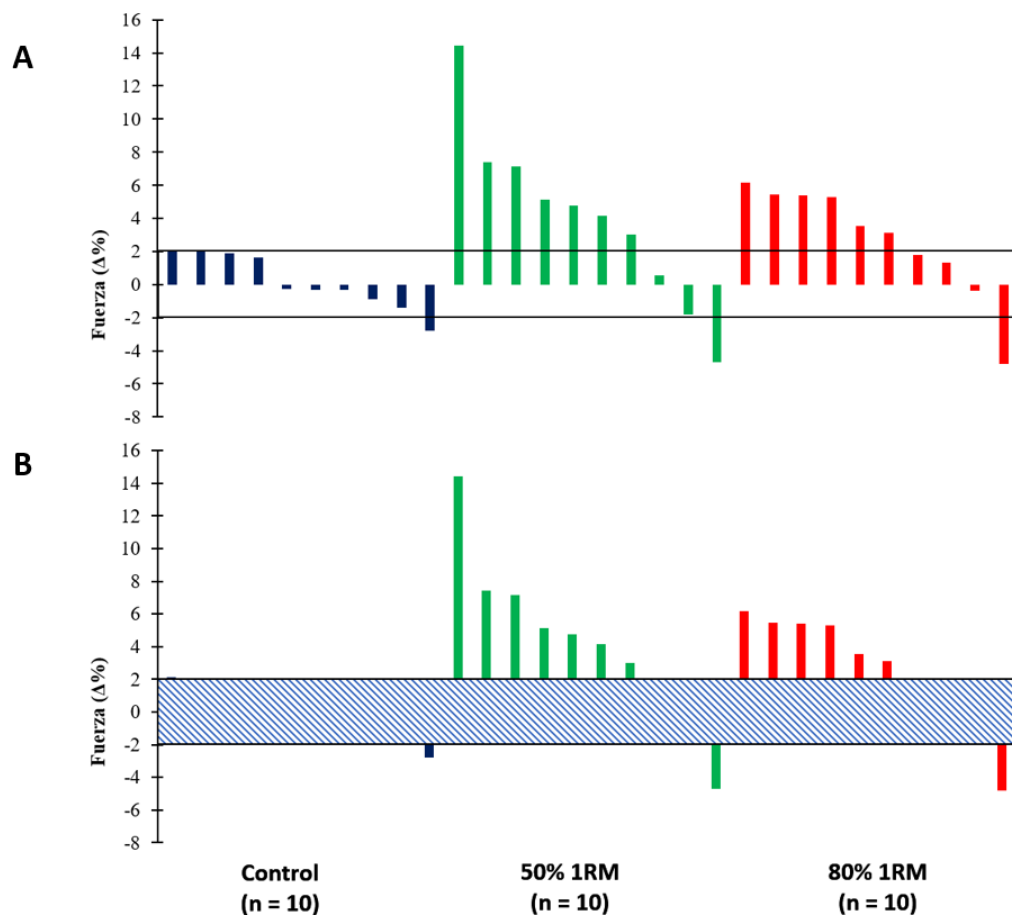
Sin embargo, ese tipo de análisis es estadísticamente inapropiado porque el resultado no es independiente de cómo se agruparon los individuos (e.g., niveles bajos de fuerza antes de la intervención) y por lo tanto viola un supuesto fundamental para probar las diferencias de grupo. Así, en esencia, si cada persona recibe el programa de entrenamiento físico para mejorar la fuerza, es probable que esta forma de análisis *post hoc* produzca el hallazgo postulado (i.e., se vea que aumentó la fuerza), pero no debido a un efecto biológico que se pueda explicar por la fisiología; es decir, se puede encontrar el efecto deseado de aumento en la fuerza, pero artificialmente o con trampa. Por tal motivo, comparar las diferencias grupales en la variable de interés cuando los grupos han sido categorizados por la misma variable de interés, viola el supuesto de independencia estadística y asegura que se observarán las diferencias de grupo, algo conocido como regresión a la media^[544], lo que conduce a una interpretación errónea de los datos y a conclusiones sesgadas.

La graficación de las respuestas individuales es importante para describir los hallazgos del estudio y brindar a los lectores una mejor interpretación de lo que ha ocurrido^{[89][219]}. Por ejemplo, en un estudio con datos hipotéticos acerca de dos intensidades de entrenamiento contra resistencia y su efecto en la fuerza muscular después de 5 sesiones semanales llevadas a cabo durante 4 semanas, inicialmente se encontró que hubo personas que mejoraron, algunas que no mejoraron y otras empeoraron (Figura 1A). En la figura 1A se observa que el 40% de los participantes del grupo control que no realizó ejercicio, mejoraron la fuerza, pues obtuvieron $\Delta\% > 0$, y, por otra parte, el 60% la redujo, pues obtuvieron valores $\Delta\% < 0$. En el grupo de participantes que entrenó al 50% de 1RM, se puede observar que el 80% de los participantes mejoraron la fuerza, y que 20% la redujo.

Finalmente, en el grupo de participantes que entrenó al 80% de 1RM, se puede observar que el 80% de los participantes mejoraron la fuerza, y que 20% la redujo. También se pueden observar dos líneas continuas de $\Delta \pm 2\%$, que representan el CMS definido antes de empezar el estudio (i.e., *a priori*) y que se esperaría al realizar un estudio de ejercicio contra resistencia. Esto significa, que independientemente de la dirección de la respuesta (i.e., arriba de cero o debajo de cero), el umbral para definir el CMS es 2%. El CMS se debería sustentar en evidencia científica, o como se mencionó anteriormente, al menos en el ETM y en el IC90% o IC95%^{[90][345][678]}. Sin embargo, no se descarta que sea posible que, para una amplia gama de contextos, ese CMS tenga que ser definido subjetivamente con base en algún criterio de juicio de expertos.

Si se eliminara la zona de $\Delta \pm 2\%$ (vista en la Figura 1B por medio de un patrón de líneas diagonales), para obtener el cambio "puro" producido por la intervención del ejercicio, la interpretación ahora cambia sustancialmente. Se puede observar, que 10% de los participantes del grupo control empeoraron solamente por no realizar ejercicio, y 90% se mantuvo sin cambios. En conjunto, el 100% respondió de la forma esperada, pues no realizaban ejercicio. También se puede observar que en el grupo de participantes que entrenó al 50% de 1RM, 70% mejoró la fuerza, 10% redujo la fuerza y 20% no cambió (cubiertos por los umbrales $\Delta \pm 2\%$). Por lo tanto, 70% respondieron en la forma esperada (i.e., "responders") y 30% no respondieron de la forma esperada (i.e., "non-responders").

Figura 1. Cambio porcentual ($\Delta\%$) en la fuerza muscular en los participantes de tres grupos experimentales. El porcentaje de personas definidas como respondedoras o como no respondedoras dependerá del umbral de clasificación. En este ejemplo, el umbral se estableció en $\Delta \pm 2\%$, el cual se muestra con las líneas continuas en la Figura 1A y como el patrón diagonal en la Figura 1B. Los valores dentro de esos umbrales no muestran cambios, y los valores que sobrepasan el umbral, hacia arriba o abajo, representan un cambio mínimamente importante para la variable de estudio.



Finalmente, en el grupo de participantes que entrenó al 80% de 1RM, el 60% mejoró la fuerza, el 10% redujo la fuerza y el 30% no cambió (cubiertos por los umbrales $\Delta \pm 2\%$). Por lo tanto, 60% respondieron en la forma esperada ("responders") y 40% no respondieron de la forma esperada (i.e., "non-responders"). Sin embargo, la descripción anterior puede explicarse en su totalidad por la variación aleatoria intraindividual de los valores de la fuerza "verdaderos" tanto en la medición inicial o de línea de base como por las mediciones de seguimiento (e.g., medidas repetidas) y por los artefactos asociados a la regresión hacia la media^{[39][40]}. En síntesis, la variabilidad aleatoria intraindividual hace que parezca que los participantes responden diferente a una intervención, cuando en realidad puede haber diferencias interindividuales clínicamente irrelevantes en la verdadera respuesta a una intervención.




También se pueden generar gráficos más complejos en los que se puede observar no solamente una variable de una persona, sino varias variables dependientes para cada individuo^{[6][91]}. El ejemplo que se presenta en la Figura 2 muestra datos hipotéticos de los patrones individuales de dos grupos de participantes (i.e., grupo control, grupo experimental).

Las respuestas en la dirección hipotéticamente esperada (i.e., respondedores) para cada variable de interés se muestran en color verde y se definen como una respuesta positiva. Las respuestas negativas (i.e., no respondedores) o en la dirección opuesta al tratamiento, se muestran en color rojo; y las respuestas neutras, es decir, donde no hubo un cambio, se muestran en color amarillo. La creatividad de los investigadores y la asesoría en comunicación de la ciencia permiten diseñar gráficos y esquemas que permiten diseminar los hallazgos de una forma más simple.

Figura 2. Patrón de visualización de las respuestas individuales en los participantes de dos grupos experimentales.

Grupo Control					
Participante	Fuerza	Potencia	Velocidad	Vigor	Fatiga
1	Verde	Verde	Amarillo	Verde	Rojo
2	Rojo	Rojo	Amarillo	Amarillo	Rojo
3	Rojo	Rojo	Amarillo	Amarillo	Verde
4	Amarillo	Rojo	Amarillo	Rojo	Verde
5	Amarillo	Rojo	Amarillo	Rojo	Amarillo
6	Verde	Verde	Verde	Verde	Verde
7	Amarillo	Verde	Verde	Amarillo	Verde
8	Rojo	Rojo	Rojo	Rojo	Rojo
9	Verde	Amarillo	Amarillo	Amarillo	Amarillo
10	Amarillo	Amarillo	Amarillo	Amarillo	Amarillo

Grupo Experimental					
Participante	Fuerza	Potencia	Velocidad	Vigor	Fatiga
1	Verde	Rojo	Verde	Verde	Verde
2	Verde	Amarillo	Verde	Verde	Verde
3	Verde	Verde	Verde	Verde	Verde
4	Verde	Verde	Verde	Verde	Rojo
5	Verde	Verde	Amarillo	Verde	Verde
6	Verde	Verde	Verde	Verde	Verde
7	Verde	Verde	Verde	Verde	Amarillo
8	Verde	Verde	Verde	Verde	Verde
9	Amarillo	Verde	Verde	Verde	Verde
10	Amarillo	Rojo	Verde	Verde	Verde

 Respuesta positiva
 Respuesta negativa
 Respuesta neutra (no hay cambio)

3.2. Diseños de investigación con alto potencial

En las Ciencias del Movimiento Humano se utilizan diseños de investigación que permiten responder a preguntas planteadas por los investigadores. Existen libros de texto que facilitan conocer cuál es el diseño más recomendando y robusto para responder a una legítima pregunta de investigación^{[138][334][398][694]}. Esos diseños se seleccionan tratando de encontrar un equilibrio entre lo "ideal" y lo "real"; es decir, entre lo que idealmente tendríamos que hacer pero que en la realidad somos capaces de hacer. Se debe recordar que hay países muy desarrollados y otros países que no han alcanzado ese desarrollo en ciencia y tecnología; sin embargo, las ideas no poseen fronteras y son válidas independientemente del origen de ésta.

Dependiendo del objetivo de la investigación, se podrían seleccionar diseños de investigación descriptivos, correlacionales, pre experimentales o experimentales^[138]. Sin embargo, actualmente se recomienda considerar otros diseños, algunos olvidados o muy poco utilizados; otros, novedosos y vanguardistas, para responder incluso a las mismas preguntas de investigación que se han venido preguntando miles de investigadores a través de los años. Por ejemplo, hoy en día se recomienda realizar estudios de viabilidad antes de invertir tiempo y recursos en tratar de responder a una pregunta de investigación. Los estudios de viabilidad responden a la pregunta de si algo se puede hacer o no, o de si el investigador debe proceder y, de ser así, cómo debería hacerlo^[139].

De forma similar, un estudio piloto hace las mismas preguntas, pero además tiene una característica de diseño específica; reproduce un estudio futuro o parte de uno realizado, pero a una mucho menor escala. Los estudios piloto pueden ser internos o externos. En el primer caso, simplemente se ejecuta la primera parte de la prueba principal. En este tipo de estudio el diseño es fijo y se toman decisiones (e.g., después de un año) acerca de si es posible el reclutamiento para la prueba completa en el tiempo propuesto. Por su parte, en un estudio piloto externo no se espera que los sujetos se incluyan en el estudio principal, y además se pueden realizar cambios más radicales en el diseño, como cambiar la variable dependiente de interés^[139].

Actualmente, existen guías para asegurarse de un adecuado diseño de estudios de factibilidad y estudios piloto. Por ejemplo, las guías CONSORT^[247], presenta una lista de verificación que se aplica a cualquier estudio en el que se realice un estudio aleatorizado controlado (en inglés, **R**andomized **C**ontrolled **T**rial, RCT) definitivo en el futuro, o al menos una parte de éste. Presenta una escala más pequeña, independientemente del diseño del estudio (e.g., grupos, factorial, cruzado) o la terminología utilizada por los autores para describir el estudio (e.g., piloto, viabilidad, ensayo, estudio).

Esta lista no aplica directamente a los estudios piloto internos integrados en el diseño de un ensayo principal, los estudios piloto no aleatorios y los estudios de viabilidad o los estudios clínicos de fase II, pero todos estos estudios tienen algunas similitudes con los estudios piloto y de viabilidad aleatorios y muchos de los criterios también podrían aplicarse. En síntesis, los estudios de viabilidad incluyen todo el trabajo preliminar antes de realizar formalmente un estudio principal, y los estudios piloto son un subconjunto de éste.

Otro diseño de investigación actualmente recomendado en las áreas de la psicología del ejercicio y del deporte es el diseño "N de 1"^[412]. Este diseño se fundamenta en que el cambio de comportamiento, siguiendo el modelo científico "tradicional", que supone que las intervenciones y tratamientos de cambio de comportamiento funcionan de manera similar en todas las personas (i.e., los investigadores calculan un efecto promedio entre los participantes), no es el correcto porque los individuos no pueden considerarse como intercambiables; es decir, la identificación de la causa y el efecto del comportamiento en una persona no se puede aplicar a todas las personas ya que existen diferencias individuales. De esta forma, el modelo científico tradicional no representa verdaderamente un enfoque personalizado o específico para un individuo^[412].

El diseño de investigación epidemiológico es otro de los múltiples tipos de diseños que se utilizan en las Ciencias del Movimiento Humano. Este tipo de diseños pueden ser principalmente prospectivos o retrospectivos. Incluso, diversas organizaciones recomiendan seguir una serie de pasos para proponer, recolectar y analizar datos de origen epidemiológico. Por ejemplo, el Comité Olímpico Internacional propuso una serie de métodos estándares para registrar y reportar datos sobre lesiones y enfermedades en el deporte con el propósito de monitorear la salud de los atletas de alto nivel^[720]. Este tipo de iniciativas permite que los investigadores reporten un mínimo de criterios para poder ser comparados entre diversos deportes y atletas particulares.

Finalmente, dentro de los ejemplos de diseños poco utilizados o casi olvidados, se encuentran los de no inferioridad^[61], en los que se pretende demostrar que un nuevo tipo de intervención, como lo podría ser una novedosa forma de entrenar, no es inferior que una forma de entrenar tradicional de la que ya se conocen sus efectos. Este tipo de diseños representa el opuesto de los tradicionales o más comúnmente utilizados diseños de superioridad, en los que se pretende conocer si la nueva intervención es superior a un grupo de control o a un grupo de comparación. En las siguientes secciones se presentarán y discutirán algunos de estos diseños.

3.2.1. Estudios de factibilidad, piloto y prueba de concepto

Los estudios de factibilidad o viabilidad responden a la pregunta de si se puede hacer el estudio, de si los investigadores deberían realizarlo, y si fuera así, ¿cómo deberían hacerlo?^[139]. De esta forma se puede afirmar que un estudio de factibilidad incluye todo el trabajo de planificación y logística preliminar antes de realizar el estudio principal. Se considera que los estudios piloto son una categoría de los estudios de factibilidad^[139].

El propósito principal de un estudio piloto no es responder preguntas de investigación específicas, sino evitar que los investigadores inicien un estudio a gran escala sin un conocimiento adecuado de los métodos propuestos^[448]. Un estudio piloto se pregunta si el futuro estudio es viable; sin embargo, la diferencia con un estudio de viabilidad radica en que su diseño reproduce un estudio futuro o parte de uno realizado, pero a mucho menor escala^[139]. Así, se considera que un estudio piloto es un prólogo metodológico de un estudio de mayor tamaño^[268]. Los estudios piloto internos son la primera parte del estudio de mayor escala, y se supone que el diseño es fijo y que se pueden tomar decisiones después de un tiempo (e.g., un año) acerca de si es posible el reclutamiento para el estudio completo. Por otra parte, en los estudios piloto externos, no se espera que los participantes se incluyan en el estudio principal, y, además, se pueden realizar cambios más radicales en el diseño (e.g., cambiar la variable de interés principal)^[139].

115

Es así como los estudios piloto permiten a los investigadores probar métodos (e.g., tamaño de muestra, estrategia de reclutamiento y retención de los participantes, recolección y almacenaje de muestras biológicas, entrenamiento del personal, obtención del consentimiento informado) antes de comprometerse con la fidelidad requerida en estudios más amplios que requieren muchos recursos^{[416][448]}. Otro aspecto relevante se refiere al análisis de los datos, en donde se recomienda obtener las estadísticas descriptivas y estimaciones de los intervalos de confianza, que posteriormente pueden utilizarse para estimar los cambios esperables o el tamaño de la muestra para el estudio principal^[416]. Finalmente, los estudios piloto también pueden mostrar la reacción de los participantes a los procedimientos de medición de las variables de interés, el tiempo real necesario para recolectar datos, la voluntad de los participantes para seguir el protocolo del estudio y posible variación no anticipada en la respuesta^[268].

Para Foster^[268], un estudio piloto no es sinónimo de un tamaño de muestra pequeño, porque existen estudios piloto con muestras mínimas de 10 personas y hasta de más de 40 participantes^{[183][739]}. Se supone que un estudio piloto bien planificado y ejecutado puede identificar posibles variables confusoras que no se conocían previamente y evaluar la fuerza de las relaciones entre las variables clave para ayudar en el cálculo del tamaño de la muestra^[448]. Algunos autores utilizan el término "piloto" cuando publican estudios completos con tamaños de muestra pequeños para desviar la atención de los revisores o editores de revistas, lo cual es inapropiado. El uso del término "piloto" es inapropiado a menos que el estudio haya sido diseñado para probar métodos de investigación y que se reporten los resultados en términos de la viabilidad de los métodos examinados. También es importante señalar que la potencia estadística de los estudios piloto siempre es menor que la de un estudio confirmatorio, ya que en principio lo que interesa es medir y evaluar otros aspectos de la posible investigación^[448]. Así, los resultados del estudio piloto son siempre preliminares (i.e., no irrefutables) y se tienen que interpretar con la mayor precaución posible^[416].

Para facilitar su identificación en las bases de datos y buscadores en línea, se recomienda que los estudios de este tipo deban identificarse claramente utilizando las palabras “viabilidad”, “factibilidad” o “piloto”^[139]. Además, se recomienda que los investigadores reporten los objetivos y métodos relacionados con la factibilidad, así como si el estudio apenas se está preparando para un futuro ensayo controlado aleatorio. En la tabla 1 se muestran ejemplos de estudios piloto en las Ciencias del Movimiento Humano que cumplen esas recomendaciones.

Tabla 1. Ejemplos de estudios de factibilidad y estudios piloto recientes reportados en la literatura científica.

Autores	Título	Objetivo(s) o propósito(s)
Estudios de Factibilidad		
Mattle et al. ^[471]	Viabilidad de la hipnosis sobre el rendimiento en la competencia de tiro con rifle de aire.	Investigar el efecto de una sola intervención de hipnosis sobre el rendimiento de tiro en una competencia de rifle de aire comprimido a una distancia de 10 m.
Metcalf et al. ^[490]	Entrenamiento de ejercicios de intervalos de velocidad guiados por computadora y eficiencia en el tiempo para mejorar la salud en el lugar de trabajo: un estudio de viabilidad de métodos mixtos aleatorios en empleados de oficina.	El objetivo de este estudio fue utilizar métodos mixtos para investigar la viabilidad, aceptabilidad y efectividad de una intervención de ejercicio de alta intensidad y corta duración cuando se aplica sin supervisión en un entorno laboral.
Voorn et al. ^[731]	Ejercicio aeróbico individualizado en enfermedades neuromusculares: un estudio piloto sobre la factibilidad y la eficacia preliminar para mejorar la condición física.	Evaluar la factibilidad y la eficacia preliminar de un programa de ejercicio aeróbico individualizado de acuerdo con una guía de entrenamiento recientemente desarrollada (B-FIT) para mejorar la aptitud física en personas con enfermedades neuromusculares.
Estudios Piloto		
Corbett et al. ^[183]	Intervención piloto de ejercicio aeróbico para jóvenes en riesgo de enfermedad mental grave.	<ol style="list-style-type: none"> 1. Examinar la viabilidad de una intervención de ejercicio. 2. Determinar qué mejoría se observó después de la participación en un programa de ejercicio aeróbico de intensidad moderada a alta.
Daly et al. ^[194]	Factibilidad, uso y disfrute de un programa de ejercicios en el hogar transmitido a través de una aplicación de ejercicios para la salud musculoesquelética en adultos mayores que viven en la comunidad: estudio piloto prospectivo a corto plazo.	Evaluar la factibilidad, el uso y el disfrute de una aplicación de prescripción de ejercicios basada en la web como una plataforma para que los profesionales del ejercicio realicen y supervisen de forma remota un programa de ejercicios multicomponente casero y personalizado (transmitido a través de tabletas) a los adultos mayores que viven de forma independiente en la comunidad.
		Continúa...

Tabla 1. Continuación.

Autores	Título	Objetivo(s) o propósito(s)
DeNysschen et al. ^[204]	Ejercicio con educación nutricional para mejorar la calidad de vida de adolescentes y adultos jóvenes sobrevivientes de cáncer: un estudio piloto.	Determinar si 8 semanas de educación sobre ejercicio y nutrición en adolescentes y adultos jóvenes sobrevivientes de cáncer afecta positivamente la fatiga, la calidad de vida, el estado físico relacionado con la salud y la ingesta dietética.
Ekrami et al. ^[245]	Las ventosas húmedas inducen una acción antiinflamatoria en respuesta al ejercicio vigoroso en atletas de artes marciales: un estudio piloto.	Investigar los posibles efectos antiinflamatorios de las ventosas húmedas antes de una prueba de ejercicio de moderada a vigorosa intensidad en atletas de artes marciales.
Higuti et al. ^[340]	Efectos de escuchar música y practicar ejercicio físico sobre aspectos funcionales y cognitivos en adultos mayores con demencia institucionalizados: Estudio piloto.	Determinar los efectos de escuchar música y practicar ejercicio físico sobre aspectos funcionales y cognitivos en adultos mayores con demencia institucionalizados.
Porrati-Paladino & Cuesta-Barriuso ^[576]	Eficacia del ejercicio pliométrico y excéntrico para el salto y la estabilidad en jugadoras de fútbol: un estudio piloto controlado aleatorio, ciego simple.	Verificar la efectividad de un programa de entrenamiento pliométrico combinado con ejercicios excéntricos, en comparación con el entrenamiento excéntrico solo, para mejorar la estabilidad de las extremidades inferiores y la capacidad de salto en jugadoras de fútbol federadas de entre 18 y 30 años.
Washabaugh et al. ^[739]	El entrenamiento de resistencia funcional altera diferencialmente la cinética de la marcha después de la reconstrucción del ligamento cruzado anterior: un estudio piloto.	Probar cómo 3 tipos de dispositivos de entrenamiento de resistencia funcional (una rodillera que resiste la extensión o abrazadera unidireccional), una rodillera que resiste la extensión y la flexión (abrazadera bidireccional) y una banda elástica que tira hacia atrás del tobillo (banda elástica) alteran de forma aguda la cinética de la marcha en individuos lesionados.

Los estudios de prueba de concepto, prueba de principio o en inglés Proof-of-concept (PoC), consisten en probar un determinado método o idea para demostrar su viabilidad o factibilidad. También pueden consistir en una demostración para verificar que algún concepto o teoría tiene potencial práctico. Una prueba de concepto suele ser pequeña y puede completarse o no.

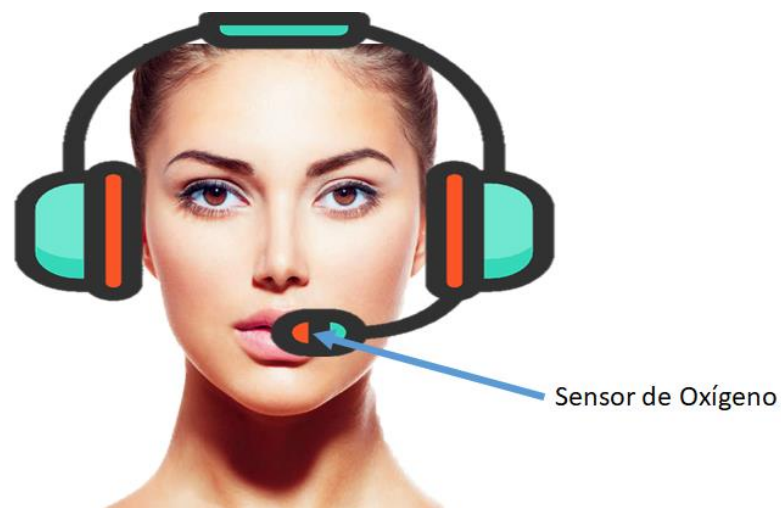
Los estudios de PoC se pueden diseñar para probar nuevas ideas y determinar si los hallazgos preliminares pueden ser prometedores para un estudio a mayor escala. Por ejemplo, los estudios de PoC se pueden diseñar para probar la viabilidad de un producto nutricional novedoso para la salud o el rendimiento deportivo, o incluso para comprobar si algún fundamento fisiológico puede funcionar en poblaciones con alguna patología^[259]. En el caso de un producto, en teoría y debido a sus características bioquímicas, podría beneficiar a las personas, pero no se sabe su efecto más allá de las pruebas preliminares realizadas en modelos animales. Por esa razón, antes de realizar

un estudio a mayor escala (i.e., muchos participantes) en seres humanos, lo conveniente es probar el concepto (i.e., de que el producto puede beneficiar) a menor escala; es decir, con menos participantes, pero cumpliendo la metodología que se seguiría en un estudio principal. Una vez realizado el estudio de PoC se puede valorar la evidencia preliminar para tomar decisiones acerca de si se realizan ajustes al protocolo.

Existen otros ambientes en los que se utilizan los estudios PoC. Por ejemplo, algunos desarrolladores de aplicaciones para teléfonos inteligentes a menudo programan fórmulas matemáticas con la finalidad de proporcionar a los usuarios algún tipo de información para su entrenamiento o monitoreo de actividad física. Sin embargo, a pesar de que la base teórica o fundamento científico sean pertinentes, la aplicación en condiciones de la vida real tal vez no lo sean^[659], y por tal razón, se aconseja realizar estudios de PoC que validen las aplicaciones. Así, se genera evidencia preliminar, la cual obviamente no es definitiva, sino que se requiere que sea consistentemente replicada para poder llegar a conclusiones robustas y definitivas con mayor seguridad.

De manera similar, las compañías que desarrollan productos para medir alguna variable fisiológica primero llevan a cabo estudios de PoC^[615]. La compañía alemana Bayern Innovative (Bavarian State Ministry of Economics and VitaScale GmbH, Nuremberg) estudió un dispositivo portátil para medir el VO_2 máx^[234]. La compañía fundamentó su dispositivo en la facilidad de transporte al compararlos con los dispositivos que requieren de una máscara que cubra la nariz y la boca de la persona, lo cual es incómodo para el usuario. De esta forma, se unieron a la compañía VitaScale, que previamente había diseñado un auricular miniaturizado, para probar si el dispositivo podía medir el VO_2 máx sin usar una máscara facial (Figura 1). Así, estudiaron a 44 personas durante ejercicio realizado en una bicicleta ergométrica y determinaron la confiabilidad y la validez de criterio del dispositivo. Los investigadores encontraron que el coeficiente de variación al medir el VO_2 máx durante la prueba incremental con el auricular fue de 6.8%, y la validez de criterio tuvo un coeficiente de variación del 4.0%, por lo que concluyeron que los auriculares podrían ofrecer una nueva tecnología para la medición del VO_2 máx debido a su diseño de bajo costo y sin máscara.

Figura 1. Esquema de un dispositivo portátil para medir el consumo máximo de oxígeno sin necesidad de una máscara. El diseño original se puede observar en los estudios de Roossien et al.^[615] y Düking et al.^[234].



Otro entorno en el que se desarrollan estudios de PoC relacionados con las Ciencias del Movimiento Humano se encuentra en el ámbito clínico. Por ejemplo, Myrcik et al.^[522], realizaron un estudio en personas que recibían cuidados paliativos para determinar si el ejercicio físico les permitía reducir el dolor, la depresión y mejorar la calidad de vida. A pesar de que se cuenta con una extensa cantidad de estudios en poblaciones sanas, los investigadores prefirieron primero realizar un estudio de PoC para obtener evidencia preliminar acerca de su posible utilidad, lo cual desde el punto de vista ético es muy positivo y relevante. Los hallazgos provenientes de este tipo de estudios deben comprenderse e interpretarse con cautela, ya que son apenas evidencias preliminares, y el siguiente paso lógico es realizar estudios clínicos aleatorizados controlados a mayor escala para ir creando un cuerpo de conocimientos consistente^[522].

En el ámbito del entrenamiento físico también se han realizado estudios PoC. Por ejemplo, se llevó a cabo un programa de entrenamiento de pies para observar sus efectos en la morfología y la fuerza muscular, así como la biomecánica de carrera en corredores recreativos sanos^[682]. Para realizar el estudio, los investigadores asignaron a 28 corredores a un grupo de control (GC) o de intervención (IN). La IN consistió en fortalecer los músculos intrínsecos del pie y su activación durante las actividades de carga de peso. Los resultados del estudio indicaron que hubo un aumento en el volumen de todos los músculos investigados y en el impulso de propulsión vertical durante la carrera en el grupo IG en comparación con el GC.

En síntesis, los estudios de viabilidad, factibilidad, piloto y PoC permiten a los investigadores planificar los recursos, logística, presupuesto y cantidad de participantes que necesitarían para un estudio completo. Además posibilitan identificar variables extrañas y variables no consideradas inicialmente en la propuesta de una investigación. Esos elementos facilitan llevar a cabo investigaciones que maximicen la utilización de los recursos para generar investigaciones de mayor calidad. Se vislumbra que en las diferentes áreas de las Ciencias del Movimiento Humano cada vez aparezcan publicadas más investigaciones de estos tipos.

3.2.2. Dosis-Respuesta

En la investigación acerca de las diferentes aristas que componen las Ciencias del Movimiento Humano, es necesario conocer el efecto de los programas de intervención en los que se utilice el ejercicio físico. Este es un problema de causa-efecto; es decir, el ejercicio o el entrenamiento es la "causa", mientras que el "efecto" es el cambio (o no) en la variable de interés para los investigadores. Esta forma de evaluar un programa se le conoce como contrafáctica^[605], en la que un efecto del programa es la diferencia entre lo que sucedió después de que se implementó el programa y lo que hubiera sucedido si el programa no se hubiera implementado, asumiendo que todo lo demás hubiera permanecido igual.

Adicionalmente, al realizar una investigación con el propósito de definir si una intervención (i.e., prescripción de ejercicio físico, metodología de entrenamiento, programa de Educación Física y Deportes) funciona para mejorar alguna variable de interés (e.g., fuerza, velocidad, capacidad aeróbica, % de grasa corporal, cogniciones), también surge la necesidad de determinar si a una mayor dosis de la intervención (e.g., más sesiones de ejercicio, más intensidad, más duración), también se encontraría una mayor respuesta en la variable de interés (e.g., mayor fuerza, mayor velocidad, mayor capacidad aeróbica, menor % de grasa corporal, mejor cognición). A este tipo de planteamiento se le considera un problema de dosis-respuesta, y es un tipo de razonamiento que permite predecir futuras respuestas y adaptaciones en los diferentes sistemas del organismo a través de cambio anatómicos, fisiológicos o conductuales^[335]. Se podría afirmar que conociendo la dosis-respuesta se podría llegar a conocer la máxima capacidad de una persona para responder a cierto estímulo; es decir, la máxima entrenabilidad que define el límite del rendimiento del ser humano. Este es el paradigma que ha guiado el trabajo de muchos entrenadores e investigadores durante décadas.

En las Ciencias del Movimiento Humano, la "dosis" se puede definir como el producto de las variables del ejercicio, las variables de entrenamiento, así como la aplicación de los principios del entrenamiento, por lo que debe operacionalizarse mediante el uso de algún marcador específico de carga interna o externa. Los marcadores específicos deben participar en los procesos biológicos que estimulan los cambios deseados (e.g., lactato, factor neurotrófico derivado del cerebro [BDNF], cambios neurocognitivos, cambios en fuerza, potencia, velocidad). La carga interna se define como una respuesta individual aguda de tipo fisiológica, biomecánica o psicológica a un factor de influencia, como las condiciones donde se lleva a cabo el entrenamiento físico, así como al trabajo realizado, también conocido como carga externa. La carga externa, a su vez, se define como el trabajo realizado por la persona independientemente de sus características internas.

La "respuesta" a la intervención se manifiesta en aquellas personas que a través del tiempo muestran cambios en la variable de interés que sobrepasan un umbral determinado para esa variable^[335]. La determinación de ese umbral puede establecerse de manera objetiva (e.g., cambio porcentual, desviación estándar, SWC), o puede ser subjetiva (i.e., criterio de los investigadores).

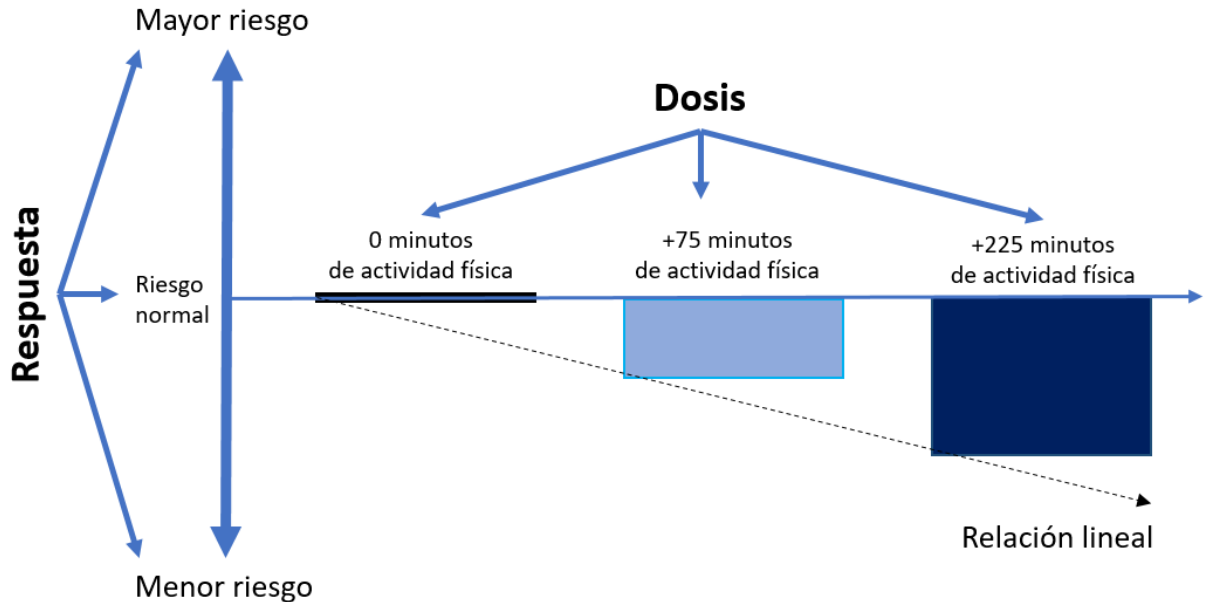
El tema de la dosis-respuesta es tan relevante hoy en día para maximizar el tiempo y los esfuerzos de las personas involucradas, que incluso existen revistas especializadas en donde se publican este tipo de investigaciones (e.g., revista "Dose-Response", editorial Sage, <https://journals.sagepub.com/home/dos>). Los estudios de entrenamiento físico también tienen acogida en esa revista; en donde, por ejemplo, se realizó una investigación para determinar si dos

tipos de entrenamiento (dosis) tenían un efecto (respuesta) en el cambio del BDNF, una proteína que actúa como factor de crecimiento de la familia de las neurotrofinas asociadas al factor de crecimiento nervioso^[281]. Los tipos de actividad física aguda (i.e., dosis) fueron una prueba gradual de ejercicio y el otro fue un entrenamiento de intervalos de alta intensidad, conocido en inglés como **H**igh-**I**ntensity **I**nterval **T**raining (HIIT), y la variable respuesta era el cambio en el BDNF. Los investigadores encontraron que independientemente de la dosis del ejercicio, los valores de BDNF no se alteraron de forma significativa, por lo que no se puede reportar un efecto de dosis-respuesta. Este tipo de estudios requiere de un seguimiento particular, es decir, intentar investigar con diferentes dosis de ejercicio para determinar cuál es la que verdaderamente produce una respuesta en la dirección teóricamente esperada. Entonces, los estudios de dosis-respuesta siempre son apreciados en las revistas científicas porque brindan mayor claridad acerca del tipo de intervención que se requiere para encontrar cambios en las variables de interés.

Por ejemplo, utilizando métodos robustos de regresión lineal múltiple, se investigó si existía una relación de dosis-respuesta entre la adherencia al ejercicio físico luego de una intervención de ejercicio de 6 meses y la respuesta producida en el control glicémico en pacientes con diabetes mellitus tipo 2^[68]. Los investigadores encontraron una relación dosis-respuesta estadísticamente significativa, en donde por cada 20% de aumento en la adherencia a tres sesiones de ejercicio prescritas por semana (e.g., de 60% a 80%, el equivalente a dos sesiones de ejercicio adicionales por mes), la hemoglobina glicosilada (HbA1c) se redujo en 0.15% ($\sim 2 \text{ mmol}\cdot\text{mol}^{-1}$; IC95%= 0.02% a 0.28%) durante 6 meses. En palabras simples, una dosis de ejercicio más alta se asoció con una mayor reducción de HbA1c. Este es un ejemplo de cómo se puede traducir a términos más comprensibles los resultados de un estudio para las personas responsables de prescribir el entrenamiento físico y así comunicarse con sus pacientes o clientes. Además, para el caso concreto del ejemplo, le permitiría al equipo de profesionales de la salud que atienden al paciente, mejorar la prescripción del ejercicio y regular mejor las dosis de los medicamentos y otras intervenciones farmacológicas o conductuales que estén utilizando con los pacientes.

Ejemplos similares se han publicado desde inicios de los años 60, en donde se ha demostrado consistentemente, que quienes realizan una mayor cantidad de actividad física semanal también tienen un menor riesgo de morir por cualquier causa^{[243][319][674]}; es decir, a mayor cantidad de actividad física acumulada durante la semana, se puede vivir más años (Figura 1). A esto se le conoce como el efecto protector del ejercicio, que ha reducido el riesgo de muerte de un 21% hasta un 77% en quienes realizan actividad física de intensidad moderada a vigorosa^[286], y aunque la relación es lineal y directa; es decir, a mayor cantidad de actividad física también hay menor probabilidad de muerte, se debe reconocer que no se sabe si un exceso de cantidad de actividad física más bien podría aumentar el riesgo de muerte, como podría ocurrir en el síndrome de muerte súbita provocada por el ejercicio^[42]. En otras palabras, solamente se ha investigado una parte de la asociación, cuando existe la posibilidad de que la asociación pueda ser curvilínea (i.e., no lineal, sino como una U invertida "∩" o con forma de J o S); es decir, que exista una cantidad de actividad física que sea positiva, pero que a partir de esa cantidad más bien sea perjudicial para la salud^[42].

Figura 1. Esquema teórico basado en la evidencia científica acerca de una relación dosis-respuesta para la actividad física y el riesgo de muerte por cualquier causa. La cantidad de minutos de actividad física es la dosis (prescripción, eje "x") y el riesgo de muerte es la respuesta (eje "y"). En el gráfico se observa que a mayor dosis hay un menor riesgo de muerte en una relación lineal.



Otra parte de la asociación dosis-respuesta del ejercicio físico y el riesgo de mortalidad se ha investigado recientemente para conocer si la actividad física de intensidad vigorosa o moderada reduce la mortalidad en la misma medida^[608]. Los investigadores realizaron un meta análisis y encontraron que los estudios prospectivos sugieren que, para la misma actividad física total, tanto las actividades físicas de intensidad vigorosa como las de intensidad moderada reducen la mortalidad por cualquier causa en la misma medida. De esta forma, la dosis-respuesta es la misma para reducir la mortalidad, independientemente de si las actividades físicas se realizan a intensidades moderadas o vigorosas.

Para enfermedades como la hipertensión arterial, se han realizado estudios acerca de la relación dosis-respuesta en los que han utilizado una combinación de técnicas de análisis estadístico avanzadas para conocer la relación de dosis-respuesta entre el nivel inicial de capacidad cardiorrespiratoria y el riesgo de hipertensión^[165]. Para analizar los datos, se calculó la reducción del riesgo para el aumento de 1-MET de capacidad cardiorrespiratoria. Se utilizó la regresión de mínimos cuadrados generalizados para calcular la dosis-respuesta específica del estudio. También se utilizó el modelo de efectos aleatorios de DerSimonian y Laird para resumir las estimaciones del riesgo relativo (RR) de dosis-respuesta específicas del estudio. Luego, se modeló la dosis-respuesta utilizando el "spline cúbico restringido", con tres configuraciones de nodos al 25%, 50% y 75% de la distribución de la capacidad cardiorrespiratoria, para capturar la naturaleza detallada de la asociación. Luego se utilizó el método de Greenland para estimar las pendientes específicas del estudio y los IC95% a partir de los logaritmos naturales de las estimaciones de riesgo con sus errores estándar correspondientes. Así, los investigadores encontraron que el RR de hipertensión fue de 0.71 (IC95%= 0.64, 0.79) para los participantes con

capacidad cardiorrespiratoria mayor en comparación con aquellos cuya capacidad cardiorrespiratoria disminuyó con el tiempo, lo que permite concluir que existe una relación negativa o inversa dependiente de la dosis entre la capacidad cardiorrespiratoria y el riesgo de hipertensión. La reducción del riesgo de hipertensión ocurre por cada incremento de 1 MET en la capacidad cardiorrespiratoria^[165].

El tema de si la forma de la asociación dosis-respuesta es lineal o no lineal (i.e., curvilínea) se ha investigado en niños^[528] y en adultos^[252]. Por ejemplo, estudios previos en niños habían asumido que la asociación entre la aptitud aeróbica y la actividad física era lineal. Los investigadores determinaron la asociación dosis-respuesta entre la actividad física y la aptitud aeróbica para evaluar si esta asociación era lineal o curvilínea y si variaba de acuerdo con el sexo, la edad y el estado de peso de los niños. Para ello, aplicaron técnicas robustas de regresión multinivel a un grupo de datos longitudinales recopilados en niños de 12, 14 y 16 años. El modelado multinivel es una extensión de la regresión múltiple ordinaria donde los datos tienen una estructura jerárquica o agrupada. Una jerarquía consta de unidades o medidas agrupadas en diferentes niveles^[600]. El análisis de los datos identificó una asociación significativa en forma de "U" invertida entre el VO_2 máx como medida de aptitud aeróbica y el nivel de actividad física medida a través de un cuestionario auto administrado^[528].

En población adulta, recientemente se estudió la asociación entre la actividad física y el riesgo de infección por SARS-CoV-2, hospitalización asociada a COVID-19, enfermedad grave y muerte por COVID-19 en adultos^[252]. Para determinar si la asociación era lineal o no lineal, se realizó una revisión sistemática y meta análisis con datos recolectados hasta marzo de 2022, en el que se resumieron los datos de 16 estudios y 1 853 610 personas. Los investigadores encontraron que quienes realizaban actividad física regular tenían un menor riesgo de infección (RR= 0.89; IC95%= 0.84 a 0.95), hospitalización (RR= 0.64; IC95%= 0.54 a 0.76), enfermedad grave (RR= 0.66; IC95% = 0.58 a 0.77) y muerte relacionada con COVID-19 (RR= 0.57; IC95%= 0.46 a 0.71) en comparación con personas inactivas. Además, se encontró que esos hallazgos seguían una relación dosis-respuesta no lineal entre la actividad física reportada en MET-min/semana y enfermedad grave y muerte por COVID-19. Esto significa que se observó un aplanamiento en la curva que describe la relación de dosis-respuesta alrededor de los 500 MET-min/semana. Esto quiere decir que a valores mayores a ese umbral, la relación ya no es lineal. Es así, como nuevamente los métodos estadísticos robustos basados en regresión permiten determinar una asociación dosis-respuesta que inicialmente se consideraba que era lineal, pero que resultó ser curvilínea. Seguramente existen otros casos en los que sería oportuno reanalizar los datos en busca de conclusiones verdaderas.

En ambientes clínicos donde intervienen los profesionales en Ciencias del Movimiento Humano, también es relevante poder determinar los puntajes o valores a partir de los cuales se pueden lograr cambios en variables de interés, como, por ejemplo, el perfil lipídico y la presión arterial. En un estudio realizado en niños y adolescentes con obesidad^[246], se determinó a través de una meta regresión, cuáles eran los valores mínimos de un puntaje estandarizado del índice de masa corporal (IMC-SDS), que eran necesarios para mejorar el perfil lipídico y los valores de la presión arterial. Los investigadores encontraron que reducciones del IMC-SDS >1 disminuían la presión arterial sistólica, >1.2 el colesterol LDL, y >0.7 la concentración de triglicéridos.

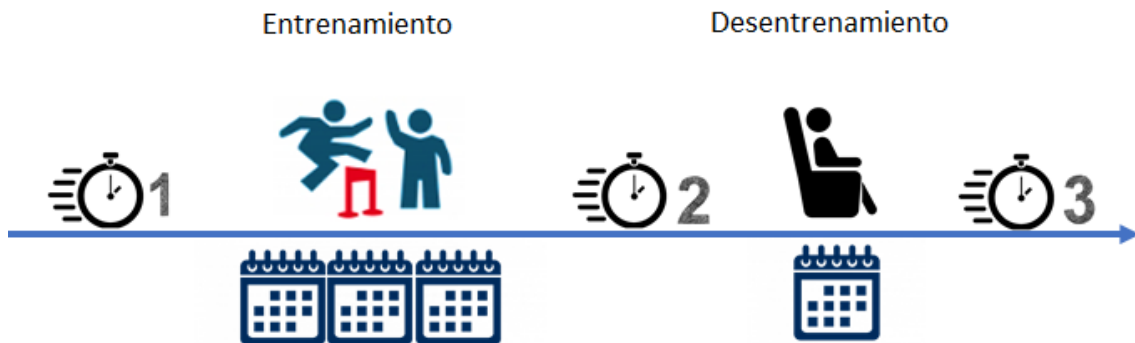
En síntesis, la enseñanza de ese meta análisis con meta regresión radica en que independientemente de la estatura de los niños y adolescentes, que se encuentran en pleno crecimiento y desarrollo, es necesario que logren mantener un peso saludable, de manera que el índice IMC-SDS pueda reducirse en al menos 0.7 unidades para mantener un perfil lipídico y de presión arterial sistólica saludable.

Finalmente, debe considerarse seriamente el tema de si la dosis de ejercicio o entrenamiento alcanza un umbral suficiente para provocar un cambio importante y significativo en una variable de interés. Por ejemplo, personas a las que inicialmente un tratamiento de entrenamiento físico no les produjo cambios en la capacidad cardiorrespiratoria, pero que fueron nuevamente entrenados utilizando una mayor dosis de ejercicio, si lograron cambiar sus valores de una forma positiva^[511]. Esto demuestra que metodológicamente se debe seleccionar la supuesta dosis que teóricamente provocaría un cambio importante en la variable de interés, y adicionalmente, estudiar el posible efecto de reducir o de aumentar esa dosis.

3.2.3. Entrenamiento-Desentrenamiento

Los estudios de entrenamiento-desentrenamiento pretenden analizar la eficacia de una intervención aplicada a una muestra con características específicas y seguidamente estudiar las consecuencias de su suspensión en las variables dependientes de interés (Figura 1). En este contexto, es de especial importancia la conceptualización del desentrenamiento, que es descrito por Araujo et al.^[31] como la interrupción de una actividad que se venía realizando, con el consecuente incremento en el riesgo de que el organismo pierda, total o parcialmente, las adaptaciones obtenidas al haber realizado el programa de ejercicios. En este ámbito de interés, los estudios se realizan principalmente utilizando dos enfoques distintos. Un enfoque se dirige en mejorar la salud y el bienestar de las personas que realizan ejercicio^{[31][60][77][136][200][265][397]}. El otro enfoque corresponde al rendimiento deportivo de las personas que se preparan para competir^{[162][537][599]}.

Figura 1. Paradigma de los estudios de entrenamiento y desentrenamiento. En la primera etapa los participantes se someten a mediciones de las variables de interés. Posteriormente completan un programa de entrenamiento físico de determinada cantidad de semanas. Al completar el programa, se miden nuevamente las variables de interés para determinar el cambio atribuido al entrenamiento. Posteriormente, los participantes ya no reciben el estímulo del entrenamiento, y, después de otro período de tiempo, nuevamente se miden las variables de interés. Con este paradigma se puede conocer y cuantificar el retroceso en las adaptaciones ganadas con el entrenamiento en los distintos sistemas del organismo humano.



Entre los diferentes autores interesados en analizar el binomio ejercicio-salud, el objetivo de los estudios se dirige a demostrar los beneficios del ejercicio, ya sea para contrarrestar los efectos generados por el padecimiento de determinadas patologías, o para incrementar la independencia y en general la calidad de vida de las personas. Además, reconociendo que en muchas ocasiones por diversos factores voluntarios o involuntarios se tiene que interrumpir la práctica del programa de ejercicio, se integra el análisis del efecto de esta suspensión para determinar si los beneficios iniciales se mantienen; es decir, si las adaptaciones anatómicas y fisiológicas perduran y durante cuánto tiempo. De esta manera, los estudios varían según la población de interés incluida, las variables dependientes estudiadas, y las características de la fase de entrenamiento y desentrenamiento. En la tabla 1 se puede apreciar el manejo de estos elementos en diferentes estudios publicados recientemente.

Tabla 1. Ejemplos de investigaciones que utilizan el paradigma de entrenamiento-desentrenamiento.

Estudio	Población	Variables medidas	Entrenamiento aplicado	Duración del desentrenamiento	Principal resultado
Araujo et al. ^[31]	VIH/SIDA.	<ul style="list-style-type: none"> Composición corporal. Acondicionamiento cardiorrespiratorio. Fuerza del tren superior e inferior. 	<ul style="list-style-type: none"> 15 semanas. Combinación de ejercicio aeróbico y contra resistencia. 	5 semanas	<ul style="list-style-type: none"> El entrenamiento genera beneficios similares en personas con y sin VIH/SIDA. Se presenta una mayor pérdida de adaptaciones en población con VIH/SIDA.
Batrakoulis et al. ^[60]	Mujeres sedentarias con sobrepeso y obesidad.	<ul style="list-style-type: none"> Fuerza muscular. Resistencia muscular. Flexibilidad. Balance estático. Movimiento funcional. Densidad mineral. 	<ul style="list-style-type: none"> 10 meses. Circuitos con ejercicios de fuerza y actividades de la vida diaria. 	5 meses	<ul style="list-style-type: none"> Programa de ejercicio mejoró el acondicionamiento musculoesquelético. Con el desentrenamiento las adaptaciones se reducen, pero no se pierden.
Biduski et al. ^[77]	Mujeres adultas mayores.	<ul style="list-style-type: none"> Capacidad funcional. Rendimiento neuromuscular de los músculos extensores de la rodilla. 	<ul style="list-style-type: none"> 12 semanas. Pilates sobre un mat. 	6 semanas	<ul style="list-style-type: none"> En general entrenamiento mejoró todas las variables. El desentrenamiento genera pérdida de ganancias en algunas variables, en otras se mantienen los beneficios.
					Continúa...

Tabla 1. Continuación.

Estudio	Población	Variables medidas	Entrenamiento aplicado	Duración del desentrenamiento	Principal resultado
Callahan et al. ^[136]	Hombres sedentarios de mediana edad.	<ul style="list-style-type: none"> • Características de las fibras musculares. • Composición corporal. • Grosor músculo. • Fuerza muscular. • Capacidad aeróbica. • Metabolismo basal. • Homeóstasis glucosa. 	<ul style="list-style-type: none"> • 6 semanas. • Ciclismo de resistencia. • Ciclismo de alta intensidad. • Ejercicios contra resistencia. 	2.5 semanas	<ul style="list-style-type: none"> • En general los entrenamientos mejoraron todas las variables. • Con el desentrenamiento las ganancias en características y fuerza muscular se mantienen y en capacidad aeróbica se pierden.
de Jesus Leite et al. ^[200]	Sobrevivientes de cáncer de mama.	<ul style="list-style-type: none"> • Composición corporal. • Fuerza muscular. • Perfil lípidos. • Marcadores inflamatorios y oxidativos. 	<ul style="list-style-type: none"> • 12 semanas • Entrenamiento contra resistencia 	12 semanas	<ul style="list-style-type: none"> • Ejercicio mejoró las diferentes variables. • Desentrenamiento no provocó cambios en las mejoras.
					Continúa...

Tabla 1. Continuación.

Estudio	Población	Variables medidas	Entrenamiento aplicado	Duración del desentrenamiento	Principal resultado
Filho et al. ^[265]	Mujeres adultas mayores.	<ul style="list-style-type: none"> • Capacidad funcional. • Fuerza muscular. • Potencia muscular. 	<ul style="list-style-type: none"> • 20 semanas. • Variantes de entrenamiento contra resistencia. 	4 semanas	<ul style="list-style-type: none"> • Diferentes tipos de entrenamiento mejoraron las variables. • Beneficios en la capacidad funcional se pierden al dejar de entrenar.
Kemmler et al. ^[397]	Hombres adultos mayores con osteosarcopenia.	<ul style="list-style-type: none"> • Masa muscular. • Densidad mineral ósea. • Fuerza muscular. • Velocidad caminata. 	<ul style="list-style-type: none"> • 18 meses. • Entrenamiento contra resistencia de alta intensidad. 	6 meses	<ul style="list-style-type: none"> • Ejercicio mejoró las diferentes variables. • Con desentrenamiento los beneficios se pierden.

De acuerdo con el diseño del estudio seguido por los diferentes autores, en ocasiones se elige estudiar un solo grupo, que es medido a través de una línea base, un período de entrenamiento y uno de desentrenamiento^{[77][200]}. O bien, se aplica un diseño en el que existen varios grupos de comparación evaluados durante el transcurso del estudio^{[31][60][265][397]}.

Una aplicación totalmente práctica de este tipo de diseños de estudio ocurre en el ámbito deportivo, ya que es reconocida la necesidad de reducir las cargas de entrenamiento en ciertos momentos de la preparación, así como la incidencia de lesiones que obligan al deportista a detener su proceso de entrenamiento normal. Consecuentemente, el análisis de las características del desentrenamiento y la cuantificación de las posibles pérdidas de las cualidades físicas o motrices, son relevantes para optimizar el desempeño deportivo. De esta manera, Chen et al.^[162], propusieron investigar el efecto de 2 semanas de desentrenamiento en la función cardiopulmonar y el acondicionamiento muscular de atletas de resistencia, concluyendo que el VO_2 máx, el tiempo de ejercicio hasta el cansancio, el volumen sistólico y la fuerza de los músculos extensores de la rodilla se reducen con la interrupción completa del entrenamiento, mientras que los valores de fuerza en los músculos flexores de la rodilla, la resistencia muscular y el porcentaje de grasa no se ven afectados después de dos semanas de desentrenamiento.

En esta misma línea de investigación, Nunes et al.^[537], demostraron que un entrenamiento neuromuscular integrado (i.e., combinación de actividades generales y específicas, entrenamiento de resistencia, estabilidad dinámica, pliometría y agilidad) de 12 semanas es capaz de mejorar la altura del salto vertical contra movimiento de jóvenes practicantes de voleibol, y que luego de 8 semanas de desentrenamiento las mejoras iniciales se logran mantener. Es decir, el desentrenamiento no afectó negativamente a los participantes, por lo que sus adaptaciones al entrenamiento se mantuvieron en el tiempo.

Räntilä et al.^[599], destacan la presencia de importantes diferencias individuales en las adaptaciones percibidas por los deportistas ante un mismo entrenamiento, por lo que consecuentemente, la posible pérdida de los beneficios generados por el desentrenamiento también debe ser analizada de manera individual. Bajo esta visión, se sometió a un grupo de 26 hombres activos a un entrenamiento de hipertrofia muscular durante 10 semanas y según el crecimiento muscular que presentaron durante el período de entrenamiento fueron clasificados en respondedores altos, medios y bajos. Luego de 6 semanas de desentrenamiento se observó que los individuos que habían ganado inicialmente más masa y fuerza muscular también tendían a perder estas adaptaciones más rápidamente.

El estudio del entrenamiento-desentrenamiento es una temática de relevancia para las Ciencias del Movimiento Humano, tanto desde la perspectiva de la salud como la del rendimiento deportivo. La generación de estudios analizando las respuestas según la población de interés, el tipo de entrenamiento administrado y la duración tanto del entrenamiento como del desentrenamiento según las variables dependientes estudiadas, son importantes para lograr comprender las respuestas individuales ante el ejercicio y la suspensión de éste. El análisis de las disminuciones generadas por el desentrenamiento según el nivel de respuesta inicial al entrenamiento realizado, representa un aporte importante para fortalecer la evidencia que se genera en esta área.

3.2.4. Ensayos controlados aleatorizados por conglomerados

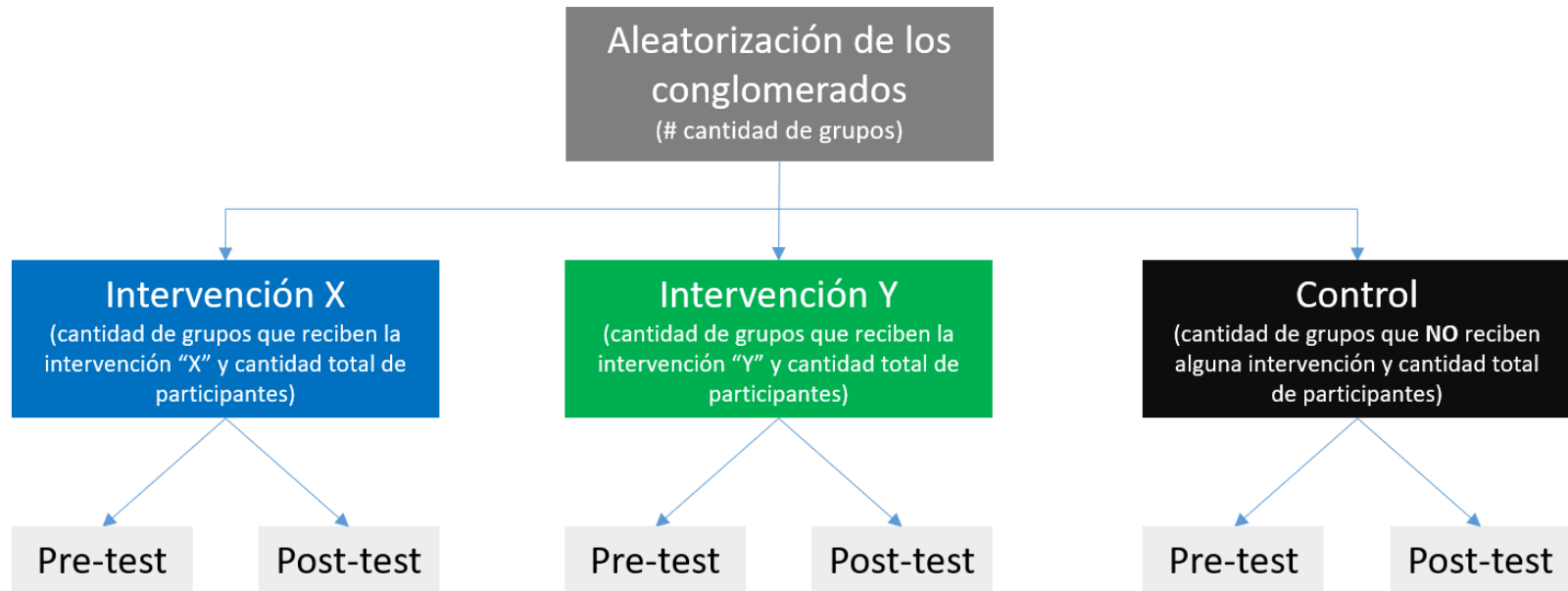
Un ensayo controlado aleatorizado por conglomerados (ECAC) (en inglés, *Cluster Randomized Controlled Trial*), es un tipo de estudio controlado aleatorio en el que se asignan al azar grupos de sujetos. En las Ciencias del Movimiento Humano se encuentran en la literatura más reciente^{[198][588][711]}. A este tipo de investigación también se le conoce como ensayos aleatorios grupales, ensayos aleatorizados grupales y ensayos aleatorizados por lugar. Los ECAC tienen la característica de que individuos particulares no son asignados aleatoriamente a alguna condición de estudio, sino más bien es el grupo de personas establecido (e.g., clase completa de estudiantes) el que es asignado aleatoriamente a alguna condición de estudio. De forma similar a otros estudios, se sigue la metodología de CONSORT^[247] y el protocolo se puede inscribir en algún registro de ensayos aleatorizados^[588].

De Bruijn et al.^[198], realizaron un ECAC, para conocer si dos intervenciones de actividad física escolar tenían algún efecto en el rendimiento en lectura, matemáticas y ortografía en niños de educación primaria. En otras palabras, los investigadores deseaban conocer si la efectividad de las intervenciones de actividad física en el rendimiento académico dependía del tipo de actividad física involucrada. Para ello, formularon un ECAC con una condición experimental que se basaba en actividad física aeróbica de moderada a vigorosa intensidad y una intervención con actividad física cognitivamente atractiva y desafiante.

El procedimiento para el ECAC incluyó primero realizar el análisis de potencia de conglomerados (Tamaño de efecto= 0.40). Con ese análisis obtuvieron que necesitaban una muestra ≥ 40 grupos de clases (i.e., 20 escuelas), en el que hubiera al menos 25 niños por clase ($\alpha= 0.05$, $\beta= 0.90$, coeficiente de correlación intraclase $\rho= 0.10$, 1 cola). La aleatorización de los grupos de estudiantes a las dos condiciones experimentales fue realizada por un investigador independiente. Un ejemplo de la forma en que se podría plantear un ECAC como el propuesto por De Bruijn et al.^[198] se presenta en la figura 1. Sin embargo, también se pueden plantear ECAC sin grupo control^[588], aunque es evidente que presentaría una limitación metodológica.

En el diagrama se puede observar que el primer paso es la aleatorización de los conglomerados (i.e., grupos intactos de participantes) a una de las tres condiciones experimentales. En el ejemplo, se tiene una intervención "X" (e.g., entrenamiento aeróbico, una alternativa novedosa), una intervención "Y" (e.g., ejercicio concurrente, otra alternativa novedosa), y una condición control (e.g., no hace ejercicio o recibe el tratamiento convencional). El estudio incluye la medición de la(s) variable(s) de interés iniciales (i.e., pre-test) y posteriores a la intervención (i.e., post-test).

Figura 1. Diagrama de flujo para el planteamiento de un ensayo controlado aleatorio por conglomerados.



La aleatorización de los conglomerados puede realizarse por grupos completos o puede realizarse utilizando algún criterio de estratificación. Por ejemplo, en un ECAC con diferentes equipos de fútbol en Alemania^[588], la asignación de equipos se estratificó por edad y nivel de juego; es decir, dos equipos de la misma edad y nivel de juego fueron emparejados y asignados a uno de los grupos de intervención mediante el uso de un sobre oculto. En otro estudio con niños futbolistas^[12] se tomaron 94 equipos de fútbol masculino (n= 780 jugadores), los cuales fueron asignados aleatoriamente a grupos experimentales o grupos controles. Los grupos experimentales (n= 45 equipos) completaban el programa "FIFA 11+Kids" como parte del calentamiento durante los entrenamientos y los partidos al menos dos veces por semana. Por su parte, los grupos controles (n= 43 equipos) simplemente realizaron el calentamiento habitual. En este caso no se asigna a los jugadores a una condición en particular, sino a todo el equipo, lo cual cumple con la noción de conglomerado o "cluster".

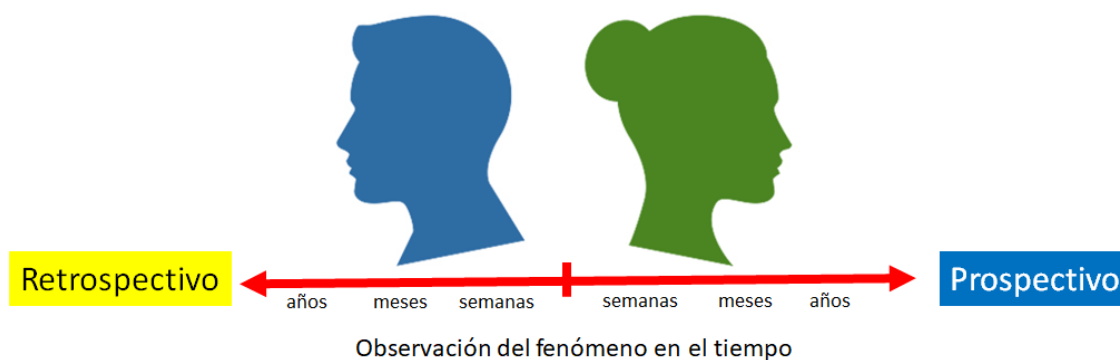
En síntesis, uno de los posibles beneficios de los ECAC en las Ciencias del Movimiento Humano es que se podría trabajar con grandes muestras de personas, lo cual beneficia la potencia estadística de los hallazgos. Sin embargo, una de las posibles debilidades de este diseño, es que grupos completos de individuos podrían no beneficiarse de alguna intervención más allá del beneficio de alguna intervención de rutina o tratamiento comúnmente recibido. Por ejemplo, en el estudio de De Bruijn et al.^[198], no se encontraron efectos significativos de las dos intervenciones de actividad física en el desempeño académico. Este tipo de diseño también oculta la respuesta individual que podría beneficiar a algunos participantes, por lo que las conclusiones deben interpretarse con cautela.

3.2.5. Longitudinal prospectivo, retrospectivo y análisis secundario de datos

Un estudio longitudinal se refiere a un diseño de investigación de tipo observacional en el que se observa y se miden características del mismo grupo de personas durante un período de tiempo previamente planificado. Esa observación se realiza repetidamente durante semanas^[292], pocos meses^[152], pero más comúnmente durante años^{[730][738]}, e incluso durante décadas^[545].

El requisito principal de un estudio longitudinal es que las personas sean medidas durante varios momentos en el tiempo. Este tipo de estudios se pueden realizar planificando obtener mediciones de las personas en el futuro, por lo cual se le definiría como longitudinal prospectivo; o se pueden realizar obteniendo información de las personas revisando expedientes en los que se hayan obtenido datos de los participantes, por lo cual se le denominaría como longitudinal retrospectivo (Figura 1).

Figura 1. Los estudios longitudinales se podrían clasificar como retrospectivos o prospectivos, siendo una característica común la observación del fenómeno de interés en el transcurso del tiempo.



Existen múltiples estudios prospectivos y retrospectivos. Estos estudios siguen el curso normal de cualquier otro tipo de estudio (e.g., aprobación de comité de ética) y además se pueden inscribir en algún tipo de registro (e.g., <https://www.clinicaltrials.gov/>). Existen guías y recomendaciones para completarlos, como, por ejemplo, las guías STROBE (**ST**rengthening the **R**eporting of **OB**servational studies in **E**pidemiology). Esta es una guía que se traduce como "Fortalecimiento del Reporte de estudios Observacionales en Epidemiología", la cual es una iniciativa internacional de colaboración de epidemiólogos, metodólogos, estadísticos, investigadores y editores de revistas involucrados en la realización y difusión de estudios observacionales^[729]. El documento, traducido al español, está disponible en el siguiente enlace de la revista *Gaceta Sanitaria* del año 2008: https://scielo.isciii.es/pdf/gsv/v22n2/articulo_especial.pdf.

Actualmente, se publican protocolos de estudios longitudinales prospectivos^{[201][388]}. Estos protocolos incluyen la estrategia metodológica que se seguirá para el reclutamiento y selección de los participantes. Incluso, algunos protocolos calculan *a priori* el tamaño de la muestra necesario para obtener hallazgos estadísticamente significativos, así como el plan de análisis de datos y de disseminación de los hallazgos.

A pesar de que existen estudios prospectivos en temas de relevancia mundial como la infección por COVID-19^[640], el más importante estudio longitudinal prospectivo realizado en la historia de las Ciencias del Movimiento Humano forma parte de un ambicioso proyecto propuesto en Harvard University, en EE.UU. Ese proyecto ha recolectado datos de exalumnos de esa universidad durante más de 80 años en múltiples áreas del desarrollo humano, como, por ejemplo, felicidad, salud mental, y por supuesto, actividad física^[545].

Para el año 1986, ya se habían obtenido datos de 16 936 exalumnos con edades comprendidas entre los 35 y los 74 años, y se había encontrado que, en un período de 12 a 16 años de seguimiento, ya habían fallecido 1413 de esos individuos. Esa información permitió a los investigadores conocer si la cantidad de actividad física que realizaban se asociaba con una mayor cantidad de años vividos o no. Los investigadores encontraron para ese grupo de personas (i.e., cohorte) que las tasas de mortalidad se reducían a medida que las personas gastaban más energía (e.g., pasando de 500 a 3500 kcal/semana) realizando múltiples tipos de actividad física. Este hallazgo se mantuvo estable aún después de controlar los efectos de enfermedades como la hipertensión, el consumo de cigarrillos, ganancias extremas de peso corporal y otros factores. La conclusión a la que llegaron los investigadores fue que un aumento en la cantidad de actividad física a lo largo de la vida proporciona una protección contra una muerte prematura de 1 a más de 2 años^[545].

En la literatura de las Ciencias del Movimiento Humano también se han reportado estudios longitudinales retrospectivos, los cuales varían en temáticas que abarcan desde aspectos relacionados con el rendimiento deportivo y lesiones^[604], hasta poblaciones específicas como adolescentes^[383] y participantes en programas de rehabilitación cardíaca^{[32][33]}. Por ejemplo, se realizó un estudio retrospectivo de 10 años de recolección de datos, para conocer la frecuencia, características y el manejo de lesiones lumbosacrales en atletas de élite^[333]. Otro ejemplo reciente de un estudio de cohorte retrospectivo se basa en datos isocinéticos y registros clínicos de lesiones de los músculos isquiotibiales de jugadores de fútbol profesional de élite brasileños^[123]. En el estudio se logró recolectar datos desde el año 2009 hasta el año 2019 y se llevó a cabo de acuerdo con la Declaración de Helsinki; además, el protocolo fue aprobado por un comité de ética y siguió las recomendaciones STROBE.

Análisis secundario de datos

Los ejemplos de estudios anteriores demuestran que, por medio de un proceso ordenado y siguiendo una metodología determinada, se puede lograr obtener información valiosa de datos que están almacenados en algún lugar. Un aspecto por considerar es que existen múltiples repositorios desde los cuales se pueden obtener datos recolectados para realizar análisis secundarios; es decir, un análisis de datos que no fue considerado como el principal cuando se propuso el estudio original. En Costa Rica, por ejemplo, el Centro Centroamericano de Población, el Instituto de Investigaciones en Salud (INISA) y la Universidad de California en Berkeley (EE.UU.), desarrollaron un estudio denominado "*Estudio Costarricense de Longevidad y Envejecimiento Saludable (CRELES)*", del cual surgieron múltiples investigaciones, incluyendo variables de la aptitud física^[140]. La base de datos es de acceso público y se puede solicitar completando un formulario en línea (<http://creles-download.demog.berkeley.edu/CRdata.pl>).

En EE.UU., por ejemplo, los Institutos Nacionales de Salud (NIH), poseen el "NIH Data Sharing Repositories" (https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html), un sitio especializado donde se comparten millones de datos de forma gratuita. Las áreas del conocimiento cubren desde la genómica hasta la población adulta mayor (e.g., <https://www.nia.nih.gov/research/dbsr/publicly-available-databases-aging-related-secondary-analyses-behavioral-and-social>), y algunas abarcan temas relacionados con el ejercicio físico, como el estudio "Entrenamiento Cognitivo Avanzado para Mayores Independientes y Vitales (ACTIVO)", desde el cual se pueden utilizar las diferentes bases de datos disponibles gratuitamente (<https://www.icpsr.umich.edu/web/ICPSR/studies/4248?q=Advanced+Cognitive+Training+for+Independent+and+Vital+Elderly+%252528ACTIVE%252529>). Este tipo de análisis secundario de datos permite responder a preguntas de investigación que no fueron consideradas inicialmente pero que surgieron *a posteriori* cuando se obtuvieron nuevas evidencias acerca de ese tema particular. Como tal, esos repositorios o bases de datos son algo así como minas de oro pendientes a ser descubiertas.

De esta forma, en algunas ocasiones se realizan investigaciones en las que se recolecta información de variables que no se utilizan para los análisis principales o primarios; pero que sin embargo, tienen el potencial de brindar información relevante posteriormente. Cuando los investigadores utilizan esos datos que no fueron analizados inicialmente, están recurriendo a una metodología llamada análisis secundario de datos. Un análisis secundario utiliza datos previamente recolectados, procesados, analizados e incluso publicados con un propósito diferente al de los objetivos originales de la investigación. En otras palabras, el análisis secundario es una forma de investigación cuyo propósito es realizar un nuevo estudio utilizando datos existentes recolectados previamente.

Las fuentes de donde provienen los datos pueden ser diversas, como por ejemplo, bases de datos electrónicas, estadísticas publicadas, registros históricos, y expedientes, entre otros. Los datos pueden provenir principalmente de dos fuentes, una interna o una externa. La fuente interna se refiere a datos que se encuentran dentro de la organización a la que pertenece el investigador; mientras que la fuente externa, serían datos que están fuera de dicha organización. Los datos pueden ser de uso público o privado; sin embargo, los investigadores siempre deben consultar si se pueden utilizar los datos y si se debe dar el reconocimiento a los investigadores o a las instituciones propietarias de esos datos.

Los estudios que realizan análisis secundario de datos fueron utilizados exitosamente durante la pandemia por COVID-19 debido al distanciamiento social y a las escasas oportunidades para recolectar datos que enfrentaron los investigadores^[665]. En las Ciencias del Movimiento Humano se publican estudios en los que claramente se identifica, desde su título, que son investigaciones en las que se realizan análisis secundarios de datos. Incluso, actualmente se estimula la inscripción o registro de los protocolos en los que se desea realizar análisis secundarios de datos; esto con el objetivo de cumplir con una serie de recomendaciones^[710] (Tabla 1).

En todo caso, realizar análisis secundarios de datos no representa un menor nivel de investigación; sino más bien la oportunidad de responder a hipótesis que no fueron planteadas inicialmente o que debido al avance del conocimiento, no fueron concebidas inicialmente^[665]. Además, el análisis secundario de datos es mucho más "económico" para los investigadores, pues toda la logística de su recolección fue realizada con anterioridad. Algunos ejemplos de investigaciones en las que se realizaron análisis secundarios de datos se presentan en la tabla 2.

Tabla 1. Recomendaciones recientes para el planteamiento e inscripción de un protocolo de análisis secundario de datos^[710].

Recomendaciones
1. Proporcione el título de trabajo de su estudio.
2. Nombre los autores de la preinscripción.
3. Enumere cada pregunta de investigación incluida en el estudio.
4. Proporcione las hipótesis de su análisis de datos secundarios. Asegúrese de que sean específicos y comprobables, y deje en claro cuál es su marco estadístico (e.g., inferencia bayesiana, NHST). En caso de que su hipótesis sea direccional, no olvide indicar la dirección. Proporcione también una justificación para cada hipótesis.
5. Nombre y describa los conjuntos de datos y, si corresponde, los subconjuntos de datos que planea usar. La información útil para incluir aquí es el tipo de datos (e.g., transversales o longitudinales), el contenido general de las preguntas y algunos detalles sobre los encuestados. En el caso de datos longitudinales, la información sobre las olas del registro de información también es útil.
6. Especifique hasta qué punto el conjunto de datos está abierto o disponible públicamente. Tome nota de cualquier barrera para acceder a los datos, incluso si están disponibles públicamente.
7. ¿Cómo se puede acceder a los datos? Proporcione un identificador persistente o enlace si los datos están disponibles en línea o proporcione una descripción de cómo obtuvo el conjunto de datos.
8. Especifique la fecha de descarga y/o acceso para cada autor.
9. Si el procedimiento de recopilación de datos está bien documentado, proporcione un enlace a esa información. Si el procedimiento de recopilación de datos no está bien documentado, describa, lo mejor que pueda, cómo se recopilaban los datos.
10. Algunos estudios ofrecen libros de códigos para describir sus datos. Si dicho libro de códigos está disponible públicamente, vincúlelo o cargue el documento. En caso contrario, proporcione otra documentación disponible. También brinde orientación sobre qué partes del libro de códigos u otra documentación son más relevantes.
11. Si va a utilizar alguna variable manipulada, identifíquela. Describa las variables y los niveles o brazos de tratamiento de cada variable (tenga en cuenta que esto no es aplicable para estudios observacionales y meta análisis). Si está colapsando grupos a través de variables, esto debe indicarse explícitamente, incluida la fórmula relevante. Si su análisis posterior depende de una verificación de manipulación, también describa sus reglas de decisión.
12. Si va a utilizar variables medidas, identifíquelas. Describa tanto las medidas de resultado como los predictores y las covariables y etiquételes. Si está utilizando una escala o un índice, indique la construcción que representa la escala/índice, de qué elementos consistirá la escala/índice, cómo se agregarán estos elementos y si esta agregación se basa en una recomendación del libro de códigos del estudio o investigación de validación. Cuando la agregación de los ítems se base en el análisis factorial exploratorio (AFE) o el análisis factorial confirmatorio (CFA), especifique también los detalles relevantes (i.e., para AFE: rotación, cómo se determinará el número de factores, cómo se seleccionará el mejor ajuste; para CFA: cómo se especificarán las cargas, cómo se evaluará el ajuste, qué términos de varianza de residuos se correlacionarán). Si está utilizando variables categóricas, indique cómo las codificará en los análisis estadísticos.
13. ¿Qué unidades de análisis (encuestados, casos, etc.) se incluirán o excluirán en su estudio? Teniendo en cuenta estos criterios de inclusión/exclusión, indique el tamaño de muestra (esperado) de los datos que utilizará para sus análisis estadísticos según su conocimiento. En las próximas preguntas, se le pedirá que refine esta estimación del tamaño de la muestra en función de sus juicios sobre los datos faltantes y los valores atípicos.
Continúa...

Tabla 1. Continuación.

Recomendaciones
14. ¿Qué sabe acerca de los datos faltantes en el conjunto de datos (i.e., tasa general de faltantes, información sobre la deserción diferencial)? ¿Cómo manejará los datos incompletos o faltantes? Con base en esta información, proporcione un nuevo tamaño de muestra esperado.
15. Si planea eliminar los valores atípicos, ¿cómo definirá qué es un valor atípico estadístico en sus datos? Proporcione también un nuevo tamaño de muestra esperado. Tenga en cuenta que este será el tamaño de muestra esperado definitivo para su estudio y utilizará este número para realizar cualquier análisis de potencia.
16. ¿Hay ponderaciones de muestreo disponibles con este conjunto de datos? Si es así, ¿los está usando o está usando sus propios pesos de muestreo?
17. Enumere las publicaciones, documentos de trabajo (en preparación, inéditos, preprints) y presentaciones de conferencias (charlas, carteles) en las que ha trabajado que se basan en el conjunto de datos que utilizará. Para cada trabajo, enumere las variables que analizó, pero límitese a las variables que sean relevantes para el análisis propuesto. Si el conjunto de datos es longitudinal, también indique qué ola del conjunto de datos analizó. Es importante destacar que algunos de los miembros de su equipo pueden haber usado este conjunto de datos y otros no. Por lo tanto, es importante especificar los trabajos anteriores de cada coautor por separado. También mencione el trabajo relevante en este conjunto de datos realizado por investigadores con los que está afiliado, ya que es posible que su conocimiento de los datos se le haya traspasado. Cuando el proveedor de los datos también tenga una descripción general de todo el trabajo que se ha realizado con el conjunto de datos, enlace a esa descripción general.
18. ¿Qué conocimiento previo tiene sobre el conjunto de datos que pueda ser relevante para el análisis propuesto? Su conocimiento previo podría provenir de trabajar con los datos primarios, de leer investigaciones publicadas anteriormente o de libros de códigos. También proporcione cualquier conocimiento relevante de los subconjuntos de los datos que no utilizará. Proporcione los conocimientos previos de cada autor por separado.
19. Para cada hipótesis, describa el modelo estadístico que usará para probar la hipótesis. Incluya el tipo de modelo (e.g., ANOVA, regresión múltiple) y la especificación del modelo. Especifique cualquier interacción y análisis post hoc y recuerde que cualquier prueba no incluida debe etiquetarse como una prueba exploratoria en el trabajo final.
20. Si corresponde, especifique un tamaño de efecto previsto o un tamaño de efecto mínimo de interés para todos los efectos probados en sus análisis estadísticos.
21. Presentar la potencia estadística disponible para detectar el tamaño del efecto previsto o el tamaño del efecto más pequeño de interés. O presentar la precisión que se obtendrá para la estimación. Utilice el tamaño de la muestra después de actualizar los datos faltantes y los valores atípicos, y justifique las suposiciones y los parámetros utilizados (e.g., explique por qué algo más pequeño que el tamaño del efecto de interés más pequeño sería teórica o prácticamente irrelevante).
22. ¿Qué criterios utilizará para hacer inferencias? Describa la información que utilizará (e.g., especifique los valores de p, los tamaños del efecto, los intervalos de confianza, los factores de Bayes, los índices de ajuste de modelos específicos), así como los criterios de corte, cuando corresponda. ¿Usará pruebas de una o dos colas para cada uno de sus análisis? Si está comparando múltiples condiciones o probando múltiples hipótesis, ¿lo tendrá en cuenta y, de ser así, cómo?
Continúa...

Tabla 1. Continuación.

Recomendaciones
23. ¿Qué hará si sus datos violan los supuestos, su modelo no converge o surge algún otro problema analítico?
24. Proporcione una serie de decisiones sobre la evaluación de la fuerza, confiabilidad o solidez de su prueba de hipótesis focal. Esto puede incluir intentos de replicación dentro del estudio, covariables adicionales, esfuerzos de validación cruzada (replicación fuera de la muestra, muestra dividida/restringida), aplicación de ponderaciones, aplicación selectiva de restricciones en un contexto SEM (e.g., comparación de estadísticas de ajuste del modelo), técnicas de ajuste de sobreajuste utilizadas (e.g., enfoques de regularización como la regresión de crestas) o algún otro método de simulación/muestreo/arranque.
25. Si planea explorar su conjunto de datos para buscar diferencias o relaciones inesperadas, describa esas pruebas o agréguelas al documento final bajo un encabezado que diferencie claramente esta parte exploratoria de su estudio de la parte confirmatoria.

Tabla 2. Ejemplos de investigaciones en las que se realizaron análisis secundarios de datos.

Referencia	Contexto	Objetivo
Ceballos-Laita et al. ^[151]	El ejercicio terapéutico es uno de los enfoques más investigados para el manejo del síndrome de fibromialgia. La educación en neurofisiología del dolor ayuda a comprender la condición del dolor, lo que lleva a cogniciones del dolor desadaptativas y estrategias de afrontamiento en pacientes con dolor crónico.	Evaluar los efectos del ejercicio terapéutico y la educación en neurofisiología del dolor versus ejercicio terapéutico de forma aislada sobre la fatiga, los trastornos del sueño y la función física a corto plazo y a los 3 meses de seguimiento en mujeres con síndrome de fibromialgia.
Gabrys et al. ^[276]	La rehabilitación cardíaca basada en ejercicios es segura y está implementada en las pautas internacionales de rehabilitación cardíaca. La evidencia de los efectos a largo plazo en la salud es escasa y rara para la investigación de los servicios de atención médica.	Evaluar la efectividad de los programas de rehabilitación cardíaca de fase III basados en ejercicios para mejorar los resultados de mortalidad y capacidad de trabajo.
Gladin et al. ^[287]	La hipercifosis es común en los adultos mayores y se asocia con una función física baja y una calidad de vida relacionada con la salud (CVRS) reducida. Pocos estudios han investigado el cambio en la función física después de una intervención de cifosis dirigida en adultos mayores con baja función física.	Explorar el cambio en la función física después de una intervención de entrenamiento de postura y ejercicio de cifosis progresiva de alta intensidad de 3 meses en adultos mayores con baja función física e hipercifosis. Los objetivos secundarios fueron explorar el cambio en la CVRS, la fuerza de la columna y la curvatura de la columna, y el cumplimiento y la seguridad de la intervención.
Sáez de Asteasu et al. ^[627]	El reposo prolongado en cama aumenta el riesgo de desarrollar deterioro cognitivo y demencia en adultos mayores hospitalizados. Los protocolos de ejercicio aplicados durante la hospitalización aguda pueden prevenir el deterioro funcional en adultos mayores, pero los beneficios del ejercicio en dominios cognitivos específicos no se han investigado previamente.	Evaluar los efectos de una intervención de ejercicio multicomponente para la función cognitiva en adultos mayores durante la hospitalización aguda.
Wu & Brannon ^[763]	La pandemia por COVID-19 ha exacerbado las tasas preexistentes de sobrepeso y obesidad en los Estados Unidos. Las tecnologías de mHealth están ganando popularidad por su potencial para reducir la obesidad, si se facilitan mediante una comunicación centrada en el paciente.	Explorar los predictores de los niveles de ejercicio de las personas con sobrepeso y obesidad durante la pandemia por COVID-19.

3.2.6. Estudios observacionales

Generalmente, en la jerga científica para la evaluación de la eficacia de una intervención en el ámbito social o sanitario, se acepta que el diseño experimental de estudios aleatorizados y controlados (RCT, por sus siglas en inglés) es la mejor alternativa para evitar sesgos en la distribución de los grupos de estudio y un buen control de variables. Sin embargo, este restringido control de variables puede generar diferencias entre lo que sucede en el ámbito experimental y lo que sucede en la realidad de la persona (i.e., baja validez externa), por lo que las conclusiones podrían no ser generalizables ni aplicables en la vida real. De acuerdo con Black^[79], existen cuatro situaciones en las que los diseños RCT podrían ser inapropiados:

- a. Cuando la cantidad de personas (n) incluidas en un estudio no es suficiente para medir con precisión reacciones adversas poco frecuentes.
- b. Cuando exista dificultad para evaluar intervenciones diseñadas para prevenir eventos raros o peligrosos (también derivado del tamaño de la muestra).
- c. Cuando las consecuencias únicamente se podrán medir a largo plazo en un futuro.
- d. Cuando la efectividad de la intervención depende de la disposición, creencias o intereses del sujeto y la asignación aleatoria pueda reducir su efectividad.

Para ese autor, en la práctica clínica hay muchos ejemplos bien conocidos de intervenciones con resultados dramáticos, cuya eficacia fue comprobada finalmente, mediante estudios observacionales.

Antes de definir el tipo de investigación utilizado para desarrollar un estudio, existen dos cuestiones medulares definitivas a considerar; en primer lugar, el problema de investigación, que determinará el diseño a utilizar y no a la inversa. En este sentido se debe considerar la forma más adecuada, razonablemente lógica y rigurosa de abordarlo. En segunda instancia, se debe considerar la manipulación o no del entorno; en este caso, los investigadores deberán definir si se mantendrán al margen del desarrollo de los acontecimientos o si decidirán intervenir en ellos.

Según Manterola y Otzen^[463], este último aspecto logra dilucidar la primera gran división de los diferentes tipos de estudios: a) estudios experimentales, y b) estudios observacionales (EO). Estos últimos corresponden a diseños de investigación cuyo objetivo es "la observación y registro" de acontecimientos sin intervenir en el curso natural de éstos. Si la decisión científica es realizar un EO, se deben definir aspectos como:

- a. Si la medición será única (i.e., estudio transversal) o a lo largo de un período de tiempo (i.e., estudio longitudinal).
- b. Si el estudio se centrará en hechos pasados (i.e., estudio retrospectivo), o si se seguirán los individuos en estudio a través del tiempo, hacia el futuro (i.e., estudio prospectivo).

- c. Si se analizarán comparativamente grupos de sujetos (i.e., estudios analíticos) o sólo se describen un grupo de variables en un momento específico sin incluir grupos de control (i.e., estudios descriptivos).

En los EO, contrario a los estudios experimentales, el investigador es un simple observador y descriptor de lo que ocurre en el ámbito que se ha propuesto analizar. Este tipo de estudios, no controlan la asignación de los sujetos a un determinado tratamiento o intervención, ni las condiciones en las que se dirige la investigación (i.e., únicamente se describen). En este sentido Black^[79] considera que los EO son útiles, pues aportan niveles de evidencia variables según el tipo de escenario del que se trate y, por ende, tienen un valor importante y complementario a la investigación experimental.

Desde la década de los 70's, Anguera-Argilaga^[25], una reconocida científica en el área, menciona que la metodología observacional se caracteriza por la flexibilidad metodológica en sus diferentes etapas y permite el estudio objetivo del comportamiento espontáneo en entornos naturales sin influencia externa, lo que le confiere un alto rigor científico. Para esta y otros investigadores^[28], pero en años recientes, esta metodología integra y conecta elementos cualitativos y cuantitativos en sus etapas de desarrollo, haciendo que un enfoque se construya sobre el otro, por lo que se puede considerar un método mixto por sí mismo. Esta integración propia de los métodos mixtos^{[29][539]} es un requisito metodológico en la investigación basada en la observación sistemática^{[63][64]}.

Hasta antes de 2019, existían algunas guías disponibles para reportar la calidad de los EO considerando la efectividad de las intervenciones, aún en ausencia de asignación aleatoria de participantes a condiciones experimentales (e.g., GRACE, STROBE, STARD, MOOSE, MINCER); sin embargo, ningún instrumento había planteado una lista de verificación de elementos mínimos necesarios, es decir un registro cuantificable y replicable que pudiera guiar a los investigadores en el diseño y análisis de la calidad de los EO en su contexto natural, sin aplicación de una intervención de tipo experimental^{[96][223][577][675][717]}.

Chacón-Moscó et al.^[153], recientemente publicaron la "**M**ethodological **Q**uality **C**hecklist for studies based on **O**bservational **M**ethodology (MQCOM)", con el objetivo de proveer a los profesionales que no son expertos en metodología observacional y carentes de experiencia para guiar su práctica al realizar este tipo de diseños, una lista para completar y verificar aspectos de orden metodológico que permite diseñar, implementar y evaluar estudios de alto rigor científico, aun cuando no son estudios experimentales y en cualquier área del conocimiento. Esta lista incluye un total de 16 ítems que se refieren a 11 criterios o dimensiones. Cada uno se ajusta a decisiones sucesivas en el proceso de investigación utilizando la metodología observacional (Tabla 1). Cada uno de los ítems son calificados, con un 1, cuando se cumple el criterio a cabalidad, 0.5 cuando el criterio se cumple de manera parcial y 0, cuando no se cumple o no se reporta.

Tabla 1. Escala de Calidad Metodológica para Estudios Observacionales: Validez de Contenido y Fiabilidad (MQCOM) (versión traducida por los autores)^[153].

<p>A continuación, se presentan una serie de ítems sobre las características metodológicas de calidad <u>mínimas</u> necesarias que debería considerar en su estudio observacional. Señale en cada ítem la opción que mejor corresponda a su estudio:</p>	
<p>Factor 1. Delimitación de objetivos</p>	
<p>Ítem 1. Se hace referencia a la metodología observacional utilizada, especificando si se trata de observación directa e indirecta:</p> <ul style="list-style-type: none"> 0) No se hace referencia a la metodología. 0.5) Sí, de forma justificada, pero sin documentar. 1) Sí, justificada y documentada. <p>Ítem 2. Delimitación de objetivo/s del estudio:</p> <ul style="list-style-type: none"> 0) No se definen objetivos. 0.5) Definidos comportamental, situacional o temporalmente (falta alguno de los tres referentes). 1) Definidos comportamental, situacional y temporalmente. <p>Ítem 3. Referencia a marco teórico:</p> <ul style="list-style-type: none"> 0) No. 0.33) Sí, sin constructos definidos operativamente. 0.67) Sí, con constructos parcialmente definidos de manera operativa. 1) Sí, con constructos completamente definidos de manera operativa. 	
<p>Factor 2. Diseño observacional</p>	
<p>Ítem 4. Criterio de unidades de observación (ideográfico: una unidad del estudio, formada por uno o varios participantes si existe entre ellos un vínculo estable; o nomotético: dos o más unidades del estudio) ^[577,725]:</p> <ul style="list-style-type: none"> 0) No se identifica. 0.5) Sí, se identifican las unidades de observación, pero sin justificar. 1) Sí, se identifican las unidades de observación justificando la elección de un planteamiento ideográfico o nomotético (de acuerdo con los objetivos del estudio). <p>Ítem 5. Criterio temporalidad (puntual: una o dos sesiones de observación; o seguimiento: más de dos sesiones de observación) ^{[577][725]}:</p> <ul style="list-style-type: none"> 0) No se identifica. 0.5) Sí, identificación del criterio de temporalidad, pero sin diferenciar. 1) Sí, identificación del criterio de temporalidad diferenciando entre seguimiento intersesional e intrasiesional. <p>Ítem 6. Criterio dimensionalidad (unidimensional: un nivel de respuesta; o multidimensional: dos o más niveles de respuesta) ^{[577][725]}:</p> <ul style="list-style-type: none"> 0) No se identifica. 0.5) Sí, identificación de la/s dimensión/es sin referencia a ningún marco conceptual. 1) Sí, identificación de la/s dimensión/es a partir de un marco conceptual. 	
<p>Factor 3. Participantes/unidades de observación</p>	
<p>Ítem 7. Criterios de inclusión y exclusión de las unidades especificados (las razones por las que algunas unidades fueron aceptadas en el estudio y otras no):</p> <ul style="list-style-type: none"> 0) Sin criterios o criterios de selección de las unidades inadecuados y/o con excepciones en su aplicación. 0.5) Incidental (por conveniencia, básicamente por accesibilidad). 1) Con criterios de selección de las unidades adecuados y aplicados a todas las potenciales unidades (muestreo intencional o <i>purposive sampling</i> –selección a partir de criterios de inclusión claramente especificados de acuerdo con los objetivos de estudio–, aleatorio, etc.). 	
<p>Continúa...</p>	

Tabla 1. Continuación

Factor 4. Instrumento de observación
<p>Ítem 8. Adecuación del instrumento de observación (combinación de formato de campo con sistema de categorías, formato de campo, sistema de categorías, o escala de estimación^[27]):</p> <p>0) No inclusión del instrumento de observación como tal (por ejemplo, solamente aportar una lista de conductas).</p> <p>0.5) Descripción del instrumento de observación sin justificar en función de los objetivos y del diseño observacional.</p> <p>1) Instrumento de observación justificado en función de los objetivos y del diseño observacional.</p> <p>Ítem 9. Manual de codificación con definición de las categorías/conductas y especificación de las dimensiones (en diseños multidimensionales):</p> <p>0) Ausencia de manual.</p> <p>0.5) Información parcial (por ejemplo, especificación de dimensiones, pero sin definición de las categorías/códigos de cada dimensión).</p> <p>1) Manual de codificación con definición de las categorías/conductas y especificación de las dimensiones (en diseños multidimensionales).</p>
Factor 5. Uso de <i>software</i>
<p>Ítem 10. Se utiliza software para el registro de datos (SDIS-GSEQ v. 4.2.1. / GSEQ 5., LINCE, MATCH VISION STUDIO, Transana, Otro (indicar); para control de calidad del dato (SDIS-GSEQ V. 4.2.1. / GSEQ 5, LINCE, HOISAN, GT, SAS, Otro –especificar-); y para el análisis de datos (SDIS-GSEQ, HOISAN, THEME v.6, R, SAS, otro –especificar-):</p> <p>0) No.</p> <p>0.5) Sí, usado parcialmente solo para alguno de los tres aspectos.</p> <p>1) Sí, para el registro de datos, para el control de calidad del dato y para el análisis de datos.</p>
Factor 6. Datos
<p>Ítem 11. Especificación del tipo de datos (datos tipo I, tipo II, tipo III y tipo IV^[49]) atendiendo a los dos criterios <i>secuenciales/concurrentes</i> (los datos secuenciales corresponden a conductas que no pueden solaparse entre sí, y pertenecen a una sola dimensión; los datos concurrentes corresponden a conductas que sí pueden coocurrir, y pertenecen a varias dimensiones) y <i>evento-base/tiempo-base</i> (evento-base implica que el parámetro utilizado en el registro es el orden de los eventos, mientras que en tiempo-base es la duración de los eventos):</p> <p>0) No especifica el tipo de datos.</p> <p>0.5) Indica el tipo de datos sin justificarlo.</p> <p>1) Especificación del tipo de datos con justificación.</p> <p>9) No aplicable (no secuencial).</p>
Factor 7. Especificación de parámetros
<p>Ítem 12. Tipo de parámetros en función del uso^{[49][446]}:</p> <p>0) Primario, o básico de registro de una sola categoría (frecuencia, orden y duración).</p> <p>0.5) Secundario, o derivado de registro de una sola categoría (<i>ratios</i> entre indicadores primarios); frecuencia media, frecuencia relativa, tasa, duración relativa, duración media y otros (especificar).</p> <p>1) Mixtos dinámicos o de transición (se trabaja con dos categorías para analizar la transición de una categoría a otra): frecuencia de transición; frecuencia relativa de transición, duración relativa de transición.</p>
Continúa...

Tabla 1. Continuación

Factor 8. Muestreo observacional
<p>Ítem 13. Delimitación de sesiones: establecimiento del criterio (temporal, comportamental o mixto) de inicio y fin de las sesiones dentro del período de observación claramente establecido, y de los criterios de aceptación de sesiones -constancia intersesional, constancia intrasesional o interrupciones temporales- ^[26]:</p> <p>0) No se especifican los criterios de delimitación de sesiones. 0.5) Establecimiento sólo de parte de los aspectos anteriores. 1) Se especifican todos los criterios de delimitación de sesiones.</p>
Factor 9. Control de calidad del dato
<p>Ítem 14. Fiabilidad interobservadores (concordancia entre los registros de diferentes observadores) / fiabilidad intraobservador (entre los registros de un mismo observador en dos momentos de tiempo):</p> <p>0) No se lleva a cabo. 0.5) Concordancia consensuada (cualitativa). 1) Sí: global (basada en indicadores primarios; frecuencia y duración), secuencial (basada en indicadores de orden -secuenciales-): correlación de Pearson, coeficiente intra-clase de Berk, etc.); punto por punto (comparando registro por registro lo que registra cada observador): % totales de acuerdo, coeficiente kappa, teoría de la generalizabilidad, etc.</p>
Factor 10. Análisis de datos
<p>Ítem 15. Análisis realizado^[724]:</p> <p>0) Ningún tipo de análisis de datos. 0.33) Únicamente análisis cualitativo. 0.67) Únicamente análisis descriptivo. 1) Análisis inferencial: relación entre datos categóricos (comparación de proporciones); búsqueda de regularidades (análisis secuencial de retardos; cadenas de Markov; detección de T-Patterns; análisis de coordenadas polares); análisis multivariante (regresión logística; log-lineal; logit-probit; análisis de correspondencias); dimensión temporal (estudios de panel; análisis de tendencias, series temporales); pruebas no paramétricas; pruebas de relación (correlación ordinal; correlación lineal; correlación múltiple); escalamiento multidimensional; otros (especificar).</p>
Factor 11. Interpretación de los resultados
<p>Ítem 16. En la Discusión:</p> <p>0) No se interpretan los resultados, ni se plantean debilidades ni desarrollos futuros. 0.5) Se interpretan los resultados en función únicamente de los objetivos, o bien en función solamente de los trabajos citados de la literatura científica, y/o no se plantean las debilidades ni desarrollos futuros. 1) Se interpretan los resultados conectándolos con los objetivos y con los trabajos citados de la literatura científica; y se plantean debilidades y desarrollos futuros.</p>

Según Chacón-Moscoso et al.^[153], los estudios basados en la metodología observacional han experimentado un enorme progreso debido a los grandes avances en software para registro y análisis de datos, y debido adicionalmente a la multiplicidad de áreas en las que se han aplicado estudios con esta metodología, entre las que destacan las Ciencias Sociales (e.g., Educación, Deportes, Comunicación) así como el área de la Salud (e.g., Nutrición, Medicina, Terapia Física) entre otros.

Recientemente, Soto-Fernández et al.^[662], crearon un software gratuito denominado LINCE PLUS^[661], adecuado para EO sistemáticos en el deporte y la salud, llevado a cabo en contextos naturales como entrenamiento deportivo, educación física y psicología, entre otros. Se pueden analizar parámetros como comportamientos, toma de decisiones o estrategias, utilizando uno o varios videos simultáneamente. El trabajo colaborativo puede lograrse mediante la utilización de videos simultáneos y múltiples observadores. Además, LINCE PLUS muestra los resultados con gráficos interactivos o, si es necesario, exporta los datos a programas de software de análisis de datos específicos como Microsoft Excel®. El programa se puede obtener en el siguiente enlace: <https://observesport.github.io/lince-plus/>.

Algunos ejemplos de EO recientes en las Ciencias del Movimiento Humano se describen a la tabla 2.

Tabla 2. Preguntas de investigación relacionadas con las Ciencias del Movimiento Humano, que han sido contestadas recientemente, mediante la utilización de estudios de tipo observacional.

Referencia	Objetivo o pregunta de investigación
Egger et al. ^[241]	Investigar las causas subyacentes y patrones regionales en la muerte súbita de jugadores de fútbol a nivel mundial, para informar y mejorar la evaluación existente y tomar medidas de prevención.
Graham et al. ^[291]	Identificar la utilización de espacios e instalaciones durante los recreos de la mañana y la hora del almuerzo de una escuela primaria culturalmente diversa, así como las decisiones según género e influencia de variables contextuales y ambientales sobre los niveles de actividad física.
Hayden et al. ^[322]	Evaluar los cambios de pre a post en la disnea (resultado primario) y otros resultados clínicamente relevantes (resultados secundarios) después de un programa de rehabilitación pulmonar con ejercicio, en pacientes según niveles iniciales de gravedad por COVID-19.
Lopes et al. ^[441]	Verificar el perfil músculo-esquelético epidemiológico de lesiones en deportistas de Jiu-Jitsu brasileños.
Machado et al. ^[453]	Comparar la fuerza, la capacidad de contracción de la musculatura del piso pélvico y la incontinencia urinaria entre mujeres practicantes de CrossFit® y no practicantes.

En síntesis, se espera que los EO posiblemente comiencen a ser cada más comunes en la literatura científica de las Ciencias del Movimiento Humano. Estos estudios no son un sinónimo de baja calidad si se atienden guías para realizarlos y se comprenden en un contexto determinado. Como tales, pueden ser los catalizadores de nuevas líneas de investigación en áreas que tradicionalmente no se han explorado.

3.2.7. Estudios de no inferioridad

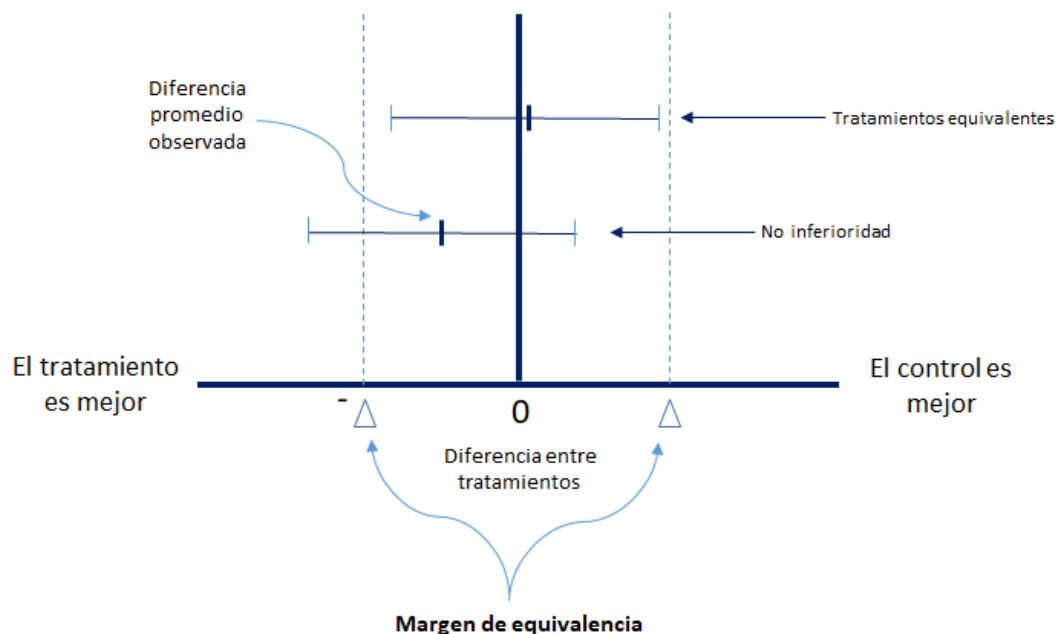
Los estudios de no inferioridad (del inglés, Non-inferiority trial), se refieren a ensayos que pretenden determinar si un nuevo tipo de tratamiento experimental (e.g., programa de entrenamiento novedoso) no es menos eficaz que un tratamiento que ya está en uso (e.g., un programa de entrenamiento tradicional)^{[312][473]}. Estos estudios se realizan con el propósito de saber si los potenciales beneficios del nuevo tratamiento pueden ser solamente marginales en comparación con otros tratamientos existentes. Los estudios de no inferioridad son más complejos de diseñar, realizar e interpretar que los estudios de superioridad o contrafactuales.

En los estudios de superioridad se considera que el tratamiento experimental es eficaz cuando demuestra su superioridad frente a un placebo, control o si demuestra una relación dosis-respuesta. Los estudios de no inferioridad se realizan para demostrar que un nuevo tratamiento es mejor que un supuesto placebo o control en situaciones en las que realizar un estudio de control con placebo no es ético^[312]. Estos estudios se pueden diseñar cuando un nuevo tratamiento puede ofrecer ventajas sobre los tratamientos estándares disponibles, ya sea porque son más seguros, tienen mejor adherencia, o incluso menor costo. Así, para determinar la no inferioridad de un tratamiento, se busca demostrar que un tratamiento experimental no es menos efectivo que un control activo por más del margen de equivalencia^[473]; es decir, de una cantidad tolerable. De esta forma, en un estudio en el que se intenta demostrar que hay una diferencia menor a una cantidad específica entre los tratamientos de control y experimentales, un diseño de no inferioridad prueba estadísticamente la hipótesis nula (H_0) de que el tratamiento experimental es inferior que el margen de equivalencia^[473]. Por ejemplo, en un estudio donde se desea un resultado de valores más altos, si el límite superior del IC95% para la diferencia de tratamiento de la condición experimental es menor que el margen de equivalencia, entonces se puede rechazar la H_0 de inferioridad (Figura 1).

En las Ciencias del Movimiento Humano existen protocolos^{[14][59]} y estudios completos de no inferioridad. Los protocolos permiten conocer *a priori*, los antecedentes teóricos y conceptuales, así como los detalles de la metodología y las comparaciones que se desean realizar. Los estudios concluidos, son los que comúnmente se encuentran en las bases de datos que proporcionan resultados y discusión.

Por ejemplo, la terapia de privación de andrógenos se usa comúnmente para tratar el cáncer de próstata en hombres; sin embargo, los efectos secundarios empeoran el funcionamiento físico y la calidad de vida de los pacientes. Actualmente, se sabe que tanto el entrenamiento contra resistencia como el ejercicio aeróbico brindan beneficios a esa población. La forma en que se prescribe el ejercicio, ya sea supervisada en algún lugar o en el hogar, puede ser importante para estos pacientes, por lo que se diseñó un estudio de no inferioridad con el que se deseaba determinar si tres modelos de prescripción del entrenamiento físico eran equivalentes en términos de beneficios en la calidad de vida y el estado físico en hombres con cáncer de próstata^[14].

Figura 1. Análisis de estudios de no inferioridad utilizando el IC95% y un margen de equivalencia (Δ) definido arbitrariamente. Si la diferencia entre las medias de los tratamientos y el IC95% están entre el margen de equivalencia, se dice que ambos tratamientos son equivalentes; pero si la media y el IC95% sobrepasan el margen de equivalencia en la dirección del tratamiento, se dice que el nuevo tratamiento no es inferior al tratamiento actual, el cual se usa como control.



Los tres protocolos de ejercicios eran la terapia convencional personal o tratamiento estándar (1:1), el entrenamiento grupal supervisado, o una aplicación telefónica para completarla en el hogar. El protocolo del estudio se completó y el trabajo finalizado se publicó posteriormente^[15], y se encontró que no había inferioridad con el tratamiento estándar de persona a persona (1:1) (Tabla 1).

Tabla 1. Variables físicas de interés de un estudio de no inferioridad, utilizando márgenes de equivalencia^[15]. Los valores representan la diferencia entre promedios y el IC95% comparando un programa de ejercicio en el hogar y un programa de ejercicios grupales contra un programa de ejercicios individualizado (1:1).

	VO₂pico (ml·kg⁻¹·min⁻¹)	Fuerza de agarre (kg)	Ponerse de pie desde una silla (repeticiones)
Margen de equivalencia →	2.5	4.5	4
Ejercicios grupales vs. 1:1	-0.7 (-8.7 a 3.8)	-0.3 (-3.3 a 2.7)	0.7 (-4.3 a 5.5)
Ejercicio en el hogar vs. 1:1	-1.8 (-4.2 a 0.6)	-3.4 (-6.3 a -0.6)	1.4 (-3.3 a 5.9)

En la tabla 1 se puede observar el margen de equivalencia para cada variable de la aptitud física evaluada, en donde se presentan en las unidades de medida de dicha variable. Así, por ejemplo, el margen de equivalencia para la capacidad aeróbica pico fue de 2.5 ml·kg⁻¹·min⁻¹, para la fuerza de agarre fue de 4.5 kg y para ponerse de pie desde una silla fue de 4 repeticiones. Esos valores se toman como referencia para determinar si la diferencia promedio con el tratamiento usual sobrepasa esos márgenes de equivalencia o no, y dependiendo de eso, se decide si el nuevo

tratamiento no es inferior al tratamiento usual. Para ello se observa también el IC95% que indicaría la confianza de la estimación.

Existen estudios de no inferioridad cuyo objetivo es comparar diferentes formas de entrenamiento. Por ejemplo, el entrenamiento de intervalos de alta intensidad (HIIT, por sus siglas en inglés), se sugiere como un tipo de entrenamiento que se realiza en un tiempo muy breve y que puede proporcionar los mismos beneficios que programas de entrenamiento que consumen más de 30 min y que son de moderada intensidad. De tal forma, recientemente se han diseñado estudios de no inferioridad para determinar si el HIIT no es inferior al entrenamiento continuo de moderada intensidad (MICT, por sus siglas en inglés)^[192].

En ese estudio la pregunta de investigación establece si la diferencia promedio de la pérdida de peso con el método HIIT no es inferior al método convencional de MICT. Existen 3 posibles respuestas. Primero, que tanto HIIT como MICT sean equivalentes; segundo, que uno sea superior al otro (i.e., HIIT > MICT o que MICT > HIIT); y tercero, que HIIT no sea inferior al MICT. Los investigadores encontraron que el HIIT parece no ser inferior a MICT en la inducción de pérdida de peso en adultos obesos, por lo que HIIT se podrían recomendar como un tipo de entrenamiento alternativo, de corta duración y que proporciona los mismos beneficios que un tipo de entrenamiento tradicional^[192]. Otros ejemplos de estudios de no inferioridad se pueden observar en la tabla 2.

Tabla 2. Ejemplo del propósito de estudios de no inferioridad en las Ciencias del Movimiento Humano.

Referencia	Propósito
Hägglund et al. ^[311]	<ul style="list-style-type: none"> a. Determinar si el entrenamiento de yoga e hidroterapia tienen el mismo efecto sobre la calidad de vida relacionada con la salud en pacientes con insuficiencia cardíaca. b. Comparar los efectos sobre la capacidad de ejercicio, los resultados clínicos y los síntomas de ansiedad y depresión entre y dentro de los dos grupos.
Langeard et al. ^[417]	<ul style="list-style-type: none"> a. Determinar si el HIIT es tan eficaz como la Hidroclorotiazida para reducir la presión arterial ambulatoria de 24 h en adultos mayores prehipertensos. b. Determinar si el HIIT se asocia con mayores beneficios cardiovasculares y cognitivos que la Hidroclorotiazida.
López-Torres Hidalgo ^[443]	Comprobar la no inferioridad del ejercicio físico supervisado frente al tratamiento farmacológico antidepressivo, en términos de reducción de los síntomas depresivos en pacientes que presentan criterios clínicos de episodio depresivo (CIE-10), a lo largo de un período de seguimiento de 6 meses.
Tousignant et al. ^[701]	Probar la no inferioridad de un programa de ejercicios basado en Tai Chi proporcionado a través de la tele rehabilitación en comparación con el mismo programa proporcionado en persona en el hogar en términos de efectividad para reentrenar el equilibrio en personas que han tenido un accidente cerebrovascular pero que no requieren rehabilitación funcional intensiva.
Werkhausen et al. ^[747]	Probar si el entrenamiento de resistencia explosiva con rango de movimiento (ROM) parcial sería tan efectivo como el entrenamiento de ROM completo utilizando un diseño de prueba de no inferioridad.

En síntesis, los estudios de no inferioridad en las Ciencias del Movimiento Humano permiten conocer la intervención mínima que sería igual a un tratamiento tradicional o previamente establecido. Este tipo de diseño de estudio es novedoso y se espera que se publiquen más a menudo debido a que existen numerosos estudios contrafactuales que han sido resumidos por medio de meta análisis. Por lo tanto, ya que se conoce la superioridad de muchas intervenciones de ejercicio y entrenamiento físico sobre una amplia variedad de variables, ahora se espera conocer la no inferioridad de nuevas intervenciones.

3.2.8. Sujeto único-estudio de caso y ABA

Para poder avanzar hacia la personalización de las intervenciones, se necesita comprender los mecanismos de acción y las respuestas individuales. Comúnmente llamados estudios de N de 1, estudios de caso único y estudios intra-persona, los estudios de sujeto único parten de un enfoque idiográfico, el cual supone que los individuos son seres irrepetibles. Su objeto es comprender la individualidad humana, por lo que puede resultar en un proceso intensivo de observación y seguimiento de una misma persona. De esta manera, el producto final de las investigaciones que aplican metodologías de estudios de sujeto único aporta al desarrollo de intervenciones, tratamientos y diagnósticos precisos e individualizados^[494].

Kinugasa^[400], destaca que los estudios de caso son similares a los estudios de sujeto único. Si bien una característica común de ambos tipos de estudios es la presencia de múltiples observaciones de una misma persona, los estudios de caso comúnmente tienen un enfoque retrospectivo en el que se estudia al individuo bajo condiciones que no fueron necesariamente sistematizadas o controladas; mientras que los diseños de investigación de sujeto único suelen utilizar enfoques cuasi experimentales. Ese investigador incluye una tabla resumen de 12 estudios de sujeto único publicados entre el año 2004 y 2011, donde evidencia que el análisis de datos utilizado suele ser predominantemente descriptivo (e.g., frecuencias, promedios, desviaciones estándar, análisis visual) o poco tradicionales en la disciplina de las Ciencias del Movimiento Humano y el deporte (e.g., análisis de series temporales, análisis de datos superpuestos).

En estos estudios, la potencia estadística no estaría determinada por el tamaño de la muestra o por el número de participantes, sino por el número de observaciones de la variable dependiente de interés, realizadas en repetidas ocasiones^[412]. Incluso, se pueden utilizar análisis robustos que posean una potencia estadística adecuada al utilizar menor cantidad de mediciones, como el método de autocorrelación de series temporales interrumpidas (ITSACORR, por sus siglas en inglés). Ese método tiene la ventaja de poder utilizar menos observaciones (e.g., a partir de 10) en comparación con otros análisis de series temporales ampliamente utilizados^[401]. Los métodos validados para análisis visual también son de particular interés en estos estudios debido a su popularidad y robustez. Sin embargo, una de sus limitaciones es que no llevan a resultados que puedan ser comparados de manera sencilla entre estudios similares^{[221][277]}.

En la literatura se destaca que los estudios de un único sujeto son particularmente idóneos para documentar características únicas de atletas de clase mundial y de casos clínicos^[750], ya que poseen el potencial para comprender los escenarios reales en los que se desenvuelve el atleta, así como estudiar cambios longitudinales en la condición física, emocional y/o indicadores fisiológicos del deportista de acuerdo con circunstancias de la temporada de entrenamiento, presencia de lesiones, y sobreentrenamiento, entre otros; y determinar qué tan adecuadas están siendo las cargas de entrenamiento; lo que representa un aporte al conocimiento existente de las ciencias aplicadas a los deportes^{[56][181][400]}, que en algunas disciplinas es escaso.

Kwasnicka y Naughton^[412] indican que en el área de estudio de cambios de conducta de actividad física y estilo de vida, los estudios de N=1 poseen gran potencial si se emplean nuevas tecnologías que permitan recolectar datos no solo día a día, sino también de manera longitudinal, y que a su vez muestren adecuada adherencia con una mínima interrupción, tal como los relojes inteligentes, los dispositivos móviles y los monitores de actividad física. Por otro lado, algunas de

las limitaciones que presentan estos estudios incluyen la posibilidad de establecer la generalizabilidad y confiabilidad de los hallazgos reportados, la falta común de análisis que permita realizar inferencias y la presencia de dificultades técnicas en cuanto al análisis de datos de un solo sujeto^[400].

Los enfoques metodológicos de sujeto único también pueden aplicarse a un pequeño grupo de individuos, en el entendido de que la naturaleza de la muestra o de la intervención no lleva necesariamente a que se considere un diseño de investigación basado en grupos. Lo mismo ocurriría en aquellos casos cuya unidad de estudio no es una persona, sino una comunidad específica, un equipo deportivo, una institución o una familia^[412]. Por ejemplo, en el estudio de Gallo et al.^[279], se describen las estrategias de entrenamiento individual de tres ciclistas de ruta en su preparación para la competencia *Giro d'Italia*, donde los ciclistas finalizaron dentro de las cinco mejores posiciones de la clasificación general. El estudio examinaba de manera retrospectiva el volumen e intensidad del entrenamiento semanal hasta la puesta a punto, así como detalles del entrenamiento en altura. Si bien en una sección del estudio se reportan los datos de manera agregada, la mayor parte del manuscrito se enfoca en describir los valores semanales obtenidos por cada atleta de acuerdo con la planificación deportiva individual.

La pregunta de investigación en estudios de N=1

Un paso esencial en cualquier estudio es establecer la pregunta de investigación. En el caso de los estudios de sujeto único, la pregunta de investigación estará orientada a la exploración de trayectorias y el registro de patrones individuales, así como la indagación de variables que anteceden o predicen un comportamiento o conducta (e.g., sedentarismo, actividad física)^[412] o que predicen el rendimiento. En síntesis, la pregunta de investigación estará orientada a examinar los cambios (o ausencia de los mismos) en una o muchas variables de interés, bajo la comprensión de qué hace única a una persona o cuáles son las respuestas específicas a un tratamiento. En las Ciencias del Movimiento Humano, estas variables dependientes de interés suelen ser conductas, percepciones, indicadores de desempeño físico, o constructos psicosociales; lo cual no es una excepción en los estudios de sujeto único^[56]. A nivel metodológico, algunas de las guías existentes para el reporte de los estudios en cuestión incluyen la extensión CONSORT para reportar ensayos de N= 1 (CENT, por sus siglas en inglés)^[727] y el lineamiento para el reporte de caso único en intervenciones conductuales (SCRIBE, por sus siglas en inglés)^[686].

Diseño de investigación A-B y A-B-A

Al igual que los estudios basados en grupos, los de sujeto único también requieren un diseño de investigación, el cual se escogerá con base en la idoneidad para responder a la pregunta de investigación. Los dos diseños predominantes para los estudios de N=1 son observacional y experimental. De manera similar a los estudios grupales, el observacional suele ser de naturaleza exploratoria, cuyas mediciones se utilizan para comprender patrones, tendencias o asociaciones en un momento dado (e.g., relaciones temporales), mientras que en el experimental existe la manipulación de una o más variables para examinar el efecto de una intervención o tratamiento sobre una variable dependiente^[412].

Así, debido a que una de las principales aplicaciones de los estudios de sujeto único es comprender la respuesta individual a una intervención, dentro de la gama de opciones existentes se encuentran los diseños A-B y A-B-A, donde el proceso puede entenderse por fases. Así, puede existir una fase de control o línea base (A) y una fase de intervención (B)^[400]. En el caso específico del diseño A-B-A, se busca llevar a la persona participante nuevamente a una condición similar a la basal o control (A) una vez que se haya suspendido la intervención (B).

Por ejemplo, un investigador puede utilizar el diseño A-B para examinar la efectividad de las sesiones de entrenamiento a partir de los cambios en indicadores fisiológicos de un triatlonista, utilizando su período de preparación general como línea base (A) y la fase de preparación precompetitiva como etapa de intervención (B). De acuerdo con Kinugasa^[400], este modelo también permitiría a los entrenadores y profesionales en Ciencias del Movimiento Humano determinar si un deportista en particular responde bien, mal, o no responde a un determinado método de entrenamiento en la fase B; por ejemplo, al entrenamiento en condiciones de calor y humedad, a las sesiones en altura o a entrenamientos utilizando distintas indumentarias.

La fundamentación teórica en los estudios de sujeto único también es relevante, ya que se debe justificar por qué este diseño es el más apropiado para examinar un fenómeno o una intervención determinada. De acuerdo con el ejemplo del párrafo anterior, la justificación teórica de que la realización de un estudio de sujeto único sería el más apropiado radica en que la evidencia previa demuestra la existencia de diferencias inter-individuales en la respuesta a los métodos de entrenamiento, en donde existen atletas que mejoran su rendimiento en pruebas de fondo en respuesta al entrenamiento de altura, mientras que el rendimiento de otros atletas no cambia^[400].

En síntesis, a pesar de que los estudios de un solo sujeto permiten un enfoque que se dirige a describir procesos particulares de la persona de estudio, describiendo las repuestas personalizadas y explicando diferencias interindividuales, su aplicación en las Ciencias del Movimiento Humano, el deporte^[400], y en la psicología y las ciencias del comportamiento^[412], continúa siendo escasa.

3.2.9. Estudios econométricos

Por primera y única vez en su historia, la Organización de las Naciones Unidas (ONU) declaró el 2005, como el “Año Internacional de la Educación Física”, brindándole un estatus sobresaliente para el desarrollo de los habitantes del planeta^[500]. Para tomar esa decisión no sólo se consideraron aspectos políticos, sino también evidencias que aún hoy en día apoyan la relevancia del movimiento del ser humano en diversos escenarios. Por ejemplo, en el año 2005 se estimaba que la productividad humana se triplicaría si a través de la actividad física las personas lograban reducir el sedentarismo, lo cual equivalía a una ganancia de US\$2 a \$5 por cada \$1 invertido en prevenir el sedentarismo. Proyecciones más recientes indican que el producto interno bruto (PIB) podría ser entre 0.15% y 0.22% más alto en comparación con la línea base si se adoptan las guías actuales de actividad física para la población^[310]. Este tipo de aseveraciones solamente podrían ser fidedignas si las apoya un análisis de datos robusto, como los que provienen de la econometría.

La econometría se puede definir como “la aplicación de métodos estadísticos y matemáticos para analizar la teoría económica con datos mediante el uso de diferentes técnicas de estimación y pruebas de hipótesis relacionadas con las teorías económicas” (p. 4)^[196]. Un modelo econométrico se construye sobre la base de un modelo económico que explica la racionalidad de la posible relación entre las variables incluidas en el análisis. Así, la econometría utiliza métodos estadísticos para el análisis de fenómenos económicos, en donde se combina “la teoría económica y la estadística para formular y analizar una pregunta económica interesante” (p. 5)^[196].

Entre esas preguntas interesantes que permite responder la econometría, se encuentran una gran cantidad directamente relacionadas con las Ciencias del Movimiento Humano, las cuales se abordan desde diferentes teorías econométricas y son analizadas con técnicas estadísticas diversas (Tabla 1). Las preguntas, y especialmente las respuestas, generan no solamente nuevo conocimiento y aportan a la generación o confirmación de teorías, sino que tienen el potencial de influir en las políticas públicas de los gobiernos y en decisiones personales de los ciudadanos. Por ejemplo, si se demuestra con datos econométricos la importancia de incrementar la cantidad de tiempo de las clases de Educación Física, es posible que se pueda convencer a políticos y tomadores de decisiones para que inviertan más en este rubro, o que contraten más docentes para que impartan lecciones de esta importante materia.

Tabla 1. Ejemplos de preguntas con potencial para cambiar políticas públicas respondidas por medio de la econometría.

Referencia	Pregunta	Métodos econométricos de estudio	Respuestas
Bachmann et al. ^[44]	¿Cuál es la asociación entre los costos de atención médica en la vejez y la aptitud cardiorrespiratoria en la mediana edad después de ajustar por factores de riesgo cardiovascular?	<ul style="list-style-type: none"> • Comparación de costos anuales de atención médica con la prueba de ANOVA Kruskal-Wallis. • Regresión de Poisson. • Modelaje de asociaciones entre los costos anuales de atención de la salud y la condición física en la mediana edad mediante regresión lineal generalizada con distribución de Poisson con función log-link. • Regresión por cuantiles. • Modelos estratificados por sexo y ajustados según la edad promedio durante el seguimiento y la edad en el examen físico inicial. • Ajuste de modelos considerado estado de tabaquismo, diabetes, colesterol total, presión arterial sistólica e índice de masa corporal (IMC). 	<ul style="list-style-type: none"> • Una mayor aptitud cardiorrespiratoria en la mediana edad está fuertemente asociada con menores costos de atención médica en un promedio de 22 años más tarde en la vida, independientemente de los factores de riesgo cardiovascular de la persona. • Para los hombres, hay una reducción del 6.8% en los costos de atención médica por cada MET alcanzado (IC95%=5.7% a 7.8%) y para las mujeres de 6.7% (IC95%= 4.1% a 9.3%).
			Continúa...

Tabla 1. Continuación.

Referencia	Pregunta	Métodos econométricos de estudio	Respuestas
Carlson et al. ^[144]	¿Cuál es la asociación entre la actividad física aeróbica realizada en el tiempo libre y los gastos de atención médica en una muestra representativa a nivel nacional de adultos de EE.UU. no institucionalizados con y sin ajuste por estado de obesidad?	<ul style="list-style-type: none"> • Modelo econométrico de 4 partes: <ul style="list-style-type: none"> ○ Dos modelos <i>Probit</i> para predecir la probabilidad de tener un gasto en salud positivo. ○ Dos modelos lineales generalizados con un enlace de registro y distribución gamma para predecir los gastos totales de atención médica por separado para los adultos con un gasto positivo de atención médica. • Pruebas de Park modificadas para determinar la especificación de distribución adecuada para los modelos lineales generalizados. 	<ul style="list-style-type: none"> • El 11.1% (IC95%= 7.3 a 14.9%) de los gastos adicionales de atención de la salud se asoció con una actividad física inadecuada. • El aumento de la actividad física de los adultos puede reducir los gastos de atención médica. • Después de ajustar las principales covariables, la diferencia media de gasto anual per cápita para adultos inactivos en comparación con los adultos activos fue de US\$1437 ($\Delta=29.9\%$) y por insuficientemente activa en comparación con los adultos activos fue de US\$713 ($\Delta=15.4\%$). • Después de incluir la categoría de IMC como covariable, las medias del gasto anual y las diferencias porcentuales para adultos inactivos (US\$ 1313; $\Delta=26.6\%$) y adultos insuficientemente activos (US\$576; $\Delta=12.1\%$) comparados con adultos activos disminuyeron ligeramente, pero las diferencias siguieron siendo significativas.
			Continúa...

Tabla 1. Continuación.

Referencia	Pregunta	Métodos econométricos de estudio	Respuestas
Coughlan et al. ^[185]	¿Cuál es la asociación entre la actividad física realizada en el tiempo libre durante la edad adulta y los costos de atención médica en la vejez?	<ul style="list-style-type: none"> • Estimación de probabilidades de incurrir en un costo de atención médica durante el período de tiempo. • Ajuste de modelo lineal generalizado para los costos de atención médica totales anuales promedio basados en la prueba de Park modificada (distribución de Poisson) y la prueba de bondad de enlace (log link) con errores estándar robustos. 	<ul style="list-style-type: none"> • La carga del costo de la atención médica en la vejez podría reducirse mediante esfuerzos de promoción que respalden la participación en actividades físicas durante la edad adulta. • Los costos de atención médica en la vejez fueron mucho más bajos para quienes siempre mantuvieron niveles de actividad física moderados (US\$ 1350/año; 15.9%) o altos (US\$ 1200/año; 14.1%) y para quienes aumentaron los niveles de actividad física en la edad adulta temprana (US\$1874/año; 22.0%) o en la mediana edad (US\$824/año; 9.7%) en comparación con adultos que permanecieron constantemente inactivos desde la adolescencia hasta la mediana edad. • Los adultos que disminuyeron sus niveles de actividad física más adelante tuvieron reducciones de costos moderados (<US\$206/año; 2.4%). • Los adultos que aumentaron los niveles de actividad física durante la adultez temprana con una disminución en la mediana edad mostraron una diferencia significativa (US\$861/año; 10.1 %) en comparación con los adultos que eran constantemente inactivos.
			Continúa...

Tabla 1. Continuación.

Referencia	Pregunta	Métodos econométricos de estudio	Respuestas
Gomes et al. ^[289]	¿Cuál es la asociación entre los niveles de actividad física y las trayectorias de los costos médicos y farmacéuticos durante 12 años en mujeres australianas de mediana edad?	<ul style="list-style-type: none"> • Modelos de regresión por cuantiles crudos y ajustados para estimar las diferencias en los costos medianos para cada una de las variables de actividad física. • Análisis ajustados para los factores de confusión edad, educación, estado civil, área de residencia, tener tarjeta de salud, tabaquismo, alcohol e IMC. 	<ul style="list-style-type: none"> • Ayudar a las mujeres a mantenerse activas en la mediana edad podría generar ahorros personales y gubernamentales considerables. • Mantenerse físicamente activas se asoció con un plan de atención médica 40% más bajo y un plan de medicinas 30% más bajo durante tres años en las mujeres australianas. • Costos de entre AU\$426 y AU\$3013 más bajos entre las mujeres que estuvieron constantemente activas que entre las que siempre estuvieron inactivas.
			Continúa...

Tabla 1. Continuación.

Referencia	Pregunta	Métodos econométricos de estudio	Respuestas
Kang & Xiang ^[385]	¿Cuál es la relación entre la actividad física y la utilización y los costos de los servicios de salud entre adultos ≥ 18 años en EE.UU?	<ul style="list-style-type: none"> • Regresión logística y test de Wald para probar la relación entre la actividad física y la utilización de 10 servicios preventivos. • Se estimaron modelos de dos partes para la utilización y los costos de medicamentos recetados, consulta médica, para pacientes ambulatorios, para pacientes hospitalizados, pacientes en la sala de emergencias y en el hogar. <ul style="list-style-type: none"> ○ La primera parte involucró la estimación de la probabilidad de usar un tipo específico de servicios de salud usando una regresión logística. ○ La segunda parte involucró el ajuste de un Modelo Lineal Generalizado (GLM) con distribuciones log link, Poisson, binomial negativa y gama, para la cantidad total de utilización de los servicios y los costos asociados condicionados a tener un número positivo de visitas. 	<ul style="list-style-type: none"> • Promover la actividad física regular puede reducir los costos de atención de la salud a través de la disminución de la demanda de servicios de atención secundaria y terciaria. • En comparación con adultos sedentarios, los adultos físicamente activos incurren en una utilización significativamente menor de hospitalizaciones (0.09 vs. 0.12 visitas/persona), uso de sala de emergencias (0.18 vs. 0.19 visitas/persona), atención médica domiciliaria (1.21 vs. 1.92 visitas/persona) y medicamentos recetados (12.66 vs. 13.75 recetas/persona), y además gastaron \$27 menos per cápita en visitas al consultorio, \$351 menos en hospitalizaciones y \$52 menos en visitas de atención médica domiciliaria.
			Continúa...

Tabla 1. Continuación.

Referencia	Pregunta	Métodos econométricos de estudio	Respuestas
Kelly et al. ^[396]	¿Cuál es el costo-beneficio (rentabilidad económica) de un programa de promoción de la actividad física como intervención de salud pública que se implementa a nivel nacional en Irlanda?	<ul style="list-style-type: none"> • La rentabilidad se determinó comparando los costos (directos e indirectos) del programa de 12 semanas con sus beneficios. • Se estudió el impacto en los años de vida ajustados por calidad (AVAC). 	<ul style="list-style-type: none"> • El programa de promoción de la actividad física es rentable en el apoyo a una cohorte de hombres “en riesgo” que logra mejoras significativas en la aptitud aeróbica, la pérdida de peso y la reducción de la circunferencia de la cintura. • El costo total por participante (€125.82 para cada uno de los 501 participantes de la intervención), los AVAC ganados (5.3% de mejora de la salud por participante) y los costos estimados de la relación de AVAC de €3723 representan una mejoría en la relación costo-efectividad en comparación con otras guías AVAC.

La información proveniente de estudios econométricos también permite al personal de atención de la salud poseer valiosos insumos para asesorar a sus pacientes con el fin de convencerlos, con datos objetivos, para que mantengan o aumenten su actividad física, especialmente durante la edad adulta. También, las personas más estructuradas y ordenadas pueden usar este tipo de análisis para considerar las posibles implicaciones financieras para la atención médica en la edad adulta mayor^[185]. De esta manera, las personas pueden planificar los potenciales gastos en salud y disfrutar de una mejor calidad de vida, especialmente en la adultez mayor, que es cuando más dinero se invierte en la atención de la salud.

Es necesario distinguir entre estudios económicos y estudios econométricos. Los estudios económicos permiten conocer el costo de alguna intervención. Por ejemplo, Babey et al.^[43] evaluaron varios tipos de estrategias para estimular la actividad física en los jóvenes. Entre esas oportunidades se encontraban los programas para después de finalizadas las clases escolares, los programas para antes de la ir a la escuela, las clases regulares de Educación Física, extender la jornada de las clases de Educación Física, y los descansos activos breves durante la jornada escolar.

Para realizar ese estudio económico, los investigadores utilizaron los costos (e.g., salario del personal, suministros y materiales, equipo, costos de las instalaciones, costos de transporte), el alcance, los efectos sobre las ganancias de la actividad física, la rentabilidad y otros beneficios potencialmente crecientes y encontraron que únicamente dos programas fueron superiores en términos de alcance y costo por estudiante: a) extender el día escolar con participación obligatoria en las clases de Educación Física y b) ofrecer descansos cortos (10 min) para la actividad física durante el horario regular de clases. Los costos de los programas extracurriculares por estudiante son altos, de aproximadamente \$2867 por año escolar por niño dentro de las instalaciones escolares (rango de \$2000 a \$3800) y de \$3047 por año escolar por niño fuera de las instalaciones escolares (rango de \$2180 a \$3980) y los programas tienen un alcance menor, pero ofrecen beneficios (como el cuidado de los niños) que pueden justificar sus costos más altos. Los programas antes de la escuela no parecían viables.

Al momento de la realización de esta obra, no se encontró en la literatura una metodología estandarizada para diseñar estudios econométricos. Dos revisiones sistemáticas encontraron que algunos estudios se enfocan únicamente en los costos directos de atención médica, otros en los costos indirectos estimados y otros en los costos del hogar^{[212][532]}. Incluso dentro de las metodologías hay enfoques distintos; por ejemplo, para los estudios de costos directos, algunos estudios utilizan un enfoque de fracción atribuible a la población (FAP) con costos de atención médica estimados atribuibles a la inactividad física.

También se realizan estudios de evaluación económica completa como los análisis de costo-efectividad, análisis de costo-beneficio y análisis de costo-utilidad. Hay otro enfoque basado en la población o enfoque basado en la persona que contempla el costo de la enfermedad, así como otro que estudia el retorno de la inversión^[532]. Otros estudios utilizan un enfoque econométrico, que además tiende a producir estimaciones más altas que los que los que utilizan el enfoque FAP.

Por otra parte, los estudios de costos indirectos utilizan el enfoque de capital humano, el de costo de fracción atribuible a la población y el enfoque de valor de vida estadística^[212]. Esta diversidad de metodologías hace que las estimaciones varíen sustancialmente, incluso dentro del mismo país, según los enfoques analíticos, el marco temporal y otras consideraciones metodológicas^[310]. Por lo tanto, se requiere consultar *a priori* a los expertos en teoría económica antes de plantear un estudio de esta índole, de manera que los datos recopilados puedan ser analizados bajo un contexto de significancia práctica para quienes toman decisiones.

En síntesis, la estimación de la carga económica de la inactividad física es un área de creciente importancia que requiere un mayor desarrollo. Como se ha mencionado, existe una marcada falta de coherencia en los enfoques metodológicos para su estudio. Algunas recomendaciones para futuras investigaciones podrían ser la incorporación de la interdisciplinariedad, en donde el problema de estudio pueda ser planteado por profesionales de economía y de las Ciencias del Movimiento Humano. Esta combinación permitiría tener abordajes que consideren con mayor profundidad la perspectiva social y a la vez sigan las mejores prácticas metodológicas disponibles, como por ejemplo, la consideración de posibles factores de confusión, causalidad inversa y comorbilidad, así como el reporte de limitaciones y justificaciones de los enfoques adoptados^{[212][310]}.

De esta forma, se requieren nuevos estudios en diversos países para determinar el impacto econométrico de las diferentes formas de prescripción de la actividad física, ya sea en el ámbito escolar o en el ámbito de la salud (e.g., gimnasios, clínicas). De la misma manera, los estudios econométricos deberían acompañar los hallazgos de los estudios experimentales que intentan demostrar si determinada forma de entrenamiento o de prescripción del ejercicio posee alguna influencia sobre alguna variable de interés público, como, por ejemplo, la presión arterial, la obesidad, la ansiedad o la depresión, entre otras.

3.3. Placebo y nocebo en la investigación

Los placebos, del latín *placebo* = debo complacer^[115], han sido popularmente utilizados durante muchos años como una herramienta para aliviar el dolor y otras condiciones en los campos médicos y psiquiátricos^[480]. El uso de placebos se ha extendido a otras disciplinas, incluyendo las Ciencias del Movimiento Humano (e.g., fisiología del deporte y el ejercicio, biomecánica, entrenamiento deportivo, psicología deportiva). Posiblemente la mayor atención la tiene el uso de placebos en el área de la nutrición deportiva, donde es una práctica casi obligatoria cuando se diseñan investigaciones acerca de los efectos de determinadas sustancias nutricionales con potencial ergogénico. Los estudios sobre suplementos nutricionales difieren parcialmente de los estudios con medicinas o fármacos; por ejemplo, para que una nueva medicina pueda venderse en EE.UU., la Food and Drug Administration (FDA) requiere que dicha sustancia pruebe ser segura y eficaz para el consumidor. Esto se logra luego de una serie de costosos ensayos clínicos. Por otra parte, los suplementos nutricionales no necesitan probar que son seguros o efectivos ya que no se les considera como medicinas, sino como complementos de la alimentación regular de la persona^[124].

En las Ciencias del Movimiento Humano, no existe una definición clara y contundente acerca de lo que es un placebo, de lo que debería ser, o acerca de cómo debe ser aplicado o administrado en las investigaciones en nutrición deportiva. Desde un punto de vista práctico, se podría decir que un placebo es una intervención diseñada para simular el tratamiento que se supone que produce un cambio en una variable dependiente (e.g., fuerza, potencia, velocidad). Es posible que, para una mayoría de investigadores, el efecto placebo se considere como un resultado positivo que resulta de la creencia de una persona de que una sustancia inerte es, de hecho, una droga activa, una medicina, una pastilla mágica o un tratamiento activo, por lo que se le ve como un fenómeno intrapsíquico^[197]. El propósito del placebo, por consiguiente, es el de “engañar” a los participantes para que produzcan una respuesta deseada psicológica, fisiológica o física en la variable dependiente de interés. A eso se le conoce como el efecto placebo, y se ha encontrado que cerca del 50% de los atletas indica haber experimentado un efecto placebo beneficioso, y un porcentaje similar de entrenadores reporta haber proporcionado placebos a sus atletas^[587].

Desde un punto de vista metodológico, se ha conocido durante mucho tiempo el papel de los placebos con el propósito de reducir sesgos en la investigación, como, por ejemplo, las expectativas de los participantes y de los propios investigadores acerca del posible resultado del estudio, también conocido como el Efecto de Hawthorne^[752]. Sin embargo, se conoce poco acerca de la teoría del uso del placebo en nutrición deportiva, una industria que maneja billones de dólares. En otras palabras, se sabe poco acerca de por qué se usan los placebos, cómo se puede explicar el efecto placebo y cuáles pueden ser los mecanismos fisiológicos, psicológicos o de otro tipo subyacentes de dicho efecto.

Por ejemplo, desde la neurociencia, se han encontrado investigaciones que identifican una serie de vías neurobiológicas y fisiológicas relacionadas a los efectos placebo y nocebo (i.e., efectos adversos que la persona piensa que pueden ocurrir, – sugestión negativa –), que se pueden asociar a aspectos propios del deporte como aquellos afines con el dolor, la fatiga y el control motor^[66]. Los hallazgos de esos estudios sugieren que los placebos y los nocebos dan como resultado la actividad de los sistemas de neurotransmisores opioides, endocannabinoides y dopaminérgicos, regiones del cerebro que incluyen la corteza motora y el cuerpo estriado, y efectos medibles en el

sistema nervioso autónomo. Esos hallazgos sugieren la posibilidad de una amplia gama de mecanismos distintos del placebo y el nocebo que podrían influir en el rendimiento deportivo.

Otros investigadores sugieren que el efecto placebo es el resultado de interacciones sociales^[197]. Por ejemplo, el efecto placebo puede ser explícito, como cuando un investigador describe y administra un tratamiento; o puede ser menos explícito, como con las percepciones del sujeto de investigación sobre la credibilidad, experiencia o confianza en el investigador que realiza el estudio. Así, se describe que muchos fenómenos en el deporte también tienen un origen social (e.g., apoyo de aficionados en el estadio, un equipo cohesionado, ansiedad inducida por un oponente experto, percepción de bajo rendimiento, influencia de los medios de comunicación) y son estudiados desde la psicología social, la fisiología experimental, y la antropología cognitiva y evolutiva^[197].

La literatura es escasa acerca de estudios que describan cómo se han diseñado los placebos que se usan en las investigaciones en nutrición deportiva, ni de cómo se han administrado o evaluado. Incluso, en una revisión de 83 estudios publicados en revistas científicas se encontró que no existe una forma consistente de diseñar o de describir los placebos (Tabla 1).

Tabla 1. Criterios recopilados en publicaciones científicas que describen cómo se diseña un placebo.

Criterio de descripción del placebo	Frecuencia	%
"similar", "idéntico", "mismo", "indistinguible"	23	28
"sabor"	19	23
"apariencia", "forma", "aspecto", "parecido"	13	16
"color"	9	11
"textura", "consistencia"	9	11
"dulzura"	2	2
"verbal"	2	2
"sin calorías", "sin nutrientes"	2	2
"inerte", "sin actividad biológica"	2	2
"mismo peso"	1	1
"suavizado"	1	1
Total	83	100

Es más, en el área médica donde fueron utilizados por primera vez, aún existe un fuerte debate acerca de su eficacia, utilidad e incluso la ética de su administración^{[351][672][703]}. En una amplia gama de estudios clínicos, los placebos han tomado diferentes formas, colores y vías de administración (e.g., oral, intramuscular, intravenosa). Por ejemplo, se han utilizado placebos sólidos (e.g., pastillas, cápsulas), líquidos o en forma de gel, visuales (e.g., luces), sonidos (e.g., música o sonidos), verbales (e.g., apoyo verbal, "hacer porras", "hacer barra", "alentar al participante"), y la sugestión^{[199][481]}.

Los carbohidratos (CHO) han sido, por mucho, las sustancias más ampliamente estudiadas en nutrición deportiva (Tabla 2) y no existe un consenso acerca del uso específico, consistente y sistemático de un tipo de placebo.

Tabla 2. Composición e ingredientes de los placebos encontrados en 48 investigaciones publicadas en revistas científicas.

Sustancia estudiada	Ingredientes del placebo		Frecuencia	%
Carbohidratos (CHO)	Celulosa	Celulosa, Meilcelulosa	2	4.17
	Edulcorantes	Cristal Light®	12	25
		Nutrisweet® (aspartame)		
		Endulzado artificialmente (limón, bebida baja en calorías, sabor a naranja, agua, sabor a cítricos), edulcorante sin CHO		
	CHO	Carbonato de calcio	16	33.33
		Almidones		
		Glucosa		
Maltodextrinas, maltodextrosa				
Lactosa				
Proteínas	Soya	Frijol de Soya	1	2.08
	Suero de leche de vaca	Proteína de suero concentrada en polvo	3	6.25
		Suero en polvo		
Otros	Electrolitos	Picolinato de Cromo	5	10.42
		Cloruro de sodio		
		Sodio		
		Potasio		
		Agua		
Combinaciones	Electrolitos, endulzado artificialmente, Na ⁺ , K ⁺ , sin CHO		9	18.75
	Cápsulas de gelatina, harina, bebida deportiva 1.7% CHO			
	Gelatina en polvo, polvo con sabor a limón, agua			
	Glicina + CHO			
	Maltodextrina, ácido cítrico, aspartame, sabor limón			
	Soya, alfatocoferol + gama tocoferol			
	Edulcorantes, líquido con sabor a naranja, agua destilada			
	Agua con sabor y gelatina			
	Leche en polvo, edulcorando artificial, agua			
Total			48	100

Obviamente el uso de diferentes tipos de placebo contribuye a aumentar la varianza de error y reducir el verdadero efecto de la sustancia de estudio; en otras palabras, la varianza verdadera atribuida al tratamiento experimental. Idealmente, un placebo tiene que ser inerte; es decir, que no posea en sus ingredientes precisamente la sustancia que se está investigando^[768]. A este tipo de placebo se le denomina placebo *puro*. Sin embargo, en muchos estudios se han reportado placebos diseñados con pequeñas cantidades o trazas de la sustancia o suplemento nutricional que está siendo investigado o incluso otras sustancias que supuestamente son inertes. A este tipo de placebo, diseñado con una estructura química o física similar a la sustancia de prueba se le conoce como placebo *activo*^[672]. Algunos ejemplos de placebos activos se muestran en la tabla 2.

Sin embargo, aún permanece la pregunta acerca de ¿qué factor incide en que un placebo sea “apropiado” o “estándar” para una situación en particular en el campo de la nutrición deportiva? Es posible que no exista una respuesta única a esa pregunta ya que se han utilizado muchos tipos de placebos aun cuando se estudia la misma sustancia. Por otra parte, se debería reportar la evaluación sensorial o subjetiva de los placebos en las investigaciones, incluyendo al menos, aspectos como el olor, sabor y textura percibida tanto por los investigadores como por los participantes del estudio^[482].

También se debería brindar información acerca de la dulzura percibida y relativa de los placebos utilizados en investigaciones con carbohidratos^[13]. Por ejemplo, la sucrosa o sacarosa (4 cal/g o 16.74 kJ/g) es considerada el estándar de oro o criterio para comparar otros azúcares^[737]; sin embargo, una debilidad de su utilización podría ser su contribución energética en sujetos que realizan ejercicio; por lo tanto, se debería evitar su uso. La celulosa (Tabla 2), por su parte, es un polímero de glucosa insoluble en agua y forma parte estructural de las plantas. También se le usa para proporcionar “bulto” a los alimentos. La celulosa no proporciona energía debido a su incapacidad para romper los enlaces 1,4-β-glucosídicos y podría usarse como un placebo activo^[482]. Edulcorantes alternativos que se usan en la industria alimenticia incluyen el sorbitol, la sacarina y el aspartame (Tabla 2). Este último es considerado un dipéptido muy dulce y bajo en calorías (4 cal/g o 16.74 kJ/g), compuesto por fenilalanina y ácido aspártico. Este edulcorante se utiliza para aumentar la dulzura de los alimentos, y funciona adecuadamente a un pH de 3 a 5, no tiene olor, tiene un color blanco cristalino, y es parcialmente soluble en agua (~1.0% a 25°C)^{[482][610]}. Por lo tanto, también podría incluirse como un placebo en los estudios relacionados con la nutrición deportiva.

Cuando se diseña un placebo se debe pensar en realizar al menos tres tipos de comparaciones en un diseño de investigación en donde se tiene: a) un nuevo suplemento nutricional (i.e., tratamiento), b) placebo, y c) condición control o de línea base (i.e., no recibe el tratamiento ni el placebo)^[617]. Este sería un diseño muy robusto desde el punto de vista de la metodología de investigación. Así, lo primero que se debería pensar es en comparar el grado en que el nuevo suplemento nutricional es superior al placebo (visto como condición control). Segundo, se debería comparar el placebo con la condición control o línea base; esta es la comparación de interés para quienes están interesados en estudiar el campo del efecto placebo. Finalmente, se puede comparar el nuevo suplemento nutricional y la condición control o de línea base. Sin embargo, al realizar esta comparación no se puede conocer el grado de un posible efecto placebo o, por el contrario, un posible efecto nocebo.

El efecto nocebo se define como “una sustancia o tratamiento inofensivo que, cuando se toma o se administra a un paciente, se asocia con efectos secundarios dañinos o empeoramiento de los síntomas debido a expectativas negativas o a la condición psicológica del paciente” (p. 863)^[5]. En otras palabras, es una respuesta psicológica negativa en la que influyen las expectativas de la persona sobre los efectos adversos al comienzo del tratamiento. El efecto nocebo impacta negativamente en los pensamientos, sentimientos y acciones de las personas, lo cual ocasiona cambios no deseados en el rendimiento físico^[357].

El efecto nocebo también se ve influenciado por el proceso de condicionamiento en el que los participantes aprenden de experiencias previas a asociar una intervención con síntomas somáticos (e.g., levantar pesas en un gimnasio provoca ansiedad en mujeres con pobre imagen corporal), o tener ciertas características psicológicas como ansiedad, depresión y la tendencia a somatizar; además de factores situacionales y contextuales^[5]. Una revisión de literatura estudió el efecto nocebo sobre el rendimiento motor, en donde se encontró un tamaño del efecto de mediano a grande (TE= 0.60) de la intervención de nocebo en el rendimiento motor^[350]. Esto significa que el efecto nocebo influye negativamente en el rendimiento motor y puede evocarse con información verbal negativa^[357].

El efecto nocebo puede ser incluso más robusto que el efecto placebo^[357], pero depende del tipo de rendimiento motor analizado, de la muestra examinada y del agente o la forma en que se presenta el nocebo^[350]. En otra revisión de literatura acerca de ayudas ergogénicas nutricionales y mecánicas^[358], se encontraron tamaños de efecto de pequeños a moderados para el efecto nocebo (TE= 0.37) y cuando se separaron por ayudas ergogénicas nutricionales (TE= 0.35) y mecánicas (TE= 0.47), lo cual sugiere que el efecto nocebo puede ejercer un efecto pequeño a moderado sobre el rendimiento deportivo.

Se deberían utilizar diseños experimentales potentes cuando se estudian nuevos suplementos en nutrición deportiva. Un diseño potente podría incluir las tres condiciones experimentales señaladas anteriormente (i.e., suplemento, placebo, control) y al menos dos mediciones de las variables dependientes de interés (e.g., pre, post)^[434]. Este es un tipo de diseño factorial de medidas repetidas en donde cada participante es aleatoriamente asignado al orden de las condiciones experimentales. En otras palabras, cada sujeto debe completar las tres condiciones experimentales, pero en diferente orden que los demás. En consecuencia, el participante "es ciego" a las condiciones experimentales, y por lo tanto, reduce el sesgo de la investigación. A la vez, sería deseable que el investigador también sea "ciego" al tratamiento que le va a administrar a los participantes. En este caso, el estudio dejaría de ser "ciego simple" y se convertiría en "doble ciego", que forma parte de un diseño aún más potente. Este tipo de diseño tiene la ventaja de que además reduce la posibilidad del efecto de aprendizaje de la prueba que se usa para medir la variable dependiente de interés, y también reduce las expectativas tanto del sujeto de estudio como del investigador. En las publicaciones de nutrición deportiva, más del 85% utilizan diseños de medidas repetidas y mixtos. Estos diseños requieren muestras de menor tamaño que los diseños de grupos independientes. Además, son más apropiados porque permiten estudiar la variabilidad dentro del mismo sujeto y no entre diferentes sujetos.

Utilizar un diseño experimental con tres condiciones (i.e., suplemento, placebo, control) como el descrito anteriormente, permite comparar el efecto del suplemento nutricional con el control, que se estima comparando el cambio en el grupo que consumió el suplemento con el grupo de control. Luego, se puede estudiar el efecto placebo, que se estima comparando el cambio en el grupo placebo con el grupo de control. Finalmente, se puede obtener el verdadero efecto del suplemento nutricional al restar el efecto placebo del efecto observado del suplemento^[434].

Este enfoque de análisis ha permitido, en un meta análisis de estudios en psicología del deporte, encontrar que casi la mitad del efecto observado del ejercicio sobre los resultados psicológicos se atribuyó a los efectos del placebo^[433]. Por lo tanto, diseños robustos con análisis planificados *a priori* proporcionan información más valiosa para la toma de decisiones acerca de un tratamiento, ya sea una intervención de ejercicio, un suplemento nutricional, una bebida deportiva o una intervención psicológica.

Se ha sugerido que tal vez no sea posible tener grupos placebo cuando se estudian las respuestas psicológicas al ejercicio^[434], aunque los placebos tradicionales no siempre son necesarios para estudiar el impacto de los mecanismos psicológicos de los efectos del placebo en las respuestas al tratamiento. Desde una perspectiva metodológica, se ha considerado que la medición de las expectativas positivas y negativas puede ayudar a explicar la variabilidad interindividual en las respuestas psicológicas al ejercicio^[435], y que la modificación y el condicionamiento de las expectativas pueden usarse para mejorar las respuestas al tratamiento y dilucidar los mecanismos neurobiológicos que median la influencia de los efectos placebo y nocebo en estas respuestas^[434].

En síntesis, existe una falta de concordancia entre el diseño y uso de placebos en nutrición deportiva, aun cuando se estudia la misma sustancia. Esta discrepancia añade varianza no deseada a un estudio, lo que aumenta el sesgo y oculta en cierta medida el verdadero efecto de un suplemento nutricional. Lo mismo podría estar ocurriendo en otras áreas de las Ciencias del Movimiento Humano, como en la psicología de la actividad física y el deporte. Por lo tanto, el aislamiento del efecto verdadero de una sustancia se ve gravemente comprometido. Se deben realizar esfuerzos para diseñar placebos puros y activos que sean apropiados, de manera que los estudios en futuros meta análisis puedan ser comparables y reduzcan el sesgo introducido por placebos diseñados deficientemente.

3.4. Modelos animales

Según Akhtar^[10], el uso de animales en la experimentación implica la intervención de tratamientos, la manipulación o alteración de funciones en animales vivos o sacrificados, así como el análisis de tejidos luego de la disección. Estos procesos tienen como objetivo generar conocimiento acerca del funcionamiento del cuerpo humano, o las implicaciones de diferentes fármacos o tratamientos sobre la biología humana o animal (Figura 1). En diversas áreas del conocimiento, la utilización de animales vivos ha progresado y es de suma relevancia en la actualidad; de acuerdo con Maldonado-Villamizar y Aquino-Guerra^[458], los resultados pueden ser extrapolados a los seres humanos, bajo el fundamento técnico de la conservación de las secuelas de ADN, que expresa metabolitos parecidos entre humanos y animales.

Figura 1. Los modelos de investigación animal se utilizan en la experimentación para comprender mecanismos biológicos que podrían extrapolarse al ser humano. Las especies animales utilizadas para la experimentación varían de acuerdo al área de estudio y de las técnicas de análisis de laboratorio empleadas.



La utilización de animales para experimentación, ha permitido grandes y acelerados avances en aspectos de primordial importancia para la salud y bienestar humano e inclusive la de los propios animales^[495]. Sin embargo, se cree que es de imperativa importancia que se consideren aspectos éticos que eviten o disminuyan el sufrimiento de las especies que se someten a estos procesos, lo que además resulta beneficioso para el propio proceso experimental, ya que las dificultades y el estrés experimentado por los animales pueden llevar a tener resultados no confiables^[48].

Aunque se utilizan muchos tipos de animales cuando se trata de experimentar con ellos, son en general los roedores los mayormente utilizados como modelo de enfermedad humana. Según Iannaccone y Jacob^[359], la rata ofrece muchas ventajas sobre el ratón y otros tipos de animales, entre las cuales se pueden mencionar:

- a) El aislamiento exitoso de células madre derivadas de embriones de rata se acelerará rápidamente.
- b) Hay una variedad de cepas genéticas que son ideales para estudiar la enfermedad cardiovascular, accidente cerebrovascular e hipertensión.
- c) La fisiología es más fácil de monitorear y se parece más a la condición humana, incluso cuando se trata de temas de cognición, memoria, aprendizaje, por lo tanto, es capaz de realizar una variedad más amplia de tareas que otros animales.
- d) Su tamaño, permite realizar procedimientos y utilizarlas como modelos de enfermedad por el tamaño proporcional de subestructuras importantes en órganos vitales y el efecto de fármacos en áreas anatómicas específicas y de gran importancia como el sistema nervioso central.
- e) Son sensibles a hormonas con histopatología y tienen etapas premalignas que se asemejan más a la enfermedad humana.

Aun considerando las ventajas y los beneficios que tiene la utilización de biomodelos animales especialmente en el ámbito de la salud, existe mucha controversia y un movimiento social importante que critica fuertemente la experimentación y que genera dudas sobre el valor predictivo de sus resultados. Además, algunas veces se ha ridiculizado el gasto utilizado en artefactos para medir variables de interés en animales. Por ejemplo, en los EE.UU., hubo una discusión acalorada en los medios de comunicación y la comunidad científica, debido a la compra de una banda sin fin para medir la caminata en camarones (Ver video en: <https://www.youtube.com/watch?v=wrhRNMgOLYo>)^[295]. En la generalidad de cada área, son los propios investigadores, los más preocupados por el bienestar de los animales con los que trabajan y rigen su actuación bajo estrictos principios bioéticos, uno de ellos conocido como "las 3 Rs". El marco ético conocido como las "3Rs", que corresponden a las palabras **Reemplazo** (i.e., solo se utilizarán animales si no hay otra alternativa), **Reducción** (i.e., se utilizará el menor número posible de ellos) y **Refinamiento** (i.e., se buscará siempre el mayor bienestar animal, tanto durante el experimento como en su estancia en el laboratorio, minimizando cualquier dolor o estrés que el animal pueda sentir), ayudará a asegurar que los estándares de atención, cuidado y manejo de los animales, siempre mejore (<https://www.animalresearch.info/en/designing-research/alternatives-and-3rs/>).

La aplicación del marco de las 3R es obligatoria por ley en algunos países como el Reino Unido y la Unión Europea siempre que se autorice el uso de animales en laboratorios. En el Reino Unido, por ejemplo, el Centro Nacional para el Reemplazo, el Refinamiento y la Reducción de Animales en la Investigación (NC3R) ayuda a la comunidad investigadora a utilizar la ciencia y la tecnología más recientes para reemplazar los estudios con animales.

Si bien es cierto, son las áreas médicas, biológicas y farmacéuticas las que mayormente utilizan biomodelos animales para generar investigación básica, el área de las Ciencias del Movimiento Humano también se ha beneficiado a lo largo del tiempo de esta práctica; algunas investigaciones recientes que impactan la práctica del ejercicio y el deporte se describen en la tabla 1.

Tabla 1. Ejemplo de investigaciones en las cuales se utilizan modelos animales.






Referencia	Área de impacto	Animales utilizados	Objetivo del estudio
Ruiz-Iglesias et al. ^[622]	Deporte y Nutrición	Ratas Wistar hembras (4 semanas de edad) 	Evaluar el efecto de una dieta enriquecida con cacao, sobre el rendimiento físico, la producción de especies reactivas de oxígeno por los macrófagos peritoneales y la función inmunológica sistémica de ratas sometidas a ejercicio agudo intenso.
Villaça et al. ^[723]	Ejercicio Físico y Diabetes	Ratones machos C57Bl/6 (6-7 semanas de edad) 	Evaluar el papel potencial del ejercicio físico como terapia coadyuvante para la función y preservación de las estructuras de los islotes de Langerhans (i.e., células del páncreas que producen insulina) en un modelo de ratón a los cuales se les provocó Diabetes tipo 1, para entender los mecanismos involucrados en estos procesos, además, se analizaron marcadores de proliferación, reparación, diferenciación y muerte celular en estos islotes.
			Continúa...

Tabla 1. Continuación.

Referencia	Área de impacto	Animales utilizados	Objetivo del estudio
Ferreira et al. ^[263]	Ejercicio Físico y Enfermedad de Parkinson	Roedores y primates (múltiples edades) 	Esta es la primera revisión sistemática que recopila toda la literatura sobre el uso de ejercicios en banda sin fin, incluidos todos los resultados que se han publicado. Identificando protocolos de ejercicio en cinta rodante, los resultados anteriores que se encuentran en modelos animales de la enfermedad de Parkinson.
Shirai et al. ^[647]	Hipertrofia muscular	Ratones machos ICR (7-8 semanas de edad) 	Examinar el efecto de entrenamiento de resistencia cardiovascular de larga y corta duración en la hipertrofia musculoesquelética de ratones, inducida por sobrecarga funcional a nivel molecular.
Vasconcelos et al. ^[719]	Accidente cerebrovascular y Ejercicio de alta intensidad	Ratones Wistar machos (Rattus norvegicus, var. Albinus sin reporte de edad) 	Comparar los efectos de dos protocolos de ejercicio físico de alta intensidad, estandarizados para la resistencia cardiovascular y la ganancia de fuerza, en ratas entrenadas antes de la isquemia cerebral inducida por la oclusión bilateral de la arteria carótida común.

En síntesis, se vislumbra un reto importante para las universidades en términos de la preparación académica de quienes desean investigar en el área de las Ciencias del Movimiento Humano utilizando modelos animales, ya sea para comprender respuestas y adaptaciones al entrenamiento físico para el rendimiento deportivo, o para comprender procesos moleculares para la salud humana. La colaboración inter y transdisciplinaria será capaz de resolver esos retos, ya que se requiere de diferentes aristas para comprender y resolver preguntas de investigación complejas. Además, se requiere de importante financiamiento para obtener los animales, alimentarlos, entrenarlos, y finalmente, utilizarlos (i.e., sacrificarlos) para los fines de la investigación evitando toda crueldad.

3.5. Diseño y publicación de protocolos de investigación

En un intento para reducir la duplicidad de esfuerzos en la investigación, reducir costos, así como para proporcionar indicadores de transparencia y promover la ética científica, se ha estimulado una cultura de publicación de protocolos de investigación. En términos simples, los protocolos permiten observar el sustento teórico y la metodología propuesta y revisada por pares académicos, con el propósito de responder a una pregunta de investigación. Cada vez existen más revistas científicas que estimulan la publicación de protocolos de investigación; algunas de éstas se presentan en la tabla 1.

Tabla 1. Lista no exhaustiva de revistas en donde se pueden publicar protocolos de investigación en las Ciencias del Movimiento Humano.

Título de la Revista	Objetivo
Systematic Reviews	Publicar revisiones sistemáticas de alta calidad e incluye <i>protocolos de revisión sistemática</i> , revisiones sistemáticas relacionadas con una definición muy amplia de salud, revisiones rápidas, actualizaciones de revisiones sistemáticas ya completadas y métodos de investigación relacionados con la ciencia de las revisiones sistemáticas, como el modelado de decisiones.
BMJ Open Sport & Exercise Medicine	Publicar investigaciones originales, revisiones (en particular, revisiones sistemáticas) y <i>protocolos de estudio</i> que se consideren coherentes y técnicamente sólidos, asegurando que las últimas investigaciones se revisen y difundan rápidamente a una audiencia global.
Plos One	Publicar investigaciones originales de las ciencias naturales, la investigación médica, la ingeniería, así como las ciencias sociales y humanidades relacionadas, que incluyen: <ol style="list-style-type: none"> a. Investigación primaria que contribuya a la base del conocimiento científico, incluyendo estudios interdisciplinarios, de replicación y resultados negativos o nulos. b. Revisiones sistemáticas cuyos métodos aseguren el muestreo completo e imparcial de la literatura existente. c. Envíos que describen métodos, software, bases de datos u otras herramientas que cumplen con los criterios de utilidad, validación y disponibilidad de la revista. d. Investigación cualitativa que se adhiere al diseño de estudio adecuado y a las pautas de presentación de informes. e. <i>Protocolos</i>, incluidos protocolos de laboratorio que describen metodologías verificadas y <i>Protocolos</i> de estudio que describen planes detallados para proyectos de investigación.
	Continúa...

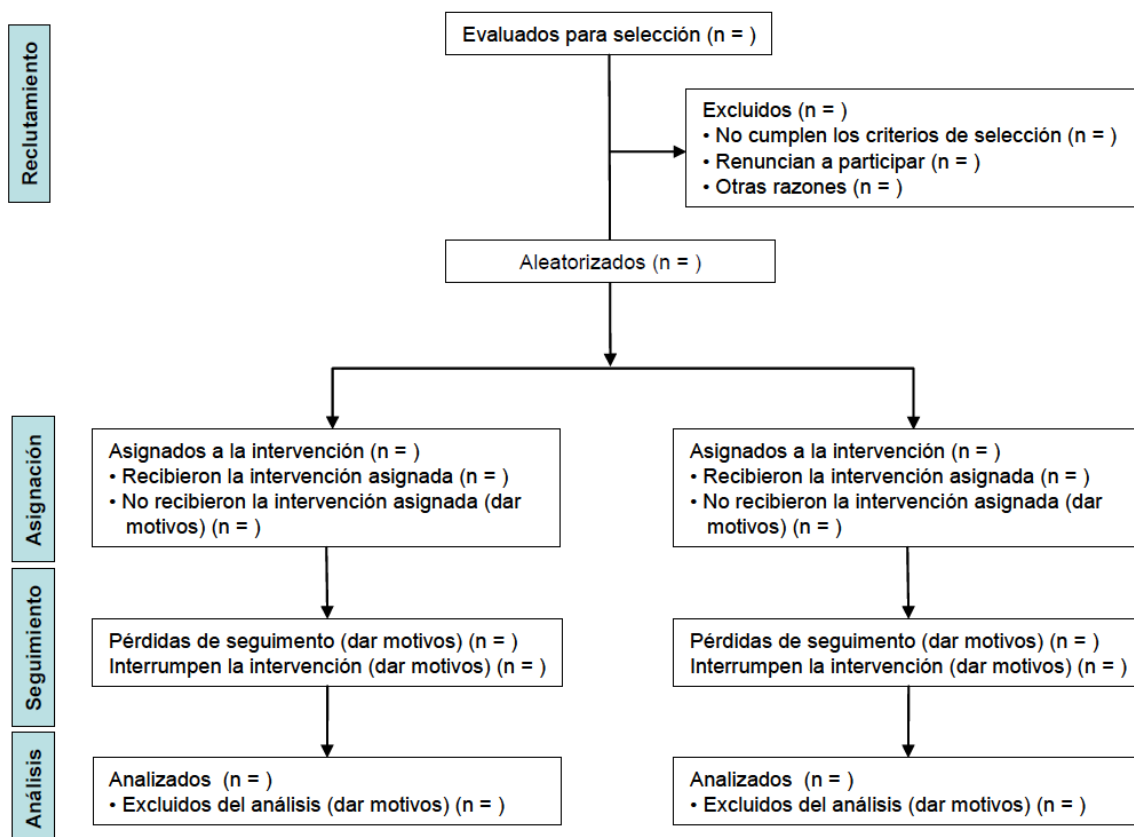
Tabla 1. Continuación.

Título de la Revista	Objetivo
BMC Geriatrics	Publicar artículos de investigación originales revisados por pares en todos los aspectos de la salud y la atención médica de las personas mayores, incluidos los efectos de los sistemas y políticas de atención médica. La revista también acepta investigación centrada en el proceso de envejecimiento, incluidos los procesos celulares, genéticos y fisiológicos y las modificaciones cognitivas. Acepta cualquier artículo para una colección editada por invitados, reporte de un caso, correspondencia, base de datos, nota de datos, software y <i>protocolo</i> de estudio.
Medicine	<p>Proporcionar a los autores una publicación continua de investigaciones originales en un amplio espectro de disciplinas y subespecialidades científicas médicas. El proceso de revisión de Medicine enfatiza la validez científica, técnica y ética de las presentaciones. La novedad o el potencial de impacto no se consideran durante la evaluación o adjudicación del manuscrito. Se aceptan los siguientes tipos de artículos:</p> <ol style="list-style-type: none"> Ensayo clínico/estudio experimental (conforme a las guías de CONSORT). Estudio observacional (compatible con las guías de STROBE). Revisión sistemática y meta análisis (cumple con las guías de PRISMA). Estudio de precisión diagnóstica (compatible con las guías de STARD). Estudio de Evaluación Económica (Cumple con las guías de CHEERS). Informe de caso clínico (cumple con las guías CARE). Revisión narrativa (de acuerdo con las guías SANRA). <i>Protocolo</i> de estudio de ensayos clínicos (cumple con las guías SPIRIT). <i>Protocolo</i> de estudio de revisión sistemática y meta análisis (cumple con las guías de PRISMA-P).

La estructura de un protocolo va a depender del tipo de investigación; es decir, si el estudio que se planea realizar es pre-experimental, experimental, una revisión de literatura narrativa, sistemática o meta analítica, entre otros. Para ello, se recomienda seguir guías en las cuales se indican los ítemes mínimos que debería contener cada una de ellas. Por ejemplo, las guías "Normas consolidadas para reporte de ensayos" (CONSORT) tienen como objetivo elaborar recomendaciones para que la publicación de ensayos clínicos sea más completa y comprensiva (Tabla 2)^{[191][247]}.

Estas recomendaciones se traducen en una lista de comprobación o de verificación (i.e., lista de verificación o "*checklist*") y un diagrama de flujo (i.e., "*flow diagram*") (Figura 1). Otras guías que se utilizan incluyen STROBE^[729], PRISMA^[34], las guías "*Standards for Reporting of Diagnostic Accuracy Studies*" (STARD)^[179], las guías "*Consolidated Health Economic Evaluation Reporting Standards*" (CHEERS)^[258], las guías "*CAsE REport*" (CARE)^[278], el instrumento "*Scale for Assessment of Narrative Review Articles*" (SANRA)^[46], y las guías "*Standard Protocol Items: Recommendations for Interventional Trials*" (SPIRIT)^[156].

Figura 1. Diagrama de flujo del *Consolidated Standards of Reporting Trials* (CONSORT) del progreso a través de las fases de un ensayo clínico aleatorizado paralelo de dos grupos (i.e., reclutamiento, asignación de la intervención, seguimiento y análisis)^[191].



También existen sitios en los cuales se pueden inscribir o registrar protocolos de investigación y hacerlos públicos a la comunidad internacional. Por ejemplo, para temas relacionados con la salud en áreas como la actividad física, el ejercicio y el deporte se recomienda inscribir el protocolo en el International Prospective Register of Systematic Reviews (PROSPERO, <https://www.crd.york.ac.uk/prospere/>). También se pueden utilizar otras plataformas como el Joanna Briggs Institute (<https://jbi.global/systematic-review-register>), el Open Science Framework (OSF, <http://osf.io/>), el Research Registry (<https://www.researchregistry.com/>), el International Platform of Registered Systematic Review and Meta-analysis Protocols (INPLASY, <https://inplasy.com/>) o el protocols.io (<https://www.protocols.io/>)^[567].

Tabla 2. Lista parcial de comprobación de *Consolidated Standards of Reporting Trials* (CONSORT) para un protocolo de un estudio clínico aleatorizado^[191].

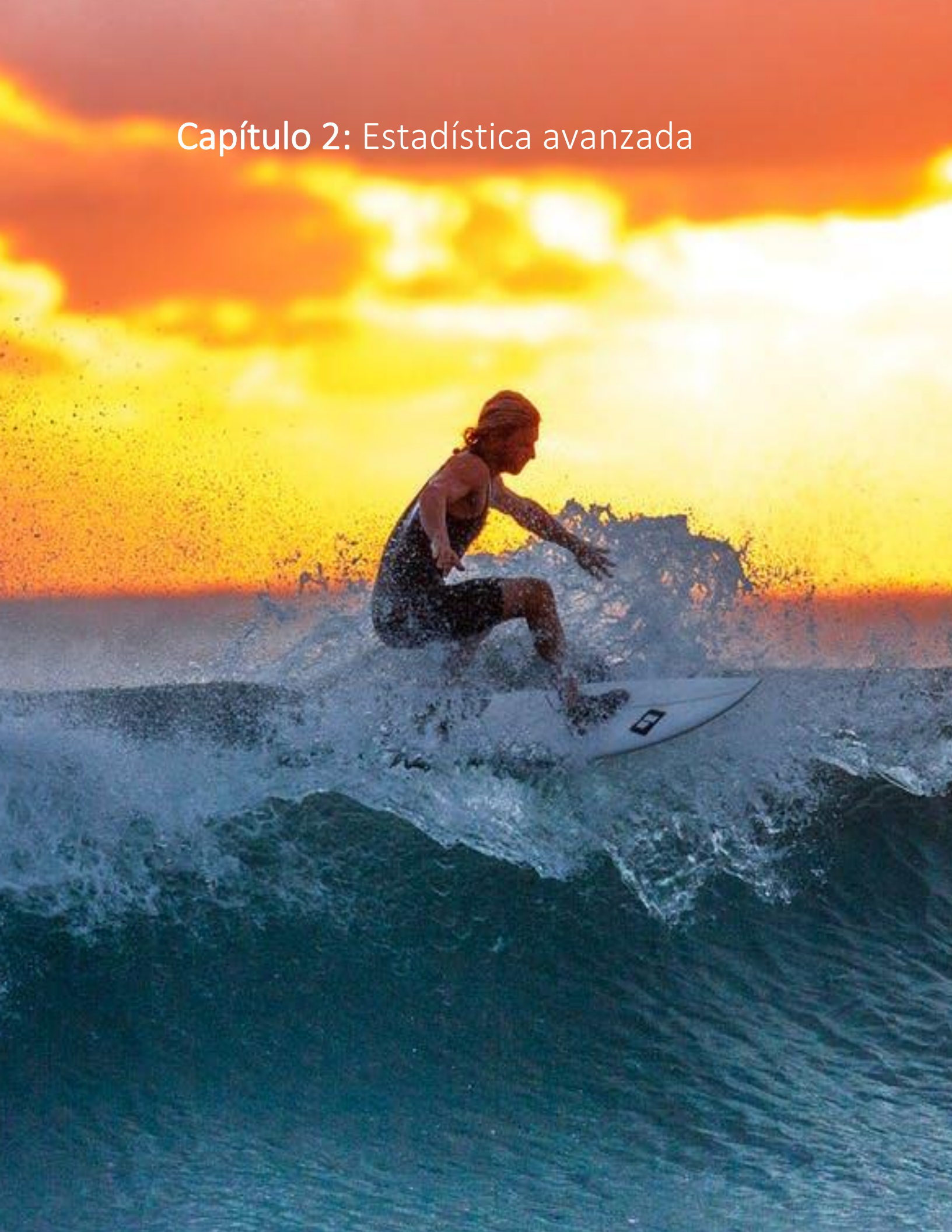
Sección/tema	Ítem nº	Ítem de la lista de comprobación	Reportado en página nº
Título y resumen	1a	Identificado en el título como un ensayo aleatorizado.	
	1b	Resumen estructurado del diseño, métodos, resultados y conclusiones del ensayo (para más detalles, ver "CONSORT for abstracts").	
Introducción			
Antecedentes y objetivos	2a	Antecedentes científicos y justificación.	
	2b	Objetivos específicos o hipótesis.	
Métodos			
Diseño del ensayo	3a	Descripción del diseño del ensayo (e.g., paralelo, factorial), incluida la razón de asignación.	
	3b	Cambios importantes en los métodos después de iniciar el ensayo (e.g., criterios de selección) y su justificación.	
Participantes	4a	Criterios de selección de los participantes.	
	4b	Procedencia (centros e instituciones) en que se obtuvieron los datos.	
Intervenciones	5	Las intervenciones para cada grupo con detalles suficientes para permitir la replicación, incluidos cómo y cuándo se administraron realmente.	
Resultados	6a	Especificación <i>a priori</i> de las variables respuesta (o desenlace) principal(es) y secundarias, incluidos cómo y cuándo se evaluaron.	
	6b	Cualquier cambio en las variables respuesta tras el inicio del ensayo, junto con el motivo de la(s) modificación(es).	
Tamaño muestral	7a	Cómo se determinó el tamaño muestral.	
	7b	Si corresponde, explicar cualquier análisis intermedio y las reglas de interrupción.	
Aleatorización:			
Generación de la secuencia	8 ^a	Método utilizado para generar la secuencia de asignación aleatoria.	
	8b	Tipo de aleatorización; detalles de cualquier restricción (como bloques y tamaño de los bloques).	
			Continúa...

Tabla 2. Continuación

Sección/tema	Ítem n°	Ítem de la lista de comprobación	Reportado en página n°
Mecanismo de ocultación de la asignación	9	Mecanismo utilizado para implementar la secuencia de asignación aleatoria (como contenedores numerados de modo secuencial), describiendo los pasos realizados para ocultar la secuencia hasta que se asignaron las intervenciones.	
Implementación	10	¿Quién generó la secuencia de asignación aleatoria, quién seleccionó a los participantes y quién asignó los participantes a las intervenciones?	
Enmascaramiento	11a	Si se realizó, ¿a quién se mantuvo cegado después de asignar las intervenciones (e.g., participantes, cuidadores, evaluadores del resultado) y de qué modo?	
	11b	Si es relevante, descripción de la similitud de las intervenciones.	
Métodos estadísticos	12a	Métodos estadísticos utilizados para comparar los grupos en cuanto a la variable respuesta principal y las secundarias.	
	12b	Métodos de análisis adicionales, como análisis de subgrupos y análisis ajustados.	

En síntesis, en las Ciencias del Movimiento Humano se han publicado protocolos de estudios en diversas temáticas; como, por ejemplo, en aspectos bioquímicos^[688], composición corporal^[555], factores asociados a problemas de apuestas^[73], comportamiento sedentario en el trabajo^[209], ejercicio y salud ósea^[215], lesiones deportivas^[266], recuperación muscular en adultos mayores después de realizar ejercicio contra resistencia^[325], métodos para el análisis de actividad física usando acelerometría^[601], y entrenamiento físico y diabetes tipo II^{[37][377]}. Los protocolos pueden incluir en su estructura la justificación teórica, el propósito del estudio, las hipótesis o posibles resultados esperados. Además, algunos incluyen un estudio de la potencia estadística para determinar el tamaño de la muestra necesaria para poder realizarse exitosamente, así como los procedimientos para la inclusión y exclusión de los participantes y los procedimientos que se utilizarán para recolectar datos. Otros protocolos lo que describen es la metodología para llevar a cabo una revisión de literatura o un meta análisis, por lo que nuevamente, se siguen guías o recomendaciones previamente establecidas en la literatura.

Capítulo 2: Estadística avanzada



Introducción

El rápido avance de la informática, el aumento en la capacidad de los procesadores de las computadoras y el desarrollo de nuevos programas (“software”) especializados para el análisis de datos han permitido la utilización de técnicas estadísticas novedosas para analizar datos que no podrían ser estudiados con técnicas tradicionales. Adicionalmente, la propagación de programas estadísticos gratuitos también ha permitido la comunicación y uso de códigos o sintaxis por miles de usuarios alrededor del mundo.

En este capítulo, se presentan diversas técnicas de análisis estadísticos, así como debates acerca de la manera en que comprendemos las pruebas de hipótesis y la “significancia estadística” en su forma dicotómica más tradicional (i.e., estadísticamente significativo vs. estadísticamente no significativo) y sus ventajas y desventajas^{[21][61]}. Por ejemplo, se ha indicado que las estimaciones estadísticamente significativas y no significativas están sesgadas en magnitud^[21] e incluso se recomienda abandonar esa dicotomización^[741]. Otro aspecto negativo de usar el enfoque tradicional y rígido de la significancia estadística es que incita a los investigadores a elegir datos y métodos que brinden significancia estadística para algún resultado deseado o no deseado con fines de publicación, lo cual produce un sesgo en las conclusiones.

Otro aspecto relevante para el avance de la investigación en las Ciencias del Movimiento Humano es el del tamaño de la muestra que señalan revisores, editores y asesores en estadística de revistas importantes, quienes indican que existen abundantes estudios que producen ruido porque tienen pocos participantes y baja potencia estadística^[61]. Por ejemplo, existe una discusión acerca de los tamaños de muestra que puedan ser representativos para investigaciones realizadas en atletas de élite (e.g., mujeres atletas de élite); sin embargo, precisamente por la naturaleza de dicha población, las muestras a menudo son pequeñas, lo cual implica un riesgo inminente de rechazo *a priori* en las revistas científicas^[122].

Es por ello que se ha promovido la utilización de diseños de investigación experimentales no tradicionales porque en el deporte de élite, “lo más importante son los promedios individuales, no los promedios grupales” (p. 2029)^[122]. Esto también puede ser relevante para áreas específicas y poblaciones especiales en donde la cantidad de personas con la condición sea excepcionalmente reducida.

La creciente exigencia de las revistas de alta calidad también genera cambios en la forma de analizar y de reportar los resultados de las investigaciones. Por ejemplo, existen revistas en las cuales se exige que los gráficos en los que se resumen promedios también reporten los datos de cada individuo cuando el tamaño de la muestra sea < 30 personas, que se presente la desviación estándar (i.e., \pm SD) y no el error estándar de la media (i.e., \pm SEM), que se reporte el valor de significancia exacto (e.g., $p = 0.034$) y no aproximado (i.e., $p \leq 0.05$), así como la cantidad exacta (i.e., n) de cada grupo de personas analizadas^[267].

También existen recomendaciones para que se reporten medidas del tamaño de efecto de una intervención rutinariamente utilizando formas contextualizadas para la interpretación de los mismos, incluso añadiendo intervalos de confianza, intervalos de predicción y reportando la confiabilidad de los tamaños de efecto^{[273][280]}.

Las guías o listas de verificación para el reporte de estadísticas también son recomendables para mejorar la forma en que se utilizan los análisis estadísticos. En las ciencias médicas existen guías desde hace algún tiempo^{[461][462]}. Por ejemplo, la “Lista de Verificación para la Evaluación Estadística de Artículos Médicos” (CHAMP, por sus siglas en inglés de **C**hecklist for statistical **A**ssessment of **M**edical **P**apers) contiene 30 ítems que abarcan las etapas de diseño, análisis, reporte e interpretación para ayudar a la revisión por pares de un trabajo enviado a una revista; pero además le ayuda a los investigadores a ser más conscientes del tipo de reporte que deben realizar al enviar sus manuscritos a las revistas^{[461][462]}.

Estas listas no son exhaustivas, sino que cubren la mayoría de aspectos mínimos que deberían aparecer en un artículo publicado en una revista científica. Una lista parcial de ítems de CHAMP se presenta en la tabla 1, específicamente para los ítems relacionados al análisis estadístico.

Tabla 1. Lista parcial de ítemes del Checklist for statistical assessment of medical papers (CHAMP) que abarca aspectos de análisis de datos^{[461][462]}.

Ítem	Comentario
Descripción correcta y completa de los métodos estadísticos.	<ul style="list-style-type: none"> • Se debe dedicar una parte separada en la sección de Métodos del manuscrito a la descripción de los procedimientos estadísticos. • Tanto los métodos estadísticos descriptivos como los analíticos inferenciales deben estar suficientemente descritos para que los métodos puedan ser evaluados por un revisor estadístico para juzgar su idoneidad y exhaustividad para abordar los objetivos del estudio.
Descripción de métodos estadísticos válidos utilizados y sus supuestos.	<ul style="list-style-type: none"> • La validez de los análisis estadísticos se basa en algunos supuestos. Por ejemplo: <ul style="list-style-type: none"> ○ Prueba t-student: independencia de las observaciones, normalidad y homogeneidad de la varianza. ○ Prueba χ^2: todos los valores esperados deben ser >1 y, como máximo, el 20% de los valores esperados pueden ser <5. • Evaluar supuestos con métodos como un gráfico de probabilidad normal para comprobar el supuesto de normalidad. • Aplicar una prueba estadística alternativa si se violan algunos supuestos. • Algunas pruebas estadísticas como ANOVA son sólidas frente a violaciones de leves a moderadas de algunos supuestos.
Evaluación adecuada del efecto del tratamiento o interacción entre el tratamiento y otra covariable.	<ul style="list-style-type: none"> • Usar pruebas estadísticas apropiadas para evaluar los efectos del tratamiento y las posibles interacciones. • La evaluación de los intervalos de confianza (IC) específicos de los grupos de tratamiento superpuestos puede ser engañosa. Por lo tanto, la comparación de los IC de los grupos de tratamiento no debe usarse como una alternativa a la prueba estadística del efecto del tratamiento. • Comparar los valores de p para el efecto del tratamiento en cada nivel de la covariable (e.g., hombres y mujeres) no debe usarse como alternativa para una prueba de interacción entre el tratamiento y la covariable. Por ejemplo: <ul style="list-style-type: none"> ○ Observar un valor de $p < 0.05$ en hombres y un valor de $p > 0.05$ en mujeres, se podría concluir erróneamente que el sexo fue un modificador del efecto. ○ No se puede concluir que no hay modificación del efecto si los IC de los subgrupos se superponen.
	Continúa...

Tabla 1. Continuación.

Ítem	Comentario
Uso correcto de las pruebas estadísticas de correlación y asociación.	<ul style="list-style-type: none"> • La correlación de Pearson no debe utilizarse para evaluar la concordancia entre dos métodos en estudios de comparación de métodos. Se recomiendan alternativas como el coeficiente de concordancia de Lin^[432].
Manejo adecuado de predictores continuos.	<ul style="list-style-type: none"> • Desconfiar de los estudios que han dicotomizado o categorizado las variables continuas; por lo general, esto debe evitarse. • El sesgo, la ineficiencia y la confusión residual también pueden resultar de la dicotomización/categorización de una variable continua y su uso como variable categórica en un modelo.
Los IC no deben incluir valores imposibles.	<ul style="list-style-type: none"> • Un IC válido debe excluir valores imposibles.
Comparación adecuada de las características iniciales entre las ramas de un estudio en ensayos aleatorios.	<ul style="list-style-type: none"> • En un ensayo clínico aleatorizado, cualquier diferencia en alguna característica inicial entre los grupos debería explicarse por el azar. • Reportar los valores de p no tiene sentido. La decisión sobre qué características iniciales se incluyen en cualquier ajuste debe especificarse previamente en el protocolo y debe basarse en el conocimiento del tema, no en los valores de p. • Las diferencias entre los grupos en las características iniciales deben identificarse por su tamaño y discutirse en términos de posibles implicaciones para la interpretación de los resultados.
Evaluación correcta y ajuste de variables confusoras.	<ul style="list-style-type: none"> • Un objetivo de la investigación en salud es obtener una inferencia causal (efecto causal de un tratamiento sobre el resultado). • La principal fuente de sesgo que amenaza los estudios de causalidad, incluidos los estudios observacionales y los estudios aleatorizados (con tamaños de muestra pequeño a moderado), son las variables confusoras. • Las variables confusoras se pueden controlar en la fase de diseño del estudio o en la fase de análisis de los datos (i.e., control por diseño vs. control estadístico).
Evitar la extrapolación de un modelo no respaldado por los datos.	<ul style="list-style-type: none"> • El objetivo de interés en muchos estudios de salud es predecir un resultado a partir de una o más variables explicativas utilizando un modelo de regresión. El modelo es válido solo dentro del rango de datos observados en las variables explicativas, y no se pueden hacer predicciones para personas fuera del rango. Esto se conoce como extrapolación del modelo.
	Continúa...

Tabla 1. Continuación.

Ítem	Comentario
Manejo adecuado de datos perdidos.	<ul style="list-style-type: none">• Los métodos utilizados para manejar los datos que faltan deben describirse y justificarse en relación con los supuestos establecidos sobre los datos que faltan (que faltan completamente al azar, que faltan al azar y que no faltan al azar).• Se deben realizar análisis de sensibilidad si corresponde.• Los datos faltantes pueden introducir un sesgo de selección y deben manejarse utilizando métodos apropiados, como la imputación múltiple y la ponderación de probabilidad inversa.• Los métodos ingenuos (i.e., naive) como el análisis de casos completos, la imputación única utilizando la media de los datos observados, la última observación transferida y el método del indicador faltante son estadísticamente inválidos en general y pueden conducir a un sesgo grave. Sin embargo, se estimula a los lectores a consultar otras opiniones al respecto^{[510][547]}.

Otro aspecto relevante acerca de la exigencia cada vez mayor de la presentación de los datos se debe a que se ha descrito que la mayoría de los hallazgos reportados en investigación son falsos, irrepetibles y sesgados debido que utilizan análisis estadísticos erróneos o reportan tendencias cuando la $p = 0.10$ ^{[280][461][462]}. Estas situaciones atentan contra la calidad de las investigaciones publicadas en diversas áreas del conocimiento, incluyendo la medicina y las ciencias del deporte, lo que además impide la réplica de los estudios para corroborar la consistencia de los hallazgos^[589]. Es por ello que se ha recomendado a los editores de las revistas científicas que integren a una persona profesional en estadística durante los procesos de revisión de manuscritos, ya que se sabe que ninguna publicación, aunque haya sido publicada en las mejores revistas, es inmune al error. Incluso, actualmente se recomienda estimular la colaboración entre investigadores para que dentro de los equipos se incluyan personas con formación en bioestadística, estadística, ciencia de los datos, análisis de datos, epidemiología, matemática, ciencias de la computación o economía que sean expertas en el manejo y análisis de datos^[754]. De hecho, se ha encontrado que menos del 14% de los equipos de investigación que aparecen en una revisión de 8970 publicaciones incluyen al menos a una persona con esos perfiles como coautores^[628].

Debido a que actualmente se posee la capacidad para recolectar cantidades enormes de datos con algunos dispositivos electrónicos, cada vez es más necesaria la participación de expertos en el manejo de "big data" y métodos computacionales potentes para el análisis de datos. Algunas de estas técnicas se basan en el aprendizaje de máquinas ("machine learning"), como, por ejemplo, los árboles de decisión ("decision tree"), máquinas vectoriales de soporte ("support vector machines", SVM), k vecinos más cercanos ("k-nearest neighbors", k-NN), algoritmo ingenuo de Bayes (naive Bayes), y partición de datos ("data splitting")^{[120][313]}. Por ejemplo, con la técnica de partición de datos, la totalidad de los datos disponibles se dividen aleatoriamente en dos partes; un conjunto de entrenamiento y un conjunto de prueba. El conjunto de prueba está oculto y el conjunto de entrenamiento se usa para entrenar o ajustar el modelo o algoritmo predictivo. Una vez que se desarrolla ese algoritmo o modelo, el conjunto de prueba se saca y se ejecuta a través del algoritmo para obtener una evaluación "honesta" de su desempeño sobre datos que nunca ha visto^[120]. Este procedimiento permite evaluar el modelo y obtener las estadísticas para responder a la pregunta de investigación.

Muchas de las técnicas de análisis estadístico provienen precisamente de la estadística, pero otras han sido desarrolladas y aplicadas en otras áreas del conocimiento. Por ejemplo, el reconocimiento de patrones utilizado en áreas del conocimiento como la ingeniería eléctrica podría colaborar con el análisis inteligente de datos. Sin embargo, se debe considerar que muchos de los métodos de análisis están apenas en su apogeo, por lo que pueden estar sesgados y necesiten revisarse tanto matemática como estadísticamente. Incluso, recientemente se promueve enseñar y comprender el "razonamiento estadístico" más que enfocarse en el cálculo o el uso de programas computacionales^{[669][700]}. En lo que se conoce como "la era post $p < 0.05$ ", la ganancia de esa sabiduría para razonar permite comprender con mayor profundidad el fenómeno de estudio, y a la vez, elegir los métodos de análisis más apropiados para evitar el error estadístico en cálculo y en interpretación (e.g., correlaciones espurias, ver Taylor et al.^[687]); es decir, se deben enseñar los aspectos conceptuales para posteriormente comprender los datos analizados para evitar sobreinterpretar las estadísticas^[669].

2.1. Los enfoques frecuentista y Bayesiano

El enfoque o probabilidad frecuentista se refiere a la posibilidad de no diferencia, y la premisa de partida es una población con características conocidas. Así, considerar lo que se encontraría cuando se toman muestras repetidas es el enfoque frecuentista de las pruebas estadísticas^[227]. Este es el enfoque predominante y tradicional que se enseña en los cursos introductorios de estadística inferencial, en donde a partir de una pregunta de investigación se plantean hipótesis o posibles respuestas a dicha pregunta. Posteriormente, se recolectan datos y se someten a pruebas de hipótesis por medio de técnicas de análisis estadísticos paramétricos o no paramétricos, dependiendo de la forma de la distribución de los puntajes.

En contraste, la teoría de Bayes, llamada así en honor al matemático inglés Thomas Bayes, se refiere a la modificación de una posibilidad preexistente, en donde los experimentos revisan la estimación anterior^[230]. Esta teoría vincula la probabilidad de un evento X dado Y con la probabilidad del evento Y dado X ; por ejemplo, si se sabe la probabilidad de ser fuerte dado que se entrena 3 días/semana, se podría saber, con la ayuda de más datos, la probabilidad de entrenar 3 días/semana si se es fuerte. Existen dos enfoques en las inferencias bayesianas: a) el enfoque empírico de Bayes, y b) el enfoque jerárquico de Bayes, conocido como el enfoque de Bayes completo. Ambos difieren en cómo se especifican los hiperparámetros; es decir, los parámetros en una distribución anterior^[430].

En el enfoque empírico de Bayes, los hiperparámetros *a priori* son valores fijos o estimaciones puntuales, determinados por cantidades agregadas, como las medidas de momentos de estudios relacionados. En la práctica, los valores de los hiperparámetros rara vez se conocen *a priori* a partir de muestras limitadas y es posible que difícilmente reflejen las distribuciones reales de los parámetros del modelo, lo que podría conducir a inferencias bayesianas distorsionadas^[430].

Por su parte, en el enfoque jerárquico de Bayes, la incertidumbre de los hiperparámetros se puede modelar aún más a través de propios valores previos con parámetros adicionales (i.e., hipervalores previos). Así, un aspecto clave en la estimación bayesiana es la determinación de un previo informativo efectivo (i.e., hiperprevios). En los modelos bayesianos jerárquicos, la incertidumbre de los hiperparámetros en un previo se puede modelar aún más a través de sus hiperprevios^[430].

Para usar el teorema de Bayes se requiere conocer el concepto de verosimilitud, que es la probabilidad de los datos en función de los valores de los parámetros, pero que considera que los datos son fijos y que los valores de los parámetros potenciales varían. Es decir, la función de verosimilitud es una función de los parámetros de un modelo estadístico que permite realizar inferencias acerca de su valor a partir de un conjunto de observaciones^[227].

Así, para desarrollar el enfoque bayesiano, primero se debe predecir una posibilidad o distribución previa^[399]. Esta predicción puede estar basada en una conjetura del experimentador, o en la opinión de un grupo de expertos, o es más probable que provenga de una serie de experimentos anteriores, para obtener una idea de la posible población. Luego, los datos experimentales recolectados se utilizan para modificar la distribución previa. Posteriormente, se genera una estimación revisada de los valores de la población, que es la distribución posterior. Los

efectos a menudo se expresan en términos del "intervalo creíble" para cuantificar el parámetro de población, que es un concepto similar al de intervalo de confianza.

Este enfoque utiliza los datos experimentales para definir con mayor precisión una población, y el intervalo creíble indica dónde es probable que se encuentre el posible parámetro de la población^[230]. En el marco bayesiano, se puede evaluar directamente las probabilidades posteriores de que una declaración en particular sea verdadera. Las probabilidades posteriores a menudo se denominan valores p bayesianos. A diferencia de los valores p frecuentistas clásicos, cuanto mayores son, más apoyo tiene una declaración en particular^[399].

Estudios experimentales en las Ciencias del Movimiento Humano a menudo parten de datos que permiten tener una idea del resultado esperado. Esos resultados se pueden usar para confirmar o negar estas predicciones; ese es un enfoque bayesiano para el análisis de datos, que utiliza los resultados observados para modificar o respaldar el supuesto resultado^[230]. Así, la pregunta estadística que se desea responder es: "¿Cómo modifican los datos la predicción de los efectos de esta intervención?".

El enfoque bayesiano se ha utilizado también en revisiones sistemáticas de literatura con meta análisis, en donde se dice que es robusto aún con pocos estudios^[757]. A diferencia del meta análisis frecuentista, que solo permite el rechazo de la hipótesis nula (H_0) o de la hipótesis alternativa (H_1), el enfoque bayesiano permite encontrar evidencia tanto para H_0 como para H_1 . Los resultados bayesianos, se pueden interpretar de forma más intuitiva que los resultados frecuentistas, especialmente si los resultados respaldan H_0 . Además, con los modelos bayesianos se pueden crear intervalos creíbles, lo que permite hacer inferencias acerca de los estudios no incluidos en el análisis.

Por ejemplo, se estudió el efecto de la suplementación con un arbusto (*Withania somnifera*) sobre el rendimiento físico y otras variables en individuos sanos^[93], en el cual parte del análisis incluyó un modelo jerárquico bayesiano. Los investigadores encontraron con el meta análisis tradicional que la suplementación con el arbusto fue más eficaz que el placebo para mejorar las variables relacionadas con el rendimiento físico en personas sanas. Ese hallazgo fue confirmado con el modelo bayesiano, que mostró que las intervenciones futuras podrían ser, al menos, beneficiosas para los resultados analizados considerando los intervalos creíbles del 95% del tamaño del efecto meta analítico.

En el área de la actividad física y la salud en respuesta a la pandemia por COVID-19 también se ha utilizado el enfoque bayesiano para analizar datos. Así, se realizó un estudio para determinar la cantidad de pasos y poder calcular la cantidad de actividad física de intensidad moderada y vigorosa en el contexto de la pandemia^[399]. Para ello, se les pidió a los participantes que proporcionaran datos portátiles longitudinales de dispositivos electrónicos desde diciembre de 2019 hasta junio de 2020. Se hicieron análisis del punto de cambio bayesiano de los datos de 97 personas y se encontró que hubo una fuerte disminución de 1473 pasos, con un intervalo creíble del 95% [IC95% = -2218 a -709 pasos), así como una reducción de minutos de actividad ligera de 41.9 min (IC95% = -54.3 a -29.3 min), y un aumento en los minutos de actividad física de moderada a vigorosa intensidad de 11.7 min (IC95% = 2.9 a 19.4 min). Los investigadores concluyeron que hubo modificaciones conductuales claras antes y durante el período inicial de confinamiento por COVID-19^[399].

En el área del rendimiento deportivo de élite se ha utilizado el enfoque bayesiano para predecir el tiempo ganador en natación de los Juegos Olímpicos^[762]. Se desarrolló un modelo estadístico de tiempos ganadores para eventos internacionales de natación para predecir distribuciones de tiempos ganadores y la probabilidad de ganar en los Juegos Olímpicos de 2020 y 2024. Los investigadores construyeron un conjunto de datos que incluyó el tiempo del 1^{er} y 3^{er} lugar de todos los eventos de natación individuales de los Juegos Olímpicos y Campeonatos Mundiales de 1990 a 2019^[762]. Para lograr el objetivo, los investigadores compararon diferentes modelos ajustados con inferencia bayesiana para obtener distribuciones predictivas. Dichas comparaciones se basaron en el error porcentual promedio en las predicciones fuera de la muestra de los tiempos de natación de los ganadores de las competencias. En general, se encontró un error de predicción promedio de 0.70% (IC95%= 0.59, 0.82%) y que los Juegos Olímpicos fueron en promedio un 0.5% (IC95%= 0.3, 0.7%) más rápidos que los Campeonatos del Mundo. Así, los modelos bayesianos en el alto rendimiento logran calcular la distribución predictiva posterior, que a la vez, permite a los entrenadores y atletas evaluar la probabilidad de ganar dado el tiempo de natación de un individuo y la probabilidad de ser más rápido o más lento que el tiempo ganador anterior o incluso del récord mundial^[762].

El enfoque bayesiano también ha sido comparado con el frecuentista en las Ciencias del Movimiento Humano para determinar la validez convergente de dos métodos utilizados para medir calidad del sueño^[143]. En el estudio, se compararon dos dispositivos y se llegó a la conclusión de que carecen de validez convergente; es decir, los instrumentos no son idénticos para medir la calidad del sueño; ambos son válidos, pero no se pueden intercambiar y esperar los mismos resultados.

Se ha indicado que el enfoque bayesiano es especialmente importante cuando se utilizan pruebas como análisis de correlación, t-student y cuando se tienen hallazgos significativos, ya que se le considera una alternativa metodológica para la replicación estadística que cuantifica el grado de evidencia del contraste de la H_1 con la H_0 . Adicionalmente, la interpretación es simple: débil, moderado, fuerte, muy fuerte y extremo^[376] (Tabla 1).

El enfoque de inferencia bayesiano se ha recomendado recientemente como complemento a estudios frecuentistas en los cuales se utilizan muestras de participantes reducidas, como ocurre en algunas áreas de las Ciencias del Movimiento Humano^[593]. Sin embargo, el enfoque bayesiano también se ha utilizado con muestras grandes. Por ejemplo, en Alemania, se estudiaron más de 35 mil niños para observar su competencia motriz^[407]. Los datos recolectados en el año 2000 y 2011 se ajustaron por medio de un modelo de regresión bayesiano jerárquico con diferentes intersecciones y tendencias temporales. Los investigadores encontraron una reducción anual en las pruebas de competencia motriz, en donde los niños con sobrepeso y obesidad mostraron una desventaja en todos los puntajes de las pruebas.

Tabla 1. Interpretación de los valores del factor de Bayes^[376].

Fuerza del valor	Valor	Dirección de la evidencia
Extrema	> 100	H ₁
Muy fuerte	30 a 100	H ₁
Fuerte	10 a 30	H ₁
Moderada	3.1 a 10	H ₁
Débil	1.1 a 3	H ₁
0	1	No hay evidencia
Débil	0.3 a 0.99	H ₀
Moderada	0.29 a 0.1	H ₀
Fuerte	0.09 a 0.03	H ₀
Muy fuerte	0.03 a 0.01	H ₀
Extrema	< 0.01	H ₀

Se espera que en los próximos años el enfoque bayesiano para el análisis de datos en las Ciencias del Movimiento Humano sea más relevante. La complejidad de los datos, así como los grandes volúmenes de información, permitirán ser estudiados con mayor detenimiento y compararse con el modelo frecuentista y bayesiano, lo cual beneficia a los investigadores pues podrán contar con mayor información para la toma de decisiones. Adicionalmente, se espera una coordinación entre profesionales de la Estadística y los de las Ciencias del Movimiento Humano para el planteamiento y definición de parámetros e hiperparámetros en el enfoque bayesiano; esto con el propósito de darle mayor contexto e interpretación a los datos.

2.1.1. Alternativas y complementos de $p < 0.05$

Cuando leemos en un artículo de investigación que “se encontraron diferencias estadísticamente significativas ($p < 0.05$) ...”, o que “la correlación entre las variables fue significativa ($p < 0.05$) ...”, inferimos que estamos ante un problema clásico de prueba de significancia de hipótesis. Las inferencias son conclusiones basadas en la evidencia acerca de la “verdadera naturaleza” de algo, y el enfoque tradicional en la investigación con muestras de participantes es una afirmación acerca de si el efecto es estadísticamente significativo o “real” basado en un valor de p ^[348].

Así, lo que nos enseñan en los cursos básicos de estadística es a estudiar H_0 y H_1 , y las consecuencias de rechazar o de mantener una u otra. En otras palabras, “¿Con qué fuerza favorece la evidencia la hipótesis alternativa [H_1] en relación con la hipótesis nula [H_0]?” (p. 186)^[69]. En los enunciados anteriores, cuando rechazamos H_0 estamos concluyendo con un 95% de certeza, que las diferencias o las asociaciones existen más allá del azar, y que aceptamos que pudiéramos equivocarnos al rechazar H_0 un 5% de las veces. Es decir, 5% de las veces, cometeríamos el error de rechazar la H_0 (i.e., Error tipo I, también llamado “falso positivo”)^[231].

Existen varias definiciones acerca del valor p . Algunos lo definen como “la probabilidad, bajo el supuesto de no efecto o no diferencia (hipótesis nula), de obtener un resultado igual o más extremo que el realmente observado”^[332]. Por lo tanto, si p es menor que algún nivel de significancia predefinido (normalmente 0.05), se dice que el resultado es estadísticamente significativo. La probabilidad de obtener un resultado estadísticamente significativo se denomina potencia de la prueba, que también depende del tamaño del efecto real y del tamaño de la muestra. Cuando se tienen tamaños de muestra lo suficientemente grandes, generalmente se van a encontrar resultados estadísticamente significativos, lo cual podría significar que el efecto de alguna intervención sea prácticamente trivial o sesgado^{[348][776]}.

Otra definición aceptada de p es “...la probabilidad, bajo la hipótesis nula, de observar un estadístico de prueba como extremo o más extremo que su valor observado” (p. 187)^[69]. Una interpretación errónea es tratar de asociar un valor p bajo (e.g., $p = 0.001$) con un efecto relevante. Los valores p indican la certeza estadística con la que se puede afirmar que una hipótesis es verdadera y se prueban con un nivel α ; es decir, p indica qué tan probable es que un efecto observado no esté realmente allí y que se observe más bien por casualidad. Así, el nivel α es el límite de decisión, en donde tradicionalmente los valores $p < \alpha$ se consideran significativos, mientras que los valores $p > \alpha$ generalmente se interpretan como que no existen diferencias^[98].

Sin embargo, los valores de la significancia p por sí mismos a menudo se malinterpretan, de manera que conducen a exagerar la evidencia en contra de la H_0 . Se ha indicado que “Un valor p , o significancia estadística, no mide el tamaño de un efecto o la importancia de un resultado”, por lo que evaluar la magnitud del efecto, es decir, su “importancia práctica” también es importante^{[98][740]}. En otras palabras, es imposible determinar a partir de p cómo, por ejemplo, un nuevo tipo de entrenamiento se compara con la eficacia de otros tipos de entrenamientos previamente establecidos o incluso novedosos para producir un cambio en la misma variable de interés (e.g., fuerza, potencia, velocidad)^[253].

Las inferencias basadas en categorías dicotómicas como “existe” o “no existe” no son lo mismo que la lógica frecuentista (i.e., que se refiere a la interpretación de la probabilidad a largo plazo o cuáles serían los resultados probables de un estudio, si se repite una y otra vez)^[707], en la que se usa que “la probabilidad es pequeña”, tal y como se maneja en las pruebas de hipótesis tradicionales^[233]. Por supuesto, las pruebas de hipótesis asumen que se trabaja con muestras de personas (n) que representan a la población (N), lo cual es un supuesto que muy rara vez se cumple, y que *a priori*, representa un problema conceptual^[231].

Así, desde hace más de una década, se ha criticado fuertemente el uso exagerado de p para aseverar y reforzar los hallazgos en las investigaciones científicas, algo que se conoce como la “era post $p < 0.05$ ”^{[69][776]}. Por ejemplo, en la mayoría de las publicaciones se tiende a resaltar los valores de p más favorables y por lo tanto las inferencias son exageradas y poco confiables^[365]. Por lo tanto, se han realizado múltiples esfuerzos para instaurar nuevas prácticas que complementen la información que aporta p para reducir el sesgo en las publicaciones y evitar malas interpretaciones con respecto a p ^[296]. Incluso, se ha recomendado sustituir p por estadísticos como la razón de verosimilitud, el factor de Bayes, la probabilidad *a posteriori*, los intervalos de confianza y los coeficientes de regresión^{[87][98][134][348][707]}.

La Asociación Americana de Estadística ha recomendado que, si se usa el lenguaje actual de “importancia estadística” para un descubrimiento novedoso, se puede reemplazar el umbral p de 0.05 por 0.005^[365]; es decir, es mejor referirse a los hallazgos con un valor p entre 0.05 y 0.005 como “sugerentes”, en lugar de “significativos”. La segunda recomendación se refiere a que al reportar un valor p en una prueba de la H_0 vs. H_1 , también se reporte que las probabilidades basadas en datos de que H_1 sea verdadera y que H_0 también sea verdadera son como máximo $1/[-e p \log p]$, donde \log es el logaritmo natural y e es su base constante. Finalmente, se recomienda calcular y reportar las probabilidades anteriores de H_1 a H_0 (i.e., las probabilidades de que las hipótesis sean verdaderas antes de ver los datos), así como calcular y reportar las probabilidades finales (i.e., posteriores) de H_1 a H_0 , que son las probabilidades anteriores multiplicado por las probabilidades basadas en datos^[69].

Otra forma para no descartar totalmente p , es la recomendación de utilizar “valores p de segunda generación”^[87]. Los valores p de segunda generación resuelven fallas críticas que promueven la mala interpretación de los datos, la distracción por efectos triviales y la irreproducibilidad de los hallazgos. Se basan en el uso de una H_0 compuesta, que captura efectos nulos y científicamente triviales. Ya que muchos de los hallazgos falsos son pequeños efectos que técnicamente no son nulos pero prácticamente indistinguibles del nulo, los valores p de segunda generación reducen en gran medida la probabilidad de un falso descubrimiento, por lo que se ha mencionado que promueven la transparencia, el rigor y la reproducibilidad de los resultados científicos al identificar *a priori* qué hipótesis candidatas son significativas en la práctica y al proporcionar un resumen estadístico más confiable de cuándo los datos son compatibles con H_1 y H_0 , o cuándo los datos no son concluyentes.

Algunas sugerencias de expertos y de la Asociación Americana de Estadística para el uso correcto de los valores de significancia y su posible reporte en manuscritos se muestran detalladamente en la tabla 1^{[707][741]}.

Tabla 1. Sugerecias acerca del uso e interpretación de los valores de significancia y la manera de reportarlo en la investigación científica^{[707][741]}.

Sugerencias de la Asociación Americana de Estadística respecto a la significancia y su reporte
<ul style="list-style-type: none"> • No base sus conclusiones únicamente en si se encontró que una asociación o efecto es "estadísticamente significativo"; es decir, que el valor de p superó algún umbral arbitrario (e.g., $p < 0.05$).
<ul style="list-style-type: none"> • No crea que existe una asociación o un efecto solo porque fue "estadísticamente significativo".
<ul style="list-style-type: none"> • No crea que una asociación o efecto está ausente solo porque no fue "estadísticamente significativo".
<ul style="list-style-type: none"> • No crea que el valor p encontrado brinda la probabilidad de que la casualidad por sí sola produzca la asociación o efecto observado o la probabilidad de que su hipótesis de prueba sea verdadera.
<ul style="list-style-type: none"> • No saque conclusiones acerca de la importancia científica o práctica basándose en la significancia estadística (o la falta de ella).
<ul style="list-style-type: none"> • Trate de no usar el término "estadísticamente significativo". Tampoco debería usar variantes como "significativamente diferente", "$p < 0.05$" y "no significativo".
<ul style="list-style-type: none"> • Redacte las conclusiones brindando contexto con respecto a p. Por ejemplo, si en el estudio, el valor de $p = 0.025$, se podría redactar algo como lo siguiente: "Suponiendo que la H_0 y los supuestos del modelo fueras ciertos, la probabilidad de obtener un aumento en la fuerza muscular es del 2.5%. O por ejemplo, "Ya que el nivel alfa fue de 0.04, si decidiéramos pretender que hay un efecto del programa de entrenamiento cuando en realidad no lo hay (Error tipo I), no nos equivocáramos más del 4% de las veces".
<ul style="list-style-type: none"> • Acepte la incertidumbre, sea reflexivo, abierto y modesto. Estos principios permitirán diseñar mejores estudios, con muestras más grandes, más transparentes, y pensar en tamaños de efecto de importancia práctica y en mecanismos para explicar el fenómeno de estudio.

2.1.1.1. Riesgos relativos y razón de momios

Un riesgo se define como la frecuencia de ocurrencia de un evento en relación con todos los posibles eventos^[478]. Por otra parte, los momios o probabilidades comparan la frecuencia de que ocurra un evento con la frecuencia de que no ocurra el evento. Definiciones más elaboradas son las siguientes^[478]:

- a. Riesgo relativo: Es el riesgo de un resultado en aquellos individuos con exposición (N° de individuos con resultado y exposición, dividido entre el N° total de individuos con exposición), dividido entre el riesgo de resultado en aquellos individuos sin exposición (N° de individuos con resultado, pero no exposición, dividido entre el N° total de individuos sin la exposición).
- b. Razón de momios o probabilidades: Son las probabilidades de un resultado en aquellos individuos con exposición (N° de individuos con resultado y exposición, dividido entre el N° de individuos con exposición, pero sin resultado) dividido entre las probabilidades de resultado en aquellos individuos sin exposición (N° de individuos con resultado, pero sin exposición), dividido entre el N° de individuos sin el resultado o la exposición). También se puede usar para comparar las probabilidades de tener exposición en aquellos individuos con y sin un resultado particular.

Estos dos estadísticos se reportan en estudios retrospectivos, longitudinales y transversales (se reporta como un riesgo de prevalencia o de probabilidad) y provienen de modelos de análisis de regresión binomial o Poisson con errores estándares robustos en el caso de los riesgos relativos, y de la regresión logística en el caso de la razón de momios.

La forma de calcularlos y comprenderlos es simple (Tabla 1). El número de eventos de interés "n", es el N° de personas con exposición o factor de riesgo que tienen un resultado particular, el cual se divide entre el N° total de eventos "N". En la tabla se observa que $n/N = 2/30$; es decir, 0.067, y para comprenderlo más fácilmente, el valor obtenido se puede multiplicar por 100 para expresarlo en términos porcentuales; es decir, 6.7% veces el riesgo relativo de adquirir la condición a la que estuvieron expuestas las personas. Por su parte, la razón de momios es el N° de eventos de interés "n" que se divide entre el N° de eventos que no son de interés, $N-n$. Así, se puede calcular $n/(N-n) = 2/(30-2) = 0.036$ o 36%.

Tabla 1. Cálculo del riesgo relativo y de la razón de momios.

	Cantidad de personas	Riesgo relativo (RR)	Razón de momios (OR)
Cantidad de eventos que ocurrieron (n)	2	2/30 = 0.067	2/[(30-2)] = 0.036
Cantidad de personas expuestas (N)	30		

En las Ciencias del Movimiento Humano se reportan ambos indicadores. Por ejemplo, se diseñó un estudio para conocer el efecto de un programa de ejercicio comunitario de 3 semanas sobre las complicaciones postoperatorias luego de 30 días en pacientes de alto riesgo programados para resección colorrectal, los cuales fueron comparados con pacientes de un grupo de referencia^[72]. Se encontró que la tasa de complicaciones postoperatorias fue menor en el grupo

de prehabilitación que en el grupo de atención habitual, en donde el RR fue de 0.59. Además, reportaron que el intervalo de confianza al 95% (IC95%) estuvo entre 0.37 y 0.96. Esto quiere decir que el grupo de pacientes que realizaron ejercicio tuvieron un riesgo 59% menor que el grupo control, y que ese riesgo menor podía oscilar entre 37% y 96%. Es decir, más de la tercera parte de todos los pacientes con cáncer de colon y recto que se someterían a programas de ejercicio se podrían beneficiar.

En otro estudio, se investigó la relación entre la conmoción cerebral y el historial de accidentes cerebrovasculares en exjugadores de la Liga Nacional de Fútbol Americano ≥ 50 años^[110]. Para ello, los investigadores utilizaron una regresión logística y calcularon la razón de momios ajustadas, tomando en consideración el accidente cerebrovascular para los grupos de antecedentes de conmoción cerebral, la edad, la enfermedad de las arterias coronarias e infarto del miocardio. Se encontró que la prevalencia de accidente cerebrovascular fue mayor que la población general, con un OR= 0.56. Sin embargo, tuvieron mayor riesgo los jugadores que habían tenido 10 o más conmociones cerebrales previas (OR= 5.51, IC95% = 1.61, 28.95), aquellos con enfermedad cardiovascular (OR= 2.24, IC95% = 1.01 a 4.77), y quienes tenían mayor edad (OR=1.07, IC95% = 1.02 a 1.11).

Algunos ejemplos de estudios relacionados con las Ciencias del Movimiento Humano se muestran en la tabla 2. Por ejemplo, en el estudio de Bubnjević et al.^[117], el RR para que las embarazadas físicamente activas tuvieran un parto vaginal superó el valor de referencia 1, que es el valor comúnmente utilizado como el RR de comparación. Un valor >1 indicaría que hay mayor riesgo o probabilidad; mientras que un valor <1 , indicaría que hay menor riesgo o probabilidad, que en algunos contextos se le conoce como factor de protección.

Otro estudio muy interesante se refiere al posible beneficio de poseer una mascota (e.g., perros, gatos) sobre la cantidad de actividad física en la población adulta mayor^[684]. En ese estudio se determinó la asociación entre tener perros y gatos, el inicio de la discapacidad y la mortalidad por todas las causas en una población de adultos mayores. Luego de analizar datos de 11233 personas mayores de 65 años, se encontró que en comparación con un grupo de referencia de aquellos que nunca habían tenido un perro, los adultos mayores que actualmente tenían perros tenían un OR significativamente menor de aparición de discapacidad (OR = 0.54; IC95%= 0.37, 0.79). Desde el punto de vista de las Ciencias del Movimiento Humano, ese estudio es muy valioso porque muestra que tener un perro parece proteger contra la discapacidad incidente entre los adultos mayores.

Tabla 2. Estudios publicados en temas variados de las Ciencias del Movimiento Humano en donde se reportan riesgos relativos (RR) y razones de probabilidad o momios (OR).

Referencia	Propósito	Hallazgo
Bubnjević et al. ^[117]	Meta analizar el impacto del ejercicio aeróbico continuo sobre la frecuencia del parto vaginal en mujeres embarazadas sanas con valores de índice de masa corporal (IMC) normales.	<ul style="list-style-type: none"> • Las embarazadas físicamente activas que hacían ejercicio de intensidad ligera a moderada 3 veces por semana (35-60 minutos), tuvieron un parto vaginal más frecuente que las embarazadas sedentarias (67.4 % vs. 60.5). • El RR= 1.11 (IC95% = 1.04, 1.18).
Hsu et al. ^[353]	Investigar si los hábitos de ejercicio y las conductas de control de peso pueden reducir la incidencia de osteoporosis en la población general.	<ul style="list-style-type: none"> • En la población general el ejercicio regular o el control del peso reducen la incidencia de osteoporosis. • Los sujetos con un hábito de ejercicio regular (OR= 0.709; IC95%= 0.599, 0.839) y comportamiento de control de peso (OR= 0.753; IC95%= 0.636, 0.890) mostraron menores probabilidades de osteoporosis.
Ikeda et al. ^[360]	Examinar la magnitud de las asociaciones de cambios en la actividad física sobre el riesgo de dolor lumbar a los 4 años de seguimiento en adultos mayores.	<ul style="list-style-type: none"> • Mantener una actividad física de intensidad moderada (RR= 0.59; IC95%= 0.36, 0.99) o vigorosa (RR= 0.46; IC95%= 0.27, 0.77) al menos 1–3 veces al mes reduce la prevalencia del dolor lumbar en comparación con no realizar actividad física del todo.
		Continúa...

Tabla 2. Continuación.

Referencia	Propósito	Hallazgo
Jouven et al. ^[381]	Probar la hipótesis de que entre personas aparentemente sanas, la muerte súbita es más probable que ocurra en presencia de perfiles anormales de frecuencia cardíaca (FC) durante el ejercicio y la recuperación.	<ul style="list-style-type: none"> • El riesgo de muerte súbita por infarto del miocardio aumentó en personas con una FC en reposo > 75 latidos por minuto (RR= 3.92; IC95%= 1.91, 8.00). Es decir, aumentó más del triple. • El riesgo de muerte súbita por infarto de miocardio aumentó en personas con aumento de la FC durante el ejercicio < 89 latidos por minuto (RR= 6.18; IC95%= 2.37, 16.11). Es decir, aumentó más de 6 veces. • El riesgo de muerte súbita por infarto de miocardio aumentó en personas con una disminución de la FC < 25 latidos por minuto después de la finalización del ejercicio (RR= 2.20; IC95%= 1.02, 4.74). Es decir, aumentó más del doble. • El perfil de FC durante el ejercicio y la recuperación son predictores de muerte súbita.
Magro-Malosso et al. ^[456]	El objetivo del estudio fue evaluar el efecto del ejercicio durante el embarazo sobre el riesgo de trastornos hipertensivos gestacionales (THG).	<ul style="list-style-type: none"> • Las mujeres que realizaron ejercicio aeróbico: <ul style="list-style-type: none"> ○ Tuvieron una incidencia menor de THG (RR= 0.70; IC95%= 0.53, 0.83) en comparación con los controles. ○ Tuvieron menor incidencia de hipertensión gestacional (RR= 0.54; IC95%= 0.40, 0.74) en comparación con los controles.
		Continúa...

Tabla 2. Continuación.

Referencia	Propósito	Hallazgo
Poyatos-León et al. ^[579]	Realizar una revisión sistemática y un meta análisis para examinar la influencia de las intervenciones de ejercicio físico en el tipo de parto de mujeres embarazadas sanas con niveles bajos a moderados de actividad física.	<ul style="list-style-type: none"> • El ejercicio físico durante el embarazo puede aumentar la probabilidad de parto normal (RR= 1.12; IC95%= 1.01, 1.24). • El ejercicio físico durante el embarazo puede aumentar la probabilidad de parto normal en particular cuando el ejercicio se realiza durante el segundo y tercer trimestre (RR= 1.14; IC95%= 1.01, 1.32). • El ejercicio físico durante el embarazo puede aumentar la probabilidad de parto normal incluso reduciendo el riesgo de parto por cesárea (RR= 0.66; IC95%= 0.46, 0.96).
Sillanpää et al. ^[650]	Investigar las asociaciones entre las puntuaciones de riesgo poligénico (PRP) de la actividad física y las enfermedades cardiometabólicas en la población finlandesa.	<ul style="list-style-type: none"> • En personas con mayor cantidad de actividad física hubo un menor riesgo de enfermedad de enfermedades cardiovasculares (OR por cambio de 1 SD en actividad física PRP= 0.95), hipertensión (OR= 0.93), diabetes tipo 2 (OR= 0.91) y enfermedad coronaria (OR= 0.95). • Los participantes con actividad física PRP alta tuvieron un menor riesgo de mortalidad (OR= 0.97). • Las personas genéticamente menos activas tienen un mayor riesgo de desarrollar enfermedades cardiometabólicas.
		Continúa...

Tabla 2. Continuación.

Referencia	Propósito	Hallazgo
Taniguchi et al. ^[684]	Examinar la asociación entre la tenencia de mascotas, la aparición de discapacidad y la mortalidad por todas las causas en una población de mayor edad.	<ul style="list-style-type: none"> • Los adultos mayores que actualmente tenían perros tenían un OR significativamente menor de aparición de discapacidad (OR= 0.54; IC95%= 0.37, 0.79) en comparación con un grupo de referencia de aquellos que nunca habían tenido un perro.
van Kleef et al. ^[715]	Investigar la asociación entre la actividad física y la enfermedad del hígado graso no alcohólico (NAFLD, por sus siglas en inglés).	<ul style="list-style-type: none"> • La actividad física total se asoció con una prevalencia más baja de NAFLD ajustada por factores demográficos, de estilo de vida y socioeconómicos (OR= 0.958 por 10 min/día; IC95%= 0.929, 0.986). • La actividad física más intensa se asoció más fuertemente con una menor prevalencia de NAFLD por cada por 10 min/día: <ul style="list-style-type: none"> ○ Actividades físicas ligeras: OR= 0.931 (IC95%= 0.882, 0.982) ○ Actividades moderadas: OR= 0.891 (IC95%= 0.820, 0.967) ○ Actividades vigorosas: OR= 0.740 (IC95%= 0.600, 0.906)

2.1.1.2. Precisión de las estimaciones

Una apropiada interpretación de los resultados estadísticos que se presentan en las publicaciones de los estudios es esencial para aplicar correctamente los conocimientos desarrollados y tomar las decisiones que representen las respuestas halladas ante los problemas estudiados. Así, no solamente es importante comprender el estimador de la tendencia central de los datos (e.g., promedio o media estadística), sino también su precisión. Por ello, se utilizan indicadores como la desviación estándar, el error estándar, y los hoy en día recomendados intervalos de confianza (IC) e intervalos de predicción (IP)^[348].

2.1.1.2.1. Intervalos de confianza

Un IC es una medida estadística de gran utilidad, la cual comprende un rango de valores entre los cuales, con un cierto grado de incertidumbre (e.g., 90%, 95%), se considera que se encuentra el valor verdadero de una población. Los IC generalmente se presentan como los IC calculados al 90% (IC90%) o al 95% (IC95%) de precisión. Tradicionalmente, se ha considerado como una medida muy importante, pues es un indicador de la precisión de las observaciones realizadas; no obstante, una interpretación adecuada es imprescindible^[327].

En un ejemplo del reporte de los IC, Kelley et al.^[395] describen que la prevalencia de la inactividad física en el Estado de Colorado, EE.UU., es del 23.2%, señalando que con un 95% de confianza, el valor verdadero podría encontrarse entre 19.7% y 27.2%. Así mismo, indican que la mayor prevalencia de inactividad física se encontró en el Estado de Kentucky (44.4%) con un rango esperado del valor verdadero entre el 40.0% y 48.8%.

El cálculo de un IC es muy sencillo^[327]:

$$\text{IC} = \text{Promedio de la muestra} \pm \text{valor } z \times (\text{Desviación estándar} / \sqrt{n})$$

En un ejemplo práctico en las Ciencias del Movimiento Humano, se tiene que la altura del salto contra movimiento en una muestra aleatoria de 45 nadadores universitarios fue de 26.45 ± 3.30 cm. A partir de esa información, ¿cuál es el IC para la altura del salto contra movimiento observado? De acuerdo con la fórmula anterior:

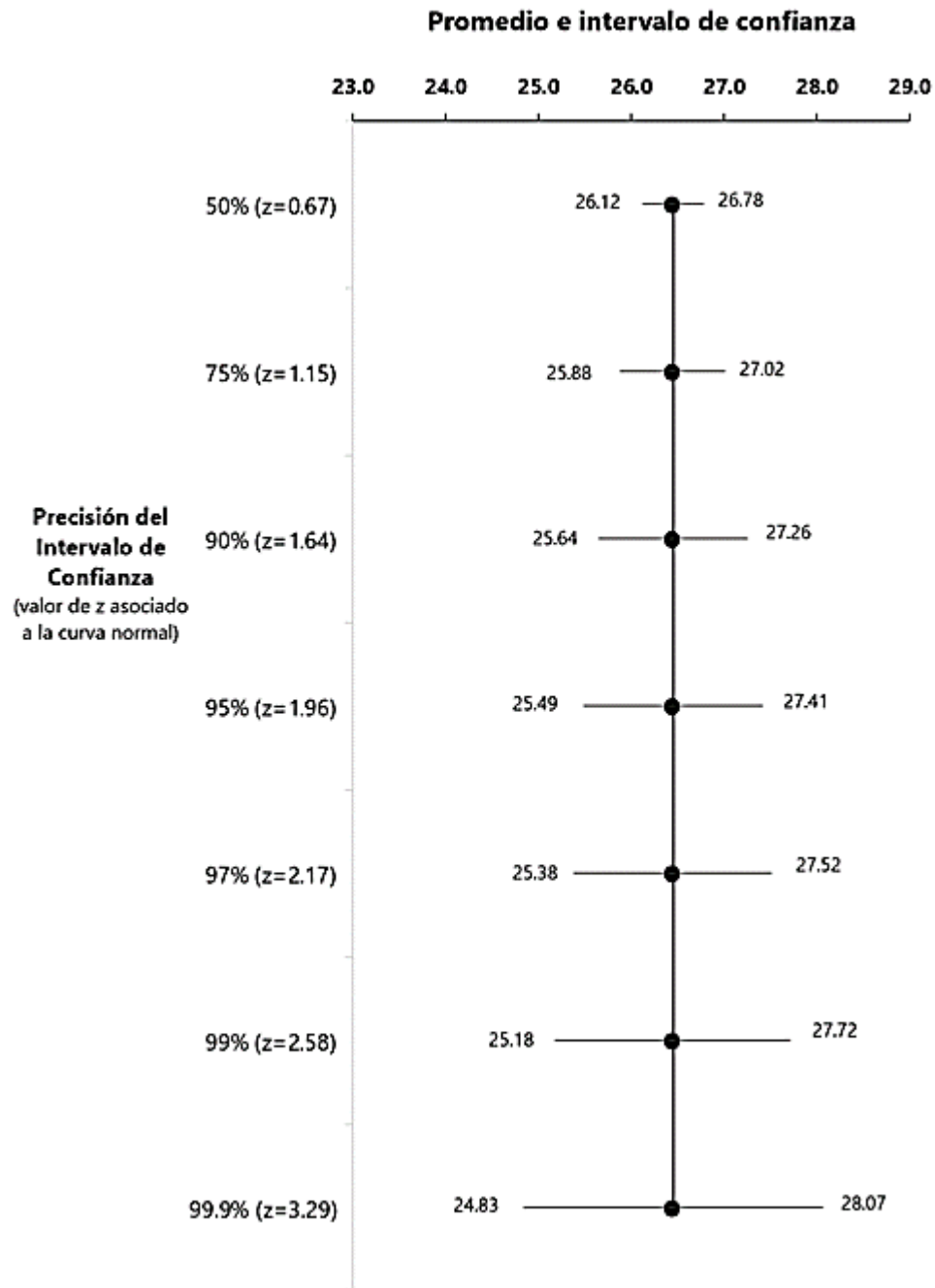
$$\text{IC95\%} = 26.45 \pm 1.96 \times (3.30 / \sqrt{45})$$

Límite inferior IC95% = 25.49 cm y Límite superior IC95% = 27.41 cm

Por lo tanto, el valor verdadero de altura del salto estaría 95% de las veces entre 25.49 cm y 27.41 cm

Para el ejemplo anterior, si se calcula el IC del 95% de la media, el valor z a utilizar sería 1.96. Sin embargo, también se pueden calcular los intervalos de confianza dependiendo del interés en la precisión. Por ejemplo, se puede usar el valor z asociado al 50%, 75%, 90%, 95%, 97%, 99% o incluso al 99.9% (Figura 1).

Figura 1. Ejemplo de intervalos de confianza calculados a diferentes niveles de precisión con base en el valor z de una distribución normal. Los valores mostrados en el gráfico son cm.



El margen de error depende del tamaño y la variabilidad de la muestra. Naturalmente, el error será menor si el tamaño de la muestra (i.e., n) es grande o la variabilidad de los datos (e.g., Desviación estándar [DE]) es pequeña. Idealmente, la DE utilizada en el cálculo debería ser la DE de la población. Sin embargo, considerando que esto es a menudo desconocido, si se cuenta con un número razonablemente grande (e.g., $n > 100$, o al menos > 30) y la muestra es aleatoria, entonces la DE de la muestra se puede usar como una buena aproximación de la DE de la población^[327].

En una visión tradicional de la estadística frecuentista se tiende a buscar e interpretar el valor de significancia p , con el objetivo de clasificar los resultados y concluir que "no hay diferencias" o "no hay asociaciones". Sin embargo, actualmente se recomienda reportar e interpretar los IC con este mismo propósito, analizando si los intervalos "tocan" o se superponen sobre cero para concluir si existen diferencias o no. Los valores p y los IC son útiles, más no se deberían utilizar con el fin de tomar una decisión dicotómica de si se apoya o rechaza una hipótesis^{[21][327][348]}.

El anterior enfoque puede llevar o cometer el error de concluir que dos estudios se contradicen solo por el simple hecho de que uno encuentre diferencias estadísticas y otro no^[21]. Un dilema fundamental al respecto es el hecho de que con una muestra lo suficientemente grande cualquier efecto puede llegar a ser estadísticamente significativo^[348].

Aunque el valor $p < 0.05$ comúnmente se ha aceptado como el umbral para rechazar la H_0 , y por lo tanto se acepta que "existen diferencias", este valor de umbral es arbitrario, ha sido utilizado históricamente y puede conducir a tomar decisiones desacertadas. Si al realizar una prueba de ANOVA para determinar la eficacia de una intervención se obtiene un valor de $p = 0.072$, bajo el concepto histórico y tradicional de $\alpha = 0.05$, se debería concluir que el tratamiento no fue efectivo y se podría finalizar el análisis con este resultado. No obstante, en contextos como el ámbito del rendimiento deportivo, donde los deportistas están cerca de su techo genético y el éxito deportivo se basa en mejoras mínimas; puede ser preferible arriesgarse al aumento de falsos positivos (i.e., rechazar H_0 y cometer un error tipo I) que protegerse firmemente contra los falsos negativos (i.e., mantener H_0 y cometer un error tipo II). Esta línea de pensamiento tiene mayor sentido cuando se considera que las pruebas estadísticas en rendimiento deportivo a menudo tienen poca potencia^[706].

Otro elemento que debe ser contemplado y que refuerza la importancia de interpretar los resultados evitando el enfoque tradicional dicotómico (i.e., significativo vs. no significativo), es la necesidad de evaluar la magnitud de los cambios registrados en función de su importancia en el mundo real y en un contexto específico. Se debe contemplar que la falta de interpretación puede llevar a concluir que un efecto que probablemente sea sustancial sea catalogado como nulo o trivial y un efecto significativo que probablemente sea trivial se malinterprete como sustancial^{[225][348]}. Por ello, recientemente se recomienda a los investigadores aprender a aceptar la incertidumbre. Una forma práctica de hacerlo es renombrar los IC como "intervalos de compatibilidad" e interpretarlos de una manera que evite el exceso de confianza. Es apropiado que los autores describan las implicaciones prácticas de todos los valores dentro del intervalo, especialmente el efecto observado (i.e., estimación puntual) y los límites.

Al hacerlo, deben recordar que todos los valores entre los límites del intervalo son "razonablemente compatibles" con los datos, dadas las suposiciones estadísticas utilizadas para calcular el intervalo. Por lo tanto, destacar un valor en particular (como el valor nulo) en el intervalo como "mostrado" no tiene sentido^{[21][706]}.

Se debe tener claro que, en la investigación científica, la palabra "significativo" no es un sinónimo de la palabra "importante", razón por la que paralelamente y como complemento a una adecuada interpretación de los resultados estadísticos, sobresale la relevancia de analizar la estimación del tamaño del efecto, y de esta manera evaluar la magnitud de los cambios o las asociaciones estudiadas en función de la naturaleza de la variable y el significado práctico y contextualizado de los datos^{[225][327][706]}.

2.1.1.2.2. Intervalos de predicción

Los IP corresponden a un cálculo que se ha integrado recientemente en los meta análisis con el objetivo de evaluar las implicaciones clínicas de un resultado según la heterogeneidad. Este es un concepto sencillo de comprender; consiste en el rango esperado en que se puede presentar un efecto verdadero entre estudios similares.

Mientras que los IC95% en un meta análisis frecuentista se interpretan como la confianza al 95% de que la media de todos los estudios se encuentra entre dos valores (i.e., límite inferior y límite superior), los IP al 95% se interpretan como lo que se podría esperar encontrar si se realizara el mismo estudio en la misma población.

Así, el IP refleja la variación en los efectos del tratamiento en diferentes entornos, incluyendo qué efecto se esperaría en futuras intervenciones^[362]. De acuerdo con Kelley y Kelley^[390], los intervalos de predicción se calculan de la siguiente manera:

$$\text{Intervalos de predicción} = \hat{\mu} \pm t^{\alpha}_{k-2} \sqrt{\text{Error estándar}(\hat{\mu})^2 + \tau^2}$$

Donde $\hat{\mu}$ es la estimación ponderada de los tamaños de efecto entre estudios, t^{α}_{k-2} es el 100 $(1-\alpha/2)$ % percentil de la distribución t con $k-2$ grados de libertad, $\text{Error estándar}(\hat{\mu})^2$ es el error estándar cuadrático estimado de $\hat{\mu}$, y τ^2 es el estimado entre la varianza del estudio.

Es importante comprender que las conclusiones basadas en los IP no deben usarse para determinar si los IC correspondientes a la estimación puntual son correctos o incorrectos, ya que los IP se basan en un efecto medio aleatorio mientras que los IC no.

En un ejemplo de la aplicación de los IP, Kelley y Kelley^[390] describen que previamente habían logrado concluir mediante la técnica del meta análisis que el entrenamiento progresivo contra resistencia, entre otros resultados, lograba reducir de manera significativa el colesterol total; no obstante, aunque dichos resultados alentaban a las personas a realizar ejercicio contra resistencia, se basaban en el análisis del efecto promedio del tratamiento y sus correspondientes IC95%, por lo que esta aproximación no generaba información respecto a cómo se distribuyen los efectos del entrenamiento de nuevos estudios individuales en relación al tamaño de efecto global del meta

análisis. De esta manera, lograron concluir que el entrenamiento progresivo contra resistencia lograba reducir el colesterol total en -5.5 mg/dl (IC95% = -9.4 a -1.6 mg/dl) y mediante los IP con una certeza del 95%, el efecto esperado en un nuevo estudio se encontraría entre -24.0 y 13.0 mg/dl.

En las investigaciones realizadas en las diferentes áreas de las Ciencias del Movimiento Humano, una correcta interpretación de los resultados de un estudio debería contemplar la integración de la valoración de p , el tamaño de efecto y el IC o de compatibilidad, así como los IP. De esta manera se contará con herramientas suficientes para obtener conclusiones apropiadas de los datos obtenidos. El reporte de indicadores de certeza y de incertidumbre, permitirá brindar información valiosa para la toma de decisiones conceptuales y de naturaleza práctica (e.g., entrenadores, atletas).

2.2. Interpretación de tamaños de efecto

Al estudiar si una intervención, como, por ejemplo, un programa de entrenamiento físico, produjo en las personas algún cambio en una variable de interés (e.g., capacidad aeróbica, depresión, atención, potencia muscular), lo que realmente nos estamos preguntando es ¿cuál es la magnitud de ese cambio?; es decir, el tamaño o la fuerza del efecto encontrado^{[502][503]}. La magnitud, de la cual existen múltiples estadísticos^{[414][425][508]}, se puede expresar por medio de un estimador cuantitativo y una interpretación cualitativa. El estimador cuantitativo puede ser el valor de la variable de interés en su forma cruda u original (e.g., $\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$, mmHg, cm, Watts, Newtons), un delta o porcentaje de cambio (i.e., $\Delta\%$), una diferencia promedio estandarizada (i.e., d de Cohen), coeficientes de regresión o un IC, entre otros^[98].

Por otra parte, la magnitud se puede expresar de una forma cualitativa para brindarle un mayor contexto al hallazgo^[62]. Por ejemplo, se puede decir que la magnitud del cambio fue trivial, pequeña, moderada o gigantesca. Batterham y Hopkins^[62] sugieren combinar el uso de IC y asignarle a la magnitud una categorización como de “posiblemente” trivial, beneficiosa o dañina, dependiendo del IC y de si ese IC se superpone sobre cero (i.e., indicaría que no hay un cambio estadísticamente significativo). Lo importante de expresarlo de forma cualitativa es que los lectores menos especializados en la temática podrán comprender la significancia del cambio en términos simples y comprensibles (Figura 1). Además, se ha indicado que las interpretaciones de los tamaños de efecto no pueden ser rígidas^[414].

Figura 1. El proceso posterior a la intervención requiere de una interpretación de la magnitud o fuerza del cambio.



La correcta interpretación de los tamaños de efecto ha cobrado especial relevancia hoy en día debido a la escasa información que proporcionan los valores de significancia p del enfoque frecuentista utilizado en las tradicionales pruebas de hipótesis (i.e., $p < 0.05$)^[98]. De hecho, Sullivan y Feinn^[677] citan textualmente a Gene Glass y a Jacob Cohen, quienes indican, respectivamente, que "La significancia estadística es lo menos interesante de los resultados. Debe describir los resultados en términos de medidas de magnitud, no solo si un tratamiento afecta a las personas, sino cuánto les afecta", y "El producto principal de una investigación es una o más medidas del tamaño del efecto, no valores P " (p. 279). Aunque no existe alguna única guía aceptada mundialmente para interpretar los tamaños del efecto, se han publicado numerosas opiniones acerca de cómo se podrían interpretar dependiendo del tipo de tamaño de efecto calculado y del contexto en que se utilicen^{[178][253][348][677]} (Tabla 1).

Tabla 1. Posibles interpretaciones para los tamaños de efecto provenientes de distintos indicadores. Los valores son sugerencias y deberían interpretarse en contextos específicos en las diferentes áreas de las Ciencias del Movimiento Humano.

Indicador	Interpretación	
	Cuantitativa	Cualitativa
d de Cohen	0.2	Pequeño
	0.5	Mediano
	0.8	Grande
	1.3	Muy grande
d de Cohen modificada por Hopkins para marcas de distancia y tiempo en el deporte	0.3	Pequeño
	0.9	Mediano
	1.6	Grande
	2.5	Muy grande
	4.0	Extremadamente grande
Δ de Glass	0.2	Pequeño
	0.5	Mediano
	0.8	Grande
Razón de momios (OR)	1.5	Pequeño
	2	Mediano
	3	Grande
Riesgo relativo (RR)	2	Pequeño
	3	Mediano
	4	Grande
r de Pearson	± 0.2	Pequeño
	± 0.5	Mediano
	± 0.8	Grande
r de Spearman	± 0.29	Pequeño
	± 0.50	Mediano
	± 0.51	Grande
Coeficiente de determinación r^2	0.04	Pequeño
	0.25	Mediano
	0.64	Grande

Existe un indicador menos utilizado en las Ciencias del Movimiento Humano, pero que sin lugar a dudas cobrará más importancia muy pronto. Este indicador, llamado Probabilidad de Beneficio (PDB) o Probabilidad de Superioridad^[414], se define como la probabilidad de que una puntuación seleccionada al azar de una población sea mayor que una puntuación seleccionada al azar de otra^[476].

En el contexto de un estudio experimental en las Ciencias del Movimiento Humano, se refiere a la probabilidad de que una persona que recibió el tratamiento (e.g., programa de ejercicios) seleccionada al azar muestre un nivel de mejora superior al de una persona del grupo control o de placebo seleccionada al azar^[253]. La PDB se calcula como un puntaje estandarizado, por medio de la ecuación: $z = \text{diferencia promedio estandarizada} / \sqrt{2}$. El valor obtenido se interpreta como que no existe posibilidad de beneficio si $z = 0.5$, que el placebo o control es mejor que el tratamiento si $z < 0.49$, o que el tratamiento es mejor que el placebo o control si $z = 0.51$ a 1.0 .

En síntesis, a medida que los investigadores conozcan más la variable dependiente de interés, la forma en que se mide, la manera en que varía y sus posibles fuentes de error, será más probable que el hallazgo derivado de una intervención sea comprendido dentro de un contexto más específico, y que la información aportada colabore en la toma de decisiones. Esta argumentación está sobre las decisiones tomadas única y exclusivamente a partir de valores de significancia estadística (i.e., p), que han mostrado ser de poca utilidad, y que actualmente se complementan con otros indicadores.

2.3. Significancia estadística vs. clínica vs. práctica

En la investigación científica, ya sea en los ámbitos de la salud o del movimiento humano, a menudo se encuentran hallazgos estadísticamente significativos, pero con escaso o nulo valor clínico o práctico^[86]. De hecho, existe una fuerte crítica hacia las revistas científicas porque tienen el sesgo de publicar únicamente hallazgos estadísticamente significativos (i.e., $p < 0.05$)^[365]; renunciando a publicar hallazgos que aunque no sean estadísticamente significativos, pueden tener el potencial de aportar información valiosa a un campo de estudio. En el área de la terapia física deportiva se ha reportado que 18% de los hallazgos estadísticamente significativos tenían un 50% de probabilidad de que el descubrimiento fuese falso; es decir, afirmar que el efecto de un tratamiento es real cuando en realidad no lo es^[86].

El aspecto tal vez más complejo sea determinar ¿qué es y qué no es clínica o prácticamente significativo o útil? Por ejemplo, en un estudio en donde se utilizó el entrenamiento contra resistencia se definió que un cambio clínicamente significativo en el colesterol HDL de las personas sería una reducción de 0.38 mmol/L^[6]. Seleccionar ese umbral es la tarea más compleja, ya que se debe asociar con la valoración del riesgo y beneficio para un paciente, o en el ámbito deportivo, el beneficio para un atleta. Así, la importancia clínica o práctica verdaderamente valiosa solamente puede determinarse tomando en consideración la diferencia promedio y su medida de variabilidad, generalmente el IC, y los umbrales relevantes de un cambio mínimo detectable (CMD) y la diferencia (clínica o práctica) mínima importante (DMI)^[6].

En la tabla 1 se presentan definiciones relevantes para determinar la significancia clínica y práctica relacionada con un ámbito en particular (e.g., clínico, deportivo); ambas definiciones tienen implícito un valor "umbral" a partir del cual se toman decisiones. Uno de los valores se refiere a la sensibilidad del cambio en la variable dependiente que no está influenciada por el error de la medición y el otro valor umbral se refiere a qué tan impactante puede ser para el beneficio de la persona.

Tabla 1. Definiciones relacionadas con el cambio en una variable dependiente que permiten establecer una relevancia más allá de la significancia estadística^{[86][182]}.

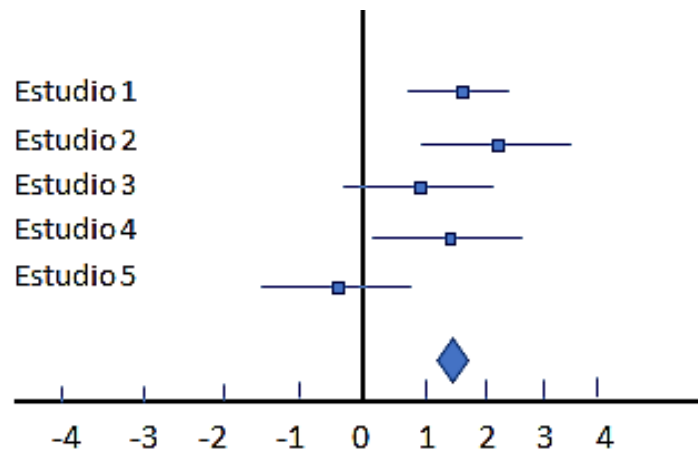
Indicador	Definición
Cambio mínimo detectable.	La cantidad de cambio en la variable dependiente que debe observarse antes de que se considere superior a los límites del error de medición.
Diferencia mínima (clínica o práctica) importante (i.e., diferencia mínima que valga la pena).	El cambio más pequeño en la variable dependiente que sería importante o les produciría un beneficio a los participantes (e.g., pacientes, clientes, atletas).

En el ámbito clínico, los cambios pueden ser definidos por el paciente o por el personal de salud; es decir, los pacientes pueden informar el cambio subjetivo y el personal de salud puede determinar un valor clínico a partir del cual observan mejoras en los pacientes^[182]. Por ejemplo, los pacientes en un ámbito clínico podrían utilizar una escala de tipo Likert para indicar su CMD con respecto a un tratamiento recibido (e.g., 1: completamente recuperado, 2: mejorado mucho, 3: ligeramente mejorado, 4: sin cambios, 5: levemente peor, 6: mucho peor).

Existen diversas formas para calcular el CMD; por ejemplo, se puede calcular para cada persona o para un grupo completo. En el caso del CMD personal, se calcula de la siguiente forma: $CMD_{individual} = 1.96 \times \sqrt{2} \times EEM$, en donde EEM representa el error estándar de la medición, el cual a su vez se calcula como la raíz cuadrada de la suma del cuadrado medio del observador y el cuadrado medio del error. El CMD grupal, se calcula como: $CMD_{grupal} = CMD_{individual} / \sqrt{n}$ de los atletas^[269]. Es importante destacar que a medida que el EEM se reduce, a la vez se tiene un valor de cambio mucho más verdadero y útil. Por lo tanto, se deben realizar esfuerzos para utilizar los mejores instrumentos de medición de las variables dependientes (i.e., válidos, confiables, sensibles), seguir los protocolos de medición a conciencia, utilizar tamaños de muestra lo más grandes posibles y análisis de datos robustos.

Otra manera para estimar el CMD se basa en decisiones subjetivas pero sustentadas en la evidencia estadística^[394]. Aquí se podría decir que convergen la experiencia y el conocimiento de los investigadores. Por ejemplo, con base en hallazgos meta analíticos, se pudo determinar que si un $CMD > 1.0$ entonces muchos pacientes pueden obtener importantes beneficios del tratamiento. Si la estimación del efecto se encuentra entre 0.5 y 1.0, el tratamiento puede beneficiar a un número apreciable de pacientes. A medida que la estimación combinada < 0.5 CMD, se vuelve progresivamente menos probable que un número respetable de pacientes obtenga beneficios importantes de un tratamiento o intervención^[380] (Figura 1).

Figura 1. Información proveniente de meta análisis para definir un cambio mínimo detectable. El valor de 1 representa el cambio mínimo detectable que podría detectarse más allá del error de la medición.



En las Ciencias del Movimiento Humano existen diversos métodos de entrenamiento destinados a mejorar la capacidad aeróbica, medida generalmente por medio del indicador $VO_2máx$ en unidades de $ml \cdot kg^{-1} \cdot min^{-1}$ o de equivalentes metabólicos de la tarea (i.e., MET por sus siglas en inglés). Cuando se integra la información epidemiológica con el conocimiento de las respuestas y adaptaciones fisiológicas del ser humano, se pueden sugerir DMI basados en evidencia. Por ejemplo, la evidencia epidemiológica indica que un valor umbral del $VO_2máx$ de $1.75 ml \cdot kg^{-1} \cdot min^{-1}$ (~ 0.5 MET) se asocia a una disminución de $\sim 8\%$ en el riesgo relativo de enfermedad cardiovascular a 11 años, y que un valor umbral de $3.5 ml \cdot kg^{-1} \cdot min^{-1}$ (~ 1 MET) a una disminución de $\sim 16\%$ en dicho riesgo relativo^[421]. Así, algunos investigadores han utilizado el

umbral de $1.75 \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ como el DMI en sus estudios^[491]. El mensaje más relevante puede ser la necesidad de buscar literatura que apoye la selección de un umbral DMI determinado. En los estudios de entrenamiento físico tratar de definir un DMI es complejo porque depende del contexto del entrenamiento, el volumen, el tipo de adaptaciones que se pretende monitorear y la variable monitoreada en sí^[118]. Para los lectores que deseen seguir un ejemplo de cálculos para determinar una DMI se les recomienda consultar el artículo de Turner et al.^[706].

Se han reportado varios DMI en estudios relacionados con adaptaciones al entrenamiento y su efecto crónico sobre varios índices de la frecuencia cardíaca^[118]. Por ejemplo, se describieron DMI para la frecuencia cardíaca de reposo ($\sim 2\%$), índices de variabilidad de la frecuencia cardíaca relacionados con el estado vagal en reposo ($\sim 3\%$), frecuencia cardíaca durante el ejercicio ($\sim 1\%$), frecuencia cardíaca de reserva posteriores al ejercicio ($\sim 7\%$), e índices de variabilidad de la frecuencia cardíaca relacionados con el tono vagal posteriores al ejercicio ($\sim 4\%$). Por supuesto que estos valores se basan en los datos recolectados, por lo que nuevamente se enfatiza en la calidad de la medición de las variables.

En otros estudios acerca de la actividad física para mejorar la salud, se ha encontrado que un DMI de $500 \text{ pasos}\cdot\text{día}^{-1}$ puede reducir el riesgo de enfermedad cardiovascular (2%), mortalidad (8-9%) y morbilidad (2%) en adultos inactivos^[620]. Incluso, si se usan acelerómetros triaxiales para medir el DMI, se ha encontrado evidencia que indica que un cambio en la aceleración media de 1.0 mg (mg: unidades miligravitacionales) se asocia a una reducción del 5% en la mortalidad por cualquier causa, y que 6 min de caminata vigorosa aumenta la expectativa de vida en ~ 4 años. Se recomienda ajustar los DMI a medida que surja mayor evidencia.

Ya que la determinación del CMD o DMI no representa una tarea fácil, se recomienda utilizar las revisiones sistemáticas de literatura y los meta análisis para obtener valores basados en la evidencia que permitan tomar decisiones informadas^{[641][738]}. Por ejemplo, en pacientes con cardiomiopatía dilatada que se sometieron a programas de rehabilitación cardíaca, se estimó que la mejoría promedio fue de 19.5% en la capacidad para realizar esfuerzo físico^[641]. Ese valor de 19.5% puede cambiarse a unidades originales (e.g., distancia recorrida, $\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$, METs) para así obtener un CMD o un DMI dependiendo del contexto del estudio. Sin embargo, los valores obtenidos en las revisiones de literatura deben someterse a verificación, tal y como lo hicieron Warkentin et al.^[738], al estudiar las recomendaciones de pérdida de peso descritas en la literatura. Los investigadores encontraron una discordancia entre las recomendaciones de pérdida de peso para mejorar la calidad de vida establecidas (i.e., DMI entre 5% y 10%), con las calculadas con datos propios, en donde encontraron que se requiere de una pérdida de peso de al menos el 20% en el transcurso de dos años para encontrar beneficios clínicamente importantes en la calidad de vida de los pacientes.

En síntesis, todavía queda por determinar cientos de CMD y DMI para guiar los procesos de investigación en las Ciencias del Movimiento Humano. La integración de la evidencia publicada y la experiencia de los entrenadores también aportará información valiosa para definir dichos umbrales. Se espera que en los próximos años se publiquen recomendaciones de umbrales aplicados a los ámbitos de la salud y del rendimiento deportivo.

2.4. Validez y confiabilidad

La validez y la confiabilidad son dos propiedades psicométricas que permiten conocer la calidad de los instrumentos, aparatos, “tests” o pruebas que consecuentemente proporcionan los datos que se recolectan en las investigaciones o en el campo del entrenamiento físico. Están asociadas a los conceptos de medición y evaluación, que significan, en términos generales, el proceso de obtener una característica (i.e., un puntaje) de un sujeto, objeto o situación, y emitir un criterio técnico acerca de la característica observada o medida, respectivamente^{[492][515]}.

Un puntaje está compuesto por su valor verdadero y por su error; y a medida que aumente el error de la medición (e.g., instrumentos poco válidos o confiables, no seguir un protocolo estandarizado, dar instrucciones poco claras, utilizar un instrumento que no está validado para una población particular, entre otros), el valor verdadero de lo que se desea medir va a disminuir, afectando la calidad del dato, su análisis, interpretación y conclusiones derivadas de este^[492].

La validez se refiere a que un instrumento de medición mida lo que se supone que debe medir; mientras que la confiabilidad, se refiere a la consistencia en el tiempo de la medición^{[515][777]}. La validez también se la define como el grado en que la evidencia y la teoría respaldan la interpretación de la medición^[58]. En términos de rendimiento deportivo, se podría decir que un protocolo válido es aquel que se asemeja lo más posible al rendimiento que se está simulando, mientras que la confiabilidad o fiabilidad se refiere a la variación de un protocolo^[190].

La confiabilidad es una medida importante porque proporciona una idea de la variación biológica y técnica de un protocolo de medición^[190]. En términos de jerarquía, se considera que la confiabilidad es más importante que la validez porque no existiría validez sin confiabilidad^[777].

Cuando se mide una variable, se sabe que la distancia entre una puntuación observada y la puntuación teóricamente “real” se le denomina “error”^[777]. Ese error puede ocurrir por variaciones naturales en la variable que se mide o puede ser causada por algún tipo de error sistemático. Por lo tanto, se puede considerar conceptualmente que cualquier puntaje observado está compuesto de dos partes: una puntuación verdadera y un error.

Cuando el error es cero, el puntaje observado será igual a la puntuación teóricamente verdadera, pero como en la vida real se desconoce un puntaje verdadero, éste se puede estimar determinando el error de la medición y restándolo de la puntuación obtenida^[777]:

Resumen de las relaciones entre el puntaje medido, el puntaje verdadero y el error.	$\text{Puntaje medido (X)} = \text{Puntaje verdadero (Y)} + \text{Error (E)}$ <p>Si $X = Y + 0$, entonces $X = Y$, nos acercamos al valor real De la misma manera, podemos estimar $Y = X - E$</p>
--	---

Existe otro caso relacionado a la confiabilidad, y se refiere a cuando más de una persona realiza la observación de una característica y se desea conocer si ambas obtuvieron resultados similares o idénticos. A ese caso se le llama objetividad o concordancia entre observadores (“inter-rater”)^{[154][449][492]}. Para profundizar en este tema, se sugiere consultar una excelente revisión^[440], en donde se presentan dos formas de evidencia utilizadas para reportar la concordancia absoluta, a través de valores sin escala y con escala. Los índices o estadísticas sin escala (e.g., desviación cuadrática media [RMSD], el índice de desviación total [TDI], la probabilidad de cobertura [CP] y

los límites de acuerdo al 95% [LoA]) no dependen de la variación entre sujetos y generalmente se reportan en su unidad de medida original, mientras que los valores escalados (e.g., coeficiente de correlación intraclase [ICC, por sus siglas en inglés], coeficiente de concordancia de Lin) dependen de la variación entre sujetos y tienen un valor máximo de +1.0^[440].

La validez y la confiabilidad tienen consecuencias relevantes en los análisis estadísticos y en las conclusiones que se deriven de esos datos. Existen sitios gratuitos en internet en donde se pueden observar hojas de cálculo para estimar puntualmente la validez y la confiabilidad en las ciencias del deporte^[346]. Ambos constructos influyen en la estandarización y elaboración de normas de variables de interés en las Ciencias del Movimiento Humano. El proceso de estandarización se refiere a la sistematización estratificada de una variable (e.g., fuerza, VO₂máx, depresión, ansiedad, vigor, imagen corporal) en categorías de sexo, edad, nivel educativo, profesión, área geográfica, país, región, urbano-rural, entre otras^[170].

La normalización o elaboración de normas permite tener una interpretación válida del significado de los puntajes de una persona en una prueba estandarizada para poder compararla con la de otras personas con características similares; como, por ejemplo, cuando un puntaje obtenido en una clase de Educación Física por un estudiante se compara con las normas nacionales para esa variable de interés^[262]. Por lo tanto, la falta de validez y confiabilidad en la investigación representa un problema ético que debe atenderse responsablemente en el reporte de datos en los artículos científicos^[517].

2.4.1. Evidencia de validez

La evidencia de validez puede definirse de múltiples maneras (Tabla 1) y ha sido desarrollada conceptualmente desde el siglo pasado^[58]. Se considera que la validez se refiere al grado en que la evidencia empírica y la teoría apoyan la interpretación de las puntuaciones de las pruebas relacionadas con un uso específico^[581]. Por ejemplo, se puede obtener validez por evidencia relacionada al contenido ("content-validity") del objeto de estudio; es decir, a la revisión de la literatura y de grandes cantidades de ítems acerca de un constructo en particular (e.g., personalidad, ansiedad competitiva, imagen corporal)^[170].

En el ámbito del rendimiento humano, algunas veces a este tipo de validez se le conoce como validez de constructo ("construct validity"), y representa el grado en que un protocolo es capaz de medir un constructo hipotético, en este caso, el rendimiento físico^[190]. También se puede obtener evidencia de validez lógica ("logic validity") o aparente ("face-validity") cuando un instrumento ha sido exitosamente piloteado o probado y se han seleccionado los ítems que formarán parte de la prueba final y los que se eliminarán de ésta^{[170][190]}.

La evidencia de validez discriminante ("discriminant validity"), se refiere a la magnitud en que un comportamiento relevante o prueba se ejecuta de manera diferenciada en grupos de personas de acuerdo a expectativas o relaciones hipotéticas que se tengan de estas (e.g., personas con condiciones especiales de aprendizaje, con amputaciones, atletas de élite, personas sedentarias, personas con alguna capacidad especial)^[170].

Tabla 1. Ejemplos de evidencia de validez y análisis estadísticos reportados en publicaciones en las Ciencias del Movimiento Humano.

Referencia	Tipo de validez	Técnica de análisis estadístico
Beato et al. ^[65]	Constructo	a. Correlación de Pearson (r), interpretada como <0.1 = trivial; $0.1-0.3$ = pequeña; $0.3-0.5$ = moderada; $0.5-0.7$ = grande; $0.7-0.9$ = muy grande; y $0.9-1.0$ = casi perfecta.
Liu et al. ^[437]	Convergente y discriminante	a. Análisis factorial: <ul style="list-style-type: none"> i. Análisis factorial exploratorio (EFA): se utilizó la prueba Kaiser-Meyer-Olkin (KMO) y la prueba de esfericidad de Bartlett para medir la adecuación de las muestras si $KMO = 0.5$ y 1.0. ii. Análisis de componentes principales (PCA): rotación ortogonal (varimax) iii. Análisis factorial confirmatorio (CFA): El ajuste del modelo incluyó Chi-cuadrado (CMIN/DF) < 3.00, error cuadrático medio de aproximación (RMSEA) < 0.05, índice de ajuste incremental (IFI) > 0.900, índice de Tucker-Lewis (TLI) > 0.900, índice comparativo de ajuste (CFI) > 0.900. b. Correlación de Spearman. Se consideró como apropiado si la correlación entre cada dimensión y el valor total de la escala > 0.40 .
Grootswagers et al. ^[300]	Relativa Concurrente	a. Prueba t-student de medidas repetidas para comparar las lecturas de dinamometría manual con las obtenidas en un equipo de dinamometría marca Biodex. b. Gráficos de Bland-Altman para evaluar la concordancia entre dinamometría manual y Biodex y para inspeccionar un posible sesgo relacionado con la fuerza. c. Correlación de Pearson (r) para evaluar la validez de las lecturas máximas. d. ICC para evaluar la validez absoluta entre múltiples mediciones.
		Continúa...

Tabla 1. Continuación.

Referencia	Tipo de validez	Técnica de análisis estadístico
Duncan et al. ^[237]	Factorial	a. Análisis factorial: i. Análisis factorial confirmatorio (CFA): El ajuste del modelo incluyó valores de RMSEA entre 0.05 y 0.08, y valores de CFI y TLI > 0.90 para considerarlos como buen ajuste. ii. Se utilizó el criterio de información Bayesiano para comparar los modelos, en donde los valores bajos indicaron modelos más ajustados.
Shakespear-Druery et al. ^[642]	Concurrente	a. Se estudió la concordancia relativa con ICC y r de Spearman. i. La ICC se interpretó como virtualmente carente de acuerdo (0.00–0.10), acuerdo pequeño (0.11–0.40), acuerdo justo (0.41–0.60), acuerdo moderado (0.61–0.80) y acuerdo sustancial (≥ 0.80) ^[649] . ii. La r de Spearman se interpretó como de bajo acuerdo (< 0.30), moderado acuerdo (0.30–0.50) y alto acuerdo (> 0.50) ^[178] .

La evidencia de validez factorial ("factorial validity"), permite determinar conceptual y estadísticamente, la dirección y la magnitud en que los ítemes o partes de una prueba se agrupan y se relacionan entre sí. Este tipo de evidencia se obtiene por medio de análisis multivariados de datos, como por ejemplo en análisis factorial exploratorio y el análisis factorial confirmatorio, los cuales tienen como propósito principal reducir una cantidad de ítemes a dimensiones que compartan algo en común^[170].

La evidencia de validez de criterio ("criterion validity"), se refiere a cuando un dato obtenido con una prueba se compara con lo que se considera como el estándar de oro ("gold standard"). Entre más coincidencia o correlación exista entre los puntajes, se puede afirmar que existe evidencia de criterio^[170], como cuando se menciona el caso de la validez convergente, concordante o concurrente ("concurrent validity"), que es la magnitud en que una prueba o un instrumento nuevo se correlaciona con un instrumento existente que mide el mismo constructo o uno similar^{[170][190]}. Otro caso de validez de criterio es el de la validez predictiva ("predictive validity"), como se podría hacer al utilizar una prueba de VO₂máx para predecir el rendimiento en una prueba de carrera de maratón.

2.4.2. Evidencia de confiabilidad

El error mínimo de medición, también conocido como confiabilidad, que se genera cuando se recopilan datos utilizando una escala o instrumento que arroja datos continuos (e.g., intervalo, proporción) se puede desglosar en el sesgo sistemático (e.g., aprendizaje de la prueba, efectos de fatiga en la prueba), y en el error aleatorio (i.e., variaciones biológicas o mecánicas)^[40]. Estos dos componentes de la confiabilidad deben cuantificarse y comprenderse en el contexto del estudio (e.g., salud, rendimiento físico) para poder emitir un juicio de valor, ya que la significancia estadística ($p < 0.05$) solamente brinda información parcial.

Al estudiar la variabilidad de las puntuaciones obtenidas en al menos dos ocasiones, se puede obtener un indicador de la confiabilidad, fiabilidad, consistencia o precisión de las mediciones. Así, si la variabilidad de las mediciones obtenidas es grande, se considerará que los valores son imprecisos y, en consecuencia, poco confiables^[581]. Para lograr cuantificar la confiabilidad, generalmente se utilizan métodos de confiabilidad relativa y de confiabilidad absoluta (Tabla 2).

En cuanto a la confiabilidad relativa, se ha descrito que los métodos basados en coeficientes de correlación y regresión están muy influenciados por el rango de valores medidos, pues la correlación ignora las diferencias en los promedios y las desviaciones estándar de los dos conjuntos de datos. Así, la correlación puede ser bastante alta cuando las puntuaciones no son muy consistentes^[54]. Es por ello que se sugiere a los investigadores tener cautela al interpretarlos o utilizar otros indicadores^[40].

Especialmente, se recomienda precaución al concluir una confiabilidad relativa aceptable incluso si una correlación es > 0.9 , que tradicionalmente ha sido aceptado como un coeficiente alto. También se previene al extrapolar los resultados de una correlación test, re-test a una nueva muestra de individuos involucrados en un experimento, o cuando se comparan las correlaciones test, re-test entre diferentes estudios de confiabilidad.

Tabla 2. Ejemplos de evidencia de confiabilidad y análisis estadísticos reportados en publicaciones en las Ciencias del Movimiento Humano.

Referencia	Tipo de confiabilidad	Técnica de análisis estadístico
Beato et al. ^[65]	Test, Re-test	<p>a. Coeficiente de correlación intraclase (ICC) para la confiabilidad relativa y su respectivo IC95%. El ICC se interpreta como excelente si > 0.9, bueno si $0.9 > 0.8$, aceptable si $0.8 > 0.7$, cuestionable si $0.7 < 0.6$, pobre si $0.6 < 0.5$, e inaceptable si < 0.5^[40]. Otros^[409] sugieren interpretarlo como pobre si < 0.50, moderado si está entre 0.50 y 0.75, bueno si está entre 0.75 y 0.90, y excelente si > 0.90.</p> <p>b. Coeficiente de variación (CV%) para la confiabilidad absoluta. Interpretado como bueno ($< 5\%$) y aceptable (5%-10%)^[184].</p>
Blagrove et al. ^[80]	Test, Re-test	<p>a. Error típico de la medición (ETM).</p> <p>b. Cambio mínimamente detectable (CMD) y el IC95%, calculado como $ETM \times 1.96 \times \sqrt{2}$.</p> <p>c. Tamaños de efecto interpretados como trivial si < 0.20; pequeño si 0.20–0.59; moderado si 0.60–1.19 y grande si ≥ 1.20.</p> <p>d. ICC con IC95%.</p> <p>e. Coeficiente Kappa (κ) como medida de concordancia e interpretado como ligero si ≤ 0.20; justo si 0.21–0.40; moderado si 0.41–0.60; sustancial si 0.61–0.80 y casi perfecto si > 0.80.</p>
Burland et al. ^[127]	Entre sesiones	<p>a. $ICC_{(3,1)}$ con IC95% por medio de un modelo de efectos mixtos para la concordancia absoluta^[409], con valores interpretados como excelente si > 0.75, bueno si 0.60–0.74, aceptable si 0.40–0.59, y pobre si < 0.40^{[169][170][606]}.</p>
		Continúa...

Tabla 2. Continuación.

Referencia	Tipo de confiabilidad	Técnica de análisis estadístico
Chalak et al. ^[154]	Concordancia entre dos investigadores	b. Coeficiente Kappa (κ). Se interpreta que si $\kappa < 0.40$ es pobre, si $\kappa = 0.40-0.59$ es aceptable, si $\kappa = 0.60-0.74$ es bueno, y si $\kappa = 0.75-1.00$ es excelente ^[170] . El valor $\kappa = 1$ se considera como un acuerdo perfecto entre dos investigadores ^[176] . Se prefiere un κ casi perfecto para casos en que se requiere tomar decisiones muy precisas (e.g, ^[154]).
Liu et al. ^[437]	Consistencia interna	a. Método de las mitades. Se interpreta que si $r > 0.70$ es satisfactorio. c. Alfa de Cronbach. Se interpreta que si $\alpha = 0.41-0.60$ es moderado, si $\alpha = 0.61-0.80$ que existe acuerdo, si $\alpha = 0.81-1.00$ es casi perfecto. En un contexto clínico ^[170] , se interpreta que si $\alpha < 0.70$ es inaceptable, si $\alpha = 0.70-0.79$ es aceptable, si $\alpha = 0.80-0.89$ es bueno, y si $\alpha \geq 0.90$ es excelente.
		Continúa...

Tabla 2. Continuación.

Referencia	Tipo de confiabilidad	Técnica de análisis estadístico
Carazo-Vargas et al. ^[143]	Concordancia entre dos instrumentos	<ul style="list-style-type: none"> a. Gráfico de Bland–Altman^[84]. b. Índice sin escala RMSD: calculado como la raíz cuadrada de la suma de las desviaciones al cuadrado entre dos evaluadores o métodos, divididos por el tamaño de la muestra; $\sqrt{(\sum(X_i - Y_i)^2)/n}$, en donde X_i es el puntaje de la persona en la primera medición y Y_i es el puntaje de la persona en la segunda medición, y n es la cantidad total de personas de las que se tienen mediciones^[54]. c. Índice sin escala TDI: representa la diferencia absoluta entre mediciones o instrumentos. Es un valor que se puede definir arbitraria o estadísticamente. Se calcula como $TDI_{\#} = Q \sqrt{(\mu_d^2 + \sigma_d^2)}$, en donde Q es la inversa de la distribución acumulada para la distribución normal estándar de $(1 + \#)/2$, μ_d^2 = es el cuadrado de las diferencias promedio, y σ_d^2 = la varianza de las diferencias^[58]. d. El valor predeterminado de la diferencia aceptable para el RMSD y el TDI utilizado como evidencia de concordancia se estimó en $\# = 0.20$ (el $\#$ varía de acuerdo al contexto de la investigación). e. Índice sin escala LoA: $d \pm 1.96 \times DE_d$, en donde d son las diferencias promedio y DE_d es la desviación estándar de las diferencias^[84]. f. Índice escalado de coeficiente de correlación de concordancia de Lin (CCC)^[432]. Se interpretó como casi perfecto si >0.99, sustancial si $0.95-0.99$, moderado si $0.90-0.95$, y pobre si <0.90^[474]. g. Análisis de modelos mixtos. h. Enfoque Bayesiano.
		Continúa...

Tabla 2. Continuación.

Referencia	Tipo de confiabilidad	Técnica de análisis estadístico
Cormack et al. ^[184]	Test, Re-test	<ul style="list-style-type: none"> a. Error típico de la medición (ETM) general e IC90%. b. ETM intra diario: am vs. pm. c. ETM inter diario: día 1, día 2. d. Coeficiente de variación (CV%): se interpretó como confiable si el CV = 10%. e. Cambio mínimamente significativo (CMS): $0.2 \times DE$ entre sujetos. Se interpretó que las variables eran capaces de detectar el CMS si $ETM \leq CMS$.
Godwin et al. ^[288]	Test, Re-test	<ul style="list-style-type: none"> a. Datos transformados logarítmicamente: permite estimar errores similares en un amplio rango de valores^[346]. b. ICC con IC90%. c. Coeficiente de variación (CV%): se interpretó como pequeño si $\leq 10\%$^[106]. d. Combinación de ICC y CV%: interpretado como pequeño si $ICC > 0.67$ y $CV < 10\%$, moderado si $ICC < 0.67$ y $CV > 10\%$, y grande si $ICC < 0.67$ y $CV > 10\%$^[113]. e. Error estándar de la medición (SEM): se calculó con la fórmula $DE \sqrt{1 - ICC}$. f. Diferencia mínimamente detectable (MDD). Se calculó con la fórmula $SEM \times 1.96 \times \sqrt{2}$^[743]. g. La escala de magnitud para las estadísticas de efectos se interpretó como trivial (< 0.10), pequeña (0.10 - 0.29), moderada (0.30 - 0.49), grande (0.5 - 0.69), muy grande (0.7 - 0.89) o casi perfecto (0.9 - 0.99)^[344].
		Continúa...

Tabla 2. Continuación.

Referencia	Tipo de confiabilidad	Técnica de análisis estadístico
Grootswagers et al. ^[300]	Intra e inter evaluadores	<ul style="list-style-type: none"> a. Error estándar de medición (SEM). b. Test, re-test el mismo día y se calculó el ICC_(2,1) e IC95% c. El ICC_(2,1) se interpretó como pobre (<0.5), moderado (0.51-0.75), bueno (0.90) y excelente (>0.91).
Moreno-Pérez et al. ^[512]	Intra e inter evaluadores	<ul style="list-style-type: none"> a. Combinación de ICC e IC95%: <ul style="list-style-type: none"> i. ICC_(3,1): de efectos mixtos bidireccionales para las mediciones intra-evaluador. ii. ICC_(2,1): de efectos aleatorios bidireccionales para mediciones entre evaluadores. iii. El ICC se interpretó como malo (<0.50), moderado (0.50-0.79), bueno (0.79-0.90) y excelente (>0.90). iv. La prueba se consideró como confiable si ICC >0.75 b. SEM en valores absolutos (en N) y expresado como coeficiente de variación (%SEM) para determinar la magnitud de la variabilidad entre test y re-test. <ul style="list-style-type: none"> i. Valor de corte de %SEM < 10% para considerar la prueba como con bajo error. c. MDD: Se calculó con la fórmula SEM x 1.96 x $\sqrt{2}$

Por su parte, los métodos utilizados para describir la confiabilidad absoluta no se ven afectados por el rango de los valores medidos e incluyen el error estándar de las mediciones (SEM), el CV, los LoA y la RMSD^{[40][54]}. Se ha descrito que estas estadísticas son apropiadas para comparar la confiabilidad entre diferentes instrumentos de medición en estudios distintos. Se pueden utilizar para predecir la magnitud de un cambio “real” en atletas individuales y se emplean para estimar la potencia estadística para un experimento de medidas repetidas^[40].

La mayoría de los métodos para calcular SEM y CV representan aproximadamente el 68% del error que en realidad está presente en las mediciones repetidas de un individuo. Por su parte, los LoA representan las diferencias test re-test para el 95% de una población. El gráfico de Bland-Altman asociado muestra el error de medición de forma esquemática (i.e., permite observar el acuerdo o desacuerdo entre mediciones) y ayuda a identificar visualmente la presencia de heterocedasticidad^[84]. Si existe evidencia de heterocedasticidad o de “no normalidad”, se recomienda transformar logarítmicamente los datos y citar el sesgo y el error aleatorio como proporciones. Esto permite comparaciones simples de confiabilidad entre diferentes instrumentos de medición.

La RMSD representa cuánto cambian típicamente las puntuaciones de una medición a la siguiente, y a diferencia de los coeficientes de correlación, toma en consideración todas las diferencias entre dos conjuntos de puntuaciones, sobre la base de una definición absoluta de concordancia^[54]. El valor mínimo de RMSD es 0, que corresponde a cuando cada número en la primera medida es idénticamente igual al número correspondiente en la segunda medida.

El valor máximo del RMSD, que es una estimación puntual, es la diferencia entre las puntuaciones más altas y más bajas posibles. Este valor ocurre cuando todas las medidas son iguales al valor más alto o más bajo, y cuando los valores son opuestos extremos (i.e., si un número es 10, el otro es 1).

La evidencia de confiabilidad puede definirse por la consistencia interna; es decir, la fuerza en que se asocian los ítems de una prueba, dominio, subdominio o subprueba. Para la mayoría de las pruebas psicológicas, se calcula mediante el α de Cronbach para valores continuos y con la prueba Kuder-Richardson (KR-20) para valores dicotómicos^[170]. En otras áreas de las Ciencias del Movimiento Humano se utiliza la confiabilidad test re-test, la estabilidad entre investigadores, coeficiente κ sin ponderar (variables categóricas, modelo conservador) y κ ponderado (variables nominales, modelo más liberal), dependiendo del contexto (e.g., clínico, rendimiento deportivo)^{[169][176][177]}, el ICC para medir la dimensionalidad de los datos medidos^[170], los LoA y el CV%^[190].

Los LoA se refieren a la variación expresada en términos absolutos que se representan como un intervalo con un límite inferior y un límite superior; es decir, es un intervalo de tolerancia o de error que permite realizar una comparación subjetiva del valor promedio con respecto a sus posibles valores inferiores y superiores^[190].

En el área del rendimiento deportivo, se ha demostrado que los protocolos de tiempo hasta el agotamiento tienen un $CV > 10\%$, mientras que las pruebas contrarreloj son más confiables pues tienen un $CV < 5\%$ ^[190]. Para el ICC actualmente se recomienda una nomenclatura que considera seis formas que se reportan entre paréntesis y que significan el modelo y el tipo de coeficiente^[409]. Es importante incluir esa nomenclatura en los manuscritos enviados a publicación a revistas científicas.

Así, el $ICC_{(1,1)}$, se refiere a un modelo de una vía de efectos aleatorios, de acuerdo absoluto y con una medición o una sola persona. El $ICC_{(2,1)}$, se refiere a un modelo de dos vías de efectos aleatorios, de acuerdo absoluto y con una medición o una sola persona. El $ICC_{(3,1)}$, se refiere a un modelo de dos vías de efectos mixtos, de consistencia y con una medición de una sola persona (e.g., Burland et al. ^[127], Tabla 2). De manera similar, el $ICC_{(1,k)}$, se refiere a un modelo de una vía de efectos aleatorios, de acuerdo absoluto y con varias mediciones o el promedio de varias personas. El $ICC_{(2,k)}$, se refiere a un modelo de dos vías de efectos aleatorios, de acuerdo absoluto y con varias mediciones o el promedio de varias personas. Finalmente, el $ICC_{(3,k)}$, se refiere a un modelo de dos vías de efectos mixtos, de consistencia y con varias mediciones o el promedio de varias personas.

En síntesis, los conceptos de validez y confiabilidad interactúan entre sí y no deberían ser vistos por separado. Un protocolo de medición tiene que ser confiable y válido; pero puede que un protocolo sea confiable pero no válido. La relación entre confiabilidad y validez es tan cercana que un coeficiente de correlación de confiabilidad es el cuadrado del coeficiente de correlación de validez^[190]. El diseño y ejecución de estudios en las Ciencias del Movimiento Humano requiere de la selección adecuada de los mejores instrumentos de medición y de la cuantificación del error de medición para que las conclusiones derivadas de los mismos posean el menor sesgo posible.

2.5. Tamaño de la muestra y remuestreo

En las investigaciones científicas muy rara vez se puede utilizar una población de personas completa (i.e., N), sino que se seleccionan muestras representativas de esa población (i.e., n). Lo que ocurre con la muestra de participantes de una población determinada, se infiere a la población; y este es el fundamento de la estadística inferencial que se obtiene por medio de las pruebas de hipótesis tradicionales (i.e., frecuentistas). Sin embargo, como se ha descrito anteriormente, existen algunas limitaciones al utilizar las tradicionales pruebas de H_0 ^[348].

Por ejemplo, la H_0 casi nunca es cierta, por lo que rechazarla es relativamente poco informativo. Además, las pruebas de significancia dependen en gran medida del tamaño de la muestra; así, cuando el tamaño de la muestra es pequeño, los efectos fuertes e importantes pueden no ser significativos (i.e., se comete un error de Tipo II). Además, los IC95% del estimador son más amplios, lo que reduce la precisión de la estimación (Figura 1)^[307]. Por otra parte, cuando los tamaños de muestra son grandes, incluso los efectos triviales pueden tener valores de significancia p muy altos (e.g., $p = 0.0001$)^[425].

Figura 1. Efecto del tamaño de la muestra en los estimadores y su precisión.



Estudiar muestras pequeñas tiene aspectos positivos y negativos. Entre los aspectos positivos se puede mencionar que pocos participantes podrían significar menores costos en análisis de muestras, menor logística, y menor tiempo para medirlos y analizarlos. Incluso, las hipótesis exploratorias se podrían estudiar con pocos participantes^[307]. Por otro lado, el principal aspecto negativo de estudios con muestras pequeñas es la interpretación de los resultados; es decir, su estimador y el intervalo de confianza asociado, así como el valor de significancia p ^[307].

También existen problemas cuando los estudios no miden la cantidad de personas que deberían (i.e., n insuficiente)^[225]. Se ha descrito que la tasa de error de tipo II aumenta, por lo que si se detectan efectos estadísticamente significativos es probable que se sobrestime el tamaño del efecto de la población y una mayor proporción de efectos estadísticamente significativos serán errores de tipo I. Esto afecta la potencia, y por lo tanto, los estudios con poca potencia son menos replicables. El tamaño de muestra también se relaciona con el tamaño de efecto real que se desee encontrar, en donde cuanto más pequeño sea ese tamaño de efecto, mayor debe ser el tamaño de la muestra debido a que es más difícil distinguir entre lo que se consideraría un efecto real y una variación aleatoria^[307].

Por consiguiente, ¿cuál debe ser el tamaño de la muestra? Esa es la pregunta que se formulan los investigadores antes de comenzar un estudio, o al menos, eso deberían hacer. A eso se le llama cálculo del tamaño de la muestra *a priori*. La respuesta no es directa, y a veces es poco clara, pues existen al menos dos enfoques para determinar ese número de personas. Por una parte, se tiene el enfoque de la "potencia", y por otro lado, el enfoque de la "precisión"^[3].

Un estudio que utilice pruebas de hipótesis frecuentistas realiza un cálculo de potencia para determinar cuántos participantes se requerirían para rechazar la H_0 suponiendo que está presente un efecto de un tamaño determinado (e.g., trivial, pequeño, moderado, grande, extremadamente grande). Esto quiere decir que, por ejemplo, si existe un efecto del tratamiento (e.g., programa de entrenamiento), un cálculo de potencia determinará aproximadamente cuántos participantes se necesitarían para detectar ese efecto en un porcentaje dado del tiempo, generalmente el 80% del tiempo^[3]. Por otra parte, el enfoque de precisión implica estimar cuántos participantes serían necesarios para que el intervalo de confianza frecuentista o el intervalo creíble bayesiano resultante de un análisis estadístico tuviera una cierta amplitud. La implicación es que un intervalo de confianza más estrecho o un intervalo creíble permite una estimación más precisa de dónde podría estar el parámetro de población "verdadero" (e.g., la diferencia de promedios).

Una de las críticas al enfoque de potencia se refiere a que no identifica el tamaño mínimo de la muestra que garantizaría una estimación precisa del parámetro de la población^[472]. Por lo tanto, se ha recomendado utilizar la denominada "precisión en la estimación de parámetros", la cual estima el tamaño de la muestra con base en lo que se requiere para lograr un determinado intervalo de confianza^[3]. El rango del intervalo de confianza es proporcional al tamaño de la muestra, por lo que para dividir el intervalo a la mitad, el tamaño de la muestra debe aumentar por un factor de cuatro, lo que hace que se necesiten tamaños de muestra muy grandes para obtener una alta precisión. Esto en la práctica es poco realista, pues no siempre se cuenta con muestras grandes en ciertos ámbitos (e.g., atletas de élite), por lo que se sacrificaría la precisión de la estimación, que a su vez tendría un impacto en las conclusiones a las que se podría llegar.

Una alternativa al enfoque de potencia es usar “pruebas secuenciales”, lo cual permite no solamente calcular un tamaño de muestra fijo, sino tamaños de muestra más pequeños. Las pruebas secuenciales implican recopilar datos hasta que se cumple una regla de parada *a priori*, pero también tienen la debilidad de que producen muchos errores de tipo I. Es por ello que se recomienda utilizar métodos bayesianos; es decir, al calcular el factor de Bayes o se estima el intervalo creíble de la distribución posterior (i.e., estimación de parámetros); los enfoques bayesianos tienen la libertad de monitorear los datos con la frecuencia que deseen mientras se van recopilando^[3]. El enfoque bayesiano permite determinar cuándo debe detenerse la recopilación de datos una vez que haya pruebas sólidas (i.e., determinadas por un factor de Bayes particular) para H_0 o H_1 . Así, el investigador describe *a priori* el factor de Bayes en el que terminará la recopilación de datos y a medida que se acumulan los datos, el factor de Bayes se supervisa continuamente y, una vez que alcanza el umbral establecido, se finaliza la recopilación de datos. También se puede establecer un n mínimo y máximo y determinar la probabilidad de obtener evidencia engañosa (i.e., falsos positivos/negativos).

Cuando no haya sido posible calcular el tamaño de la muestra y se desea llegar a conclusiones más sólidas a partir de los datos recolectados, es posible utilizar el método de remuestreo (conocido en del inglés como “bootstrap”)^[137]. El remuestreo permite obtener conclusiones sólidas solo a partir de los datos recolectados sin tener que cumplir con los supuestos acerca de la población muestreada, y se les puede ejecutar utilizando programas estadísticos. El proceso permite, dada una muestra adecuada, obtener muchas características de la población a partir de una sola muestra (i.e., toma muestras aleatorias repetidas de los datos originales). Es decir, el principio en el que se basa el remuestreo es que se usa la muestra con la que se trabajó, y ésta se puede remuestrear repetidamente al azar (llamadas pseudo muestras) para estimar las características de una población de origen a la que uno no tiene acceso^[137]. Con las pseudo muestras se pueden calcular estimadores y sus respectivos intervalos de confianza.

En síntesis, independientemente del enfoque que se utilice para estimar el tamaño de la muestra, se recomienda buscar asesoría estadística y tomar en consideración la cantidad de personas que podrían abandonar el estudio debido a causas directamente (e.g., el programa de entrenamiento fue muy intenso) o indirectamente relacionadas con la intervención (e.g., enfermedad repentina, horario difícil de cumplir, un viaje)^[97]. Adicionalmente, es relevante considerar aspectos positivos y negativos de contar con una muestra pequeña y una muestra grande. Finalmente, para quienes deseen hacer los cálculos tomando en consideración el diseño del estudio, pueden consultar el artículo de Charan y Biswas^[157] y el de Santabárbara^[631], quienes además ofrecen opciones de software gratuito (el lector debe verificar si los URL están disponibles) y una sintaxis para obtener el tamaño de muestra deseada. Por ejemplo, Santabárbara^[631] ofrece un ejemplo de sintaxis o código en el programa IBM-SPSS (IBM-SPSS Inc., Armonk, NY) para determinar el tamaño de la muestra si se parte de un coeficiente de correlación que estimamos en un estudio correlacional (Cuadro 1).

Cuadro 1. Ejemplo de sintaxis para determinar el tamaño de muestra a partir de un coeficiente de correlación (r) estimado, la significancia estadística (sig) y la potencia. Los valores en negrita dentro del cuadro se pueden modificar para obtener el tamaño de la muestra. El programa SPSS crea una base de datos con la información proporcionada en la sintaxis y arrojará el valor del tamaño de la muestra. Modificado de Santabábara^[631].

```
DATA LIST FREE/r sig potencia.
begin data.
0.80 0.001 0.80
end data.
* Transforma r en el valor z: fz.
compute fz = 0.5*ln((1+r)/(1-r)).
* Calcula el valor crítico de la curva normal para el nivel de significancia: critz_sig.
compute critz_sig = abs(idf.normal(sig/2,0,1)).
* Calcula el valor crítico de la curva normal para la potencia: critz_potencia.
compute critz_potencia = abs(idf.normal(1-potencia,0,1)).
* Calcula el tamaño de la muestra (n).
compute n = rnd((((critz_sig + critz_potencia)/fz)**2)+3,1,0).
formats r (f10.2) sig (f10.4) potencia (f10.2) / n (f8).
* Muestra el tamaño de muestra necesario.
list r sig potencia n.
```

En el ejemplo mostrado en el cuadro 1, los investigadores proponen que debería ser posible encontrar un coeficiente de correlación $r = 0.80$, a un nivel $\alpha = 0.001$ y un valor $\beta = 0.80$. Al ejecutar la sintaxis, se encuentra que el tamaño de la muestra $n = 19$ personas. Esos valores se pueden modificar a conveniencia; por ejemplo, si $\alpha = 0.05$, entonces el $n = 10$ personas.

Otra opción para calcular el tamaño de la muestra es utilizar el programa *G*Power*^{[255][256]}, el cual permite determinar la cantidad de participantes que se requieren en un estudio dependiendo del diseño de la investigación, el cambio esperado, el nivel de significancia estadística y otros elementos propios de la estadística. Sin embargo, se recomienda consultar a profesionales en estadística para acompañar el proceso de la determinación del tamaño de la muestra.

2.6. Datos perdidos en la investigación

Cuando se recolectan datos para completar una investigación, ya sea en el campo o en el laboratorio, algunas veces no se logran obtener todos los datos que se planeaba analizar. En algunos momentos, ocurren eventos fortuitos que no permiten recolectar la totalidad de los datos, como, por ejemplo, el mal funcionamiento de un equipo o la poca sensibilidad para grabar datos^[150], la pérdida de algún formulario de recolección de datos, el olvido de la persona recolectora de datos, o simplemente porque el dato no se anotó con buena letra o se perdieron los datos por descuido.

La pérdida de datos afecta la potencia de los análisis estadísticos que se realicen independientemente de si el diseño involucra varios participantes o si el diseño es de sujeto único^[412]. Por consiguiente, la pérdida de datos tiene implicaciones directas en las conclusiones y las decisiones que se basan en esos análisis^[552]. En ese contexto, los investigadores afrontan una dura decisión; vuelven a recolectar datos, lo cual puede que sea imposible por la oportunidad única de contar con los participantes, buscan un método aceptado para sustituir los datos perdidos, o analizan los datos de los participantes que tienen datos completos^[356].

Existen tres mecanismos que explican la pérdida de datos^{[372][547][621]}:

- a. Los datos perdidos que faltan completamente al azar (MCAR, por sus siglas en inglés) ocurren si esa "pérdida" es independiente de todos los demás datos. En ese contexto, los participantes con datos perdidos tienen la misma distribución para el resultado primario que aquellos con datos completos. MCAR causa errores estándares grandes debido al reducido tamaño de la muestra, pero no causa sesgo o error sistemático (i.e., sobrestimación de los beneficios y subestimación de los perjuicios). Para su tratamiento se recomienda el análisis completo de casos, la imputación única y la imputación múltiple.
- b. Si la pérdida de datos depende de la variable observada o de covariables, pero no del patrón de los datos, se dice que los datos perdidos faltan al azar (MAR, por sus siglas en inglés). MAR permite la predicción de los valores faltantes en función de los participantes con datos completos. En presencia de MAR, los métodos para la sustitución de datos como la imputación múltiple o la máxima verosimilitud directa de información completa pueden conducir a resultados no sesgados.
- c. Si el mecanismo depende de los datos faltantes y esta dependencia se mantiene incluso teniendo los datos observados, entonces los datos se clasifican como perdidos no al azar (NMAR, por sus siglas en inglés). Para su tratamiento se recomienda el análisis de sensibilidad.

En la literatura se describen métodos para la sustitución de datos. Por ejemplo, en la literatura médica los métodos más frecuentemente utilizados para manejar los datos perdidos para el análisis primario son el análisis de casos completos, la imputación única, los métodos basados en modelos estadísticos (e.g., modelos mixtos, ecuaciones de estimación generalizadas) y la imputación múltiple^[372]. Actualmente, se recomiendan los métodos basados en modelos estadísticos, así como el uso de software especializado para el manejo de datos perdidos (e.g., función Quark de semTools en lenguaje R)^[510].

En el análisis de casos completos se realiza un análisis estadístico basado en los participantes que cuentan con datos completos. Consecuentemente, los participantes a los que les faltan datos se excluyen del análisis. Cuando se utiliza la imputación única, los valores perdidos se reemplazan por un valor definido por una determinada regla que decide el investigador. Una manera de hacerlo es usando el último valor siguiente; es decir, los valores faltantes de un participante se reemplazan por el último valor observado del participante. Otra forma es mediante la peor observación siguiente; es decir, los valores perdidos de un participante se reemplazan por el peor valor observado del participante. Finalmente, se puede utilizar la imputación media simple, en la que los valores perdidos se reemplazan por el promedio. Sin embargo, la imputación única subestima la variabilidad porque cada valor no observado tiene el mismo peso en el análisis que los valores observados conocidos. Por lo tanto, la imputación única es a menudo un método potencialmente sesgado y debe usarse con gran precaución^[239]. Con el método de imputación múltiple se realiza una copia de la base de datos original y se reemplazan los valores faltantes con estimaciones probables de los valores que hubieran existido en las celdas vacías si se hubieran recolectado los datos^[510].

En la literatura proveniente de las Ciencias del Movimiento Humano, también se cuenta con argumentos para utilizar métodos de sustitución de datos^{[70][299]}. Por ejemplo, los preparadores físicos encuentran el problema de la pérdida de datos debido a ausencias de los atletas a los entrenamientos, por lo que el cálculo de la carga de entrenamiento puede ser poco precisa, lo cual puede ocasionar consecuencias negativas en el proceso de entrenamiento. Se ha indicado que el método de imputación única es cuando se atribuye un valor único a un valor ausente, ya sea eligiendo la media (i.e., promedio) de todos los demás puntajes para ese individuo, utilizando el puntaje anterior o el promedio de varias mediciones anteriores, escogiendo un valor aleatorio o sustituyendo con puntajes relacionados de individuos con características similares (e.g., mismo sexo, misma condición física). Actualmente, también se utilizan técnicas de machine learning (e.g., regresión, árboles de decisión) para la sustitución de valores. Este enfoque, que es bastante práctico, también tiene sus desventajas, principalmente porque puede producir sesgos^[552]. Por ello, se prefiere usar métodos de imputación múltiple (i.e., sustituir varios datos) en lugar de imputación simple.

Otro contexto en el que se “pierden” datos es durante la realización de meta análisis, pues algunas veces los artículos de investigación no reportan valores de estadística descriptiva de forma numérica, sino que lo hacen por medio de gráficos. Debido a esto, en la elaboración de meta análisis se deben descartar esos artículos, lo que provoca una pérdida de información valiosa^[121]. Actualmente, existe software gratuito, entre éstos uno llamado *WebPlotDigitizer*^[613], que sirven para estimar datos a partir de gráficos. Este tipo de programas permiten extraer con una alta precisión los datos crudos a partir de un gráfico; sin embargo, se debe tener precaución porque estos programas son incipientes y necesitan mayor verificación para poder tener mayor certeza de su utilidad.

La sustitución de datos perdidos, ya sea por un método de reemplazo o de imputación, podría ser mal interpretado, pues no consiste en inventar datos, lo cual sería una clara ética y moralmente cuestionable decisión, y en algunos casos, una acción legalmente castigable. En caso de que se tengan que reportar datos perdidos en un manuscrito para enviar a una revista o para alguna tesis, es importante seguir las recomendaciones de Papageorgiou et al.^[547]:

- a. Reporte el número o proporción de valores faltantes para cada variable de interés. Indique también el número de participantes con datos completos. Si fuera posible, indique y analice las posibles causas que explican la pérdida de datos.
- b. Compare casos completos e incompletos: utilice una tabla o figura en la que se comparen las distribuciones de las variables de interés.
- c. Indique la metodología utilizada para el análisis de datos incompletos. Mencione el método de sustitución utilizado (e.g., análisis de caso completo, imputación múltiple, etc.), así como las suposiciones que se hicieron con respecto a la causa de la pérdida de los datos.
- d. Mencione el software y la versión del mismo que se utilizó para manejar los datos perdidos. Además, mencione cualquier cambio a la configuración o características predeterminadas del software o de las funciones que se usaron.
- e. Reporte el número de conjuntos de datos imputados y el número de iteraciones.
- f. Enumere las variables que se incluyeron en el modelo de imputación.
- g. Mencione cualquier función de orden superior de las variables imputadas.
- h. Evalúe la solidez del supuesto MAR realizando un análisis de sensibilidad para varios escenarios NMAR.

2.7. Transformación de datos

Uno de los primeros pasos al analizar los datos recolectados en un proyecto de investigación, consiste en determinar si se debe tomar la decisión de utilizar técnicas de análisis de datos inferencial paramétricas o no paramétricas. Para ello, se requiere conocer si la distribución de los datos es normal o lo más cercana a la normalidad, pues muchos de los supuestos de las pruebas paramétricas (e.g., t-student) se basan en la normalidad de los datos^[85]. Sin embargo, muchas veces los datos no siguen una distribución normal (e.g., variables de origen biológico), o la cantidad de datos es muy reducida (i.e., tamaño de la muestra), lo que atenta contra el teorema del límite central (i.e., los promedios de muestras grandes seguirán una distribución normal), y por lo tanto se deba decidir utilizar técnicas de análisis no paramétrico (e.g., Kruskal-Wallis, Mann-Whitney), las cuales no deben cumplir con ese supuesto de normalidad.

Existen opciones para manejar grupos de datos que carecen de normalidad. Algunos autores recomiendan transformar los datos, de manera que se comporten de manera normal, y así poder utilizar técnicas de análisis paramétricas. Una forma para transformar los datos es la logarítmica, conocida en inglés como *Log transformation*^[189]. También se ha utilizado la raíz cuadrada y el recíproco para transformar los datos, los cuales posteriormente son analizados en lugar de los datos originales^{[85][328]}. Sin embargo, se ha criticado la transformación de los datos, en donde se ha indicado que transformar datos es una forma de hacer trampa^[85].

La transformación logarítmica puede estabilizar la varianza; es decir, ayuda a cumplir el supuesto general de la estadística inferencial relacionado con la variabilidad (i.e., desviación estándar) de una variable dependiente (e.g., salto medido en cm), que se supone es homogénea entre grupos o entre alguna variable predictora^[189]. A pesar de que la transformación logarítmica se ha utilizado y promovido, no siempre logra transformar los datos de manera que queden normalmente distribuidos^{[189][261]}. En realidad, la transformación logarítmica colabora con que la distribución de los datos parezca más normal. Además, otra crítica a los datos analizados con la transformación logarítmica es que los intervalos de confianza calculados son difíciles de interpretar con datos transformados^[85].

Independientemente del método de transformación de datos, se recomienda estudiar la distribución de los puntajes. Posteriormente, si ese estudio indica que no están distribuidos normalmente, se puede explorar cuál es el mejor método para estudiar si la transformación de datos ha sido exitosa. Por ejemplo, se puede explorar si el método de transformación de raíz cuadrada, *Log transformation* o bootstrap producen una distribución normal de datos transformados para posteriormente aplicarle estadísticas paramétricas. En todo caso, los análisis de datos transformados deben ser interpretados en el contexto de datos transformados y no de los datos crudos originalmente recolectados.

En algunas áreas del conocimiento podría ser una práctica habitual esa transformación de datos; pero en las Ciencias del Movimiento Humano pocas veces se utilizan esos procedimientos. Algunos ejemplos se pueden observar en las áreas de la biomecánica (e.g., datos de electromiografía), datos fisiológicos como frecuencia cardíaca, o algunas pruebas de aptitud física y motriz^{[88][238][571][749]}. Se recomienda acudir a profesionales en Estadística para recibir asesoría profesional y poder analizar los datos con un mayor fundamento profesional.

2.8. Técnicas de análisis de datos poco utilizadas

Al planificar una investigación, los investigadores deben conocer el nivel de medición de las variables que van a medir. Identificar el nivel de medición es simple; hay variables que se miden con escalas categóricas y hay variables que se miden con escalas continuas. Las escalas categóricas se dividen en ordinales y nominales (e.g., categorías de sexo, estado civil, niveles de actividad física); y las escalas continuas se dividen en intervalos y razones (e.g., tiempo, distancia, temperatura, consumo de oxígeno, peso, estatura, IMC)^[501]. Con éstas dos últimas se pueden aplicar técnicas de análisis paramétrico siempre que se cumplan supuestos de normalidad. Con las escalas categóricas únicamente se pueden utilizar técnicas de análisis no paramétrico. Así, la selección de técnicas de análisis estadístico requieren conocer si la distribución de los datos posee una forma normal o similar a la normal, o si por el contrario, los datos no se distribuyen normalmente. Cuando la distribución de los datos es normal, se utilizan técnicas de análisis paramétrico; mientras que las técnicas de análisis no paramétrico se utilizan cuando se viola el supuesto de normalidad en la distribución de los datos^[501].

Un reporte no exhaustivo de las técnicas de análisis estadístico reportados en una revista de alto impacto en las Ciencias del Movimiento Humano mostró que las tres técnicas paramétricas más utilizadas fueron el análisis de varianza (ANOVA) (57.7%), la prueba t-student (34.8%) y la correlación de Pearson (29.5%); mientras que las menos reportadas fueron el análisis de covarianza (ANCOVA) (5.2%), el análisis multivariado de varianza (MANOVA) y la correlación canónica (0.3%). Las técnicas de análisis estadístico no paramétricas fueron poco utilizadas, en las que la prueba de χ^2 fue la más reportada (7.9%), seguida de la prueba de Wilcoxon (4.9%) y la prueba U de Mann-Whitney (3.0%)^[499]. Desde la publicación de ese estudio, otros investigadores han propuesto técnicas de análisis estadísticos más complejas, pero que brindan mucha mayor información y solidez a los estudios que se publican^[348].

Debido al desarrollo de paquetes estadísticos, a su vez impulsados por una mayor capacidad de procesamiento de datos informáticos, se ha podido avanzar vertiginosamente en el desarrollo y perfeccionamiento de conceptos para el análisis de datos provenientes de las diversas áreas de las Ciencias del Movimiento Humano. Hoy en día, se integran múltiples formas de análisis para estudiar los datos con el propósito de llegar a conclusiones sólidas, confiables y reproducibles que sirven para tomar decisiones. En esta sección se describirán algunas de las técnicas de análisis de datos poco conocidas, poco utilizadas y raramente reportadas en las investigaciones en las Ciencias del Movimiento Humano.

2.8.1. Análisis robustos de datos

Existen diferentes categorías de análisis de datos. Por ejemplo, se puede medir alguna variable de interés (i.e., variable dependiente) como la fuerza muscular. Ese dato, que es único, representa una característica de ese atleta, cliente o paciente en un momento determinado del tiempo. Como tal, ese dato proporciona información para algún objetivo particular para quien lo obtiene. Entonces, esa sería la categoría más elemental de análisis, un dato. Sin embargo, y especialmente hoy en día, se cuenta con una cantidad monumental de información de una persona, pues los medios tecnológicos han ido evolucionando para obtener, incluso segundo a segundo, datos de una variable para su posterior análisis, interpretación y toma de decisiones. Esa, por lo tanto, es una categoría de análisis superior y compleja que permite mapear con mayor precisión la realidad de la persona en una variable determinada o en un grupo de variables determinadas. Por lo tanto, las técnicas de análisis de datos convencionales ya no parecen ser suficientes para comprender el fenómeno del análisis de datos que enfrentan los investigadores.

El uso de datos masivos, macrodatos o más comúnmente conocidos como “big data” en diversas áreas del conocimiento ha permitido el avance vertiginoso de la ciencia. El término “Big data” se refiere a conjuntos de datos tan grandes y complejos que requieren de aplicaciones informáticas no tradicionales para su procesamiento para tratarlos adecuadamente. Por ejemplo, debido al confinamiento forzado por la pandemia de Covid-19, se realizó un estudio utilizando la herramienta de big data de Google Trends para conocer las tendencias de búsqueda de información relacionada con la actividad física antes y después del confinamiento, encontrándose un aumento en el interés y el compromiso de la población con la actividad física^[211]. El uso de big data en el área de la salud también contiene cantidades masivas de información de diversas fuentes, incluyendo los expedientes de salud electrónicos, datos administrativos y datos de dispositivos de autocontrol (e.g., monitores de glucosa sanguínea, medidores de presión arterial ambulatoria)^[540]. Así, la disponibilidad de big data ofrece posibilidades para una mejor comprensión de las enfermedades y sus posibles tratamientos; sin embargo, también se requiere que los usuarios comprendan su interacción con las máquinas y programas de computación^[611].

En las Ciencias del Movimiento Humano, en áreas como el rendimiento deportivo en el fútbol, se ha observado que los nuevos dispositivos inerciales (e.g., WiMu, FieldWiz, Catapult, OhCoach, JOHAN V5, McLloyd STv4, PLAYR, Quantrax) son capaces de recolectar miles de datos de cada jugador por segundo, durante los entrenamientos y durante los juegos oficiales^[663]. Esos dispositivos miden la distancia recorrida, las carreras de alta intensidad, saltos, aceleraciones, entre otras. Esto significa, en el fútbol, por ejemplo, que un dispositivo puede grabar 150 variables por segundo durante 90 min para cada uno de los jugadores en el campo y los sustitutos. Los millones de datos recolectados obliga a los profesionales a recurrir a técnicas robustas para el análisis de información (e.g., big data, análisis multivariados, machine learning, árboles de decisión), y es por eso que los profesionales analistas de datos en el deporte deben estar preparados para enfrentar esos retos que ya han llegado para quedarse^[754]. Por ejemplo, con base en análisis robustos de datos, se ha estimado que, para el año 2030, es probable que el fútbol europeo se juegue a velocidades más altas y con períodos más densos de esfuerzos de alta intensidad^[525]. Incluso, la analítica deportiva permitirá que la identificación y reclutamiento de jugadores talentosos sea una herramienta de uso diario en las organizaciones deportivas, aunque se deberán conocer también las necesidades tecnológicas y de recursos humanos para lograr la tarea^[611].

2.8.1.1. Inteligencia artificial

La inteligencia artificial (IA) se puede definir como la capacidad de una computadora o un robot controlado por una computadora para realizar tareas que generalmente realizan los humanos debido a que requieren inteligencia y discernimiento humanos. A la IA también se le podría definir como la capacidad de un sistema para interpretar datos externos correctamente, aprender de dichos datos y utilizar esos aprendizajes para lograr objetivos y tareas específicos a través de una adaptación flexible^[309]. Otra reciente definición indica que “es la ciencia y la ingeniería de la creación de máquinas inteligentes que tienen la capacidad de realizar tareas que de otro modo requieren intervención humana” (p. 1)^[592].

Debido a que la IA es un área en constante evolución, su uso se ha generalizado a todas las facetas de la vida humana, incluyendo la salud y las Ciencias del Movimiento Humano. La IA tiene subáreas de desarrollo, incluida el aprendizaje de máquinas, más conocido en inglés como “*Machine learning*” (ML). La IA utiliza estadísticas descriptivas de datos con alta densidad de información para medir cosas, que luego son agrupadas y analizadas (i.e., análisis de agrupamiento) por medio de algoritmos, que son un conjunto de instrucciones detalladas paso a paso o fórmulas que se utilizan para resolver un problema o completar una tarea^[764]. Es decir, se requiere de potencia informática para tomar datos individuales, agruparlos y ordenarle a la máquina, por medio de instrucciones (i.e., algoritmos), que simplifique la información para el usuario final.

Para una excelente revisión del concepto de IA y su aplicación en la medicina deportiva, se recomienda consultar el artículo de Ramkumar et al.^[592]. Los autores indican que, por ejemplo, la IA en el campo de la cirugía ortopédica y la medicina deportiva ha demostrado ser muy prometedora para predecir el riesgo de lesiones de los atletas, interpretar imágenes y evaluar los resultados de los pacientes^[592]. En el área del entrenamiento deportivo, se han realizado estudios para comprender el rendimiento en el ejercicio anaeróbico cíclico y la actividad neuromuscular utilizando algoritmos genéticos de IA. La idea de este tipo de estudios es proporcionar a los entrenadores un nuevo método factible y más efectivo de entrenamiento diario e investigar si este nuevo método de entrenamiento puede optimizar el ejercicio anaeróbico en humanos^[734]. En el área de la actividad física en niños también se ha utilizado la IA para la detección de la aptitud física de los niños asistida por computadora y la evaluación de la intervención del ejercicio basada en el modelo de IA^[764].

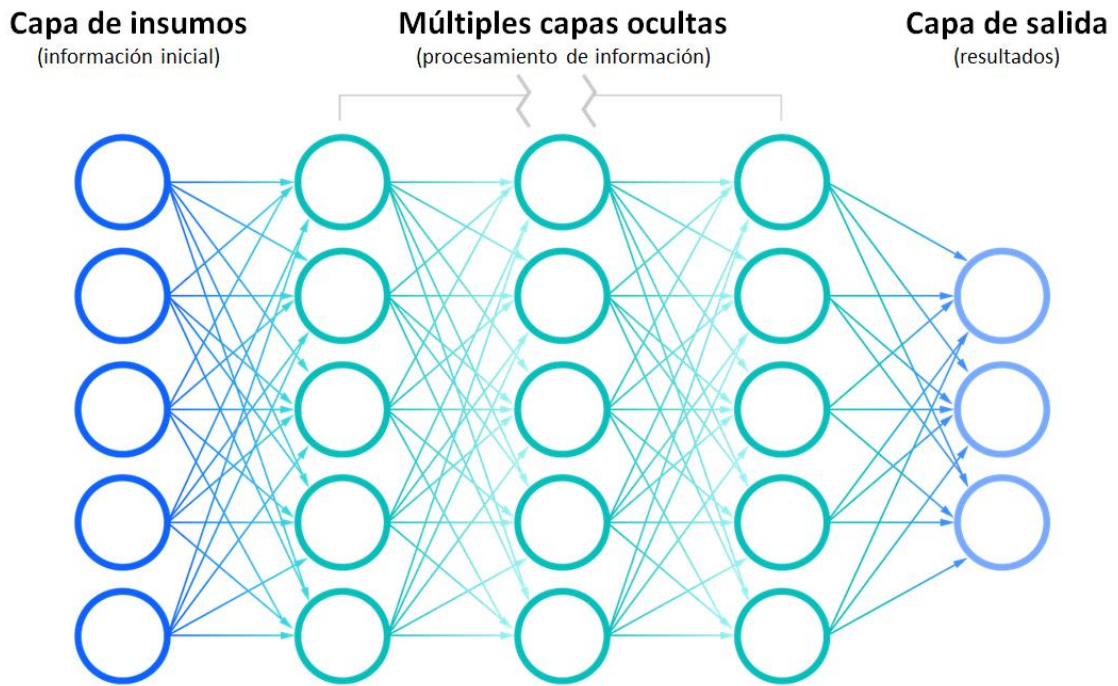
En las Ciencias del Movimiento Humano no se han implementado guías para estandarizar procedimientos al realizar estudios que involucren IA. Por el contrario, en el área de la salud, se han implementado guías que sugieren considerar aspectos fundamentales durante la inclusión de IA en estudios científicos. Por ejemplo, SPIRIT-AI, que es una parte de la iniciativa SPIRIT-AI y CONSORT-AI, representa una iniciativa internacional respaldada por la red SPIRIT y EQUATOR para ampliar la declaración SPIRIT del año 2013, con el propósito de desarrollar una guía de protocolo específica de IA de consenso. Es complementario a la declaración CONSORT-AI, cuyo objetivo es promover informes de alta calidad de estudios donde se utilice IA^[187].

La práctica futura de los especialistas en Ciencias del Movimiento Humano, incluyendo la medicina deportiva, requerirá de un conocimiento práctico fundamental acerca de las fortalezas, limitaciones y aplicaciones de las herramientas basadas en IA para así aprovechar el potencial para automatizar tareas y mejorar los conocimientos basados en datos para optimizar la atención de atletas y pacientes^[592]. Al ser un área inter y transdisciplinaria, los nuevos profesionales en Ciencias del Movimiento Humano deben buscar formación complementaria proveniente de las ciencias de la computación y la informática para poder comprender elementos básicos de la IA y así poder establecer una comunicación eficaz con otros profesionales.

2.8.1.2. Redes neuronales

El vertiginoso avance de hardware y software permiten utilizar herramientas de la computación e informática tradicionalmente reservadas para las ingenierías para aplicarlas a las Ciencias del Movimiento Humano. Una de las aplicaciones de la IA son las redes neuronales (RN), también conocidas como redes neuronales artificiales (ANN, por sus siglas en inglés) o redes neuronales simuladas (SNN, por sus siglas en inglés). Las RN representan un subconjunto del aprendizaje automático y forman parte del núcleo de los algoritmos de aprendizaje profundo ("deep learning"). Su nombre y estructura están inspirados en el cerebro humano, imitando la forma en que las neuronas biológicas se envían señales entre sí (Figura 1).

Figura 1. Esquema conceptual de las redes neuronales. Para que la red neuronal funcione, se requiere brindarle información inicial, y por medio de procesos matemáticos (capas ocultas o nodos), se brindan resultados de clasificación. El proceso completo asemeja el funcionamiento de una red de conexión biológica mediada por neuronas.



Las RN generalmente están compuestas de dos fases; una de "entrenamiento" y otra de clasificación o predicción^[20]. En la primera fase, la red aprende a autoajustar sus pesos o ponderaciones en respuesta a un insumo ("input") y a un producto ("output") proporcionado por el usuario. En la segunda fase, la red aplica el aprendizaje adquirido a los insumos para generar un producto. Para ello, se recomienda utilizar una estrategia por medio de la cual se entrene la RN con ejemplos de patrones conocidos.

El uso de las RN no es completamente novedoso en las Ciencias del Movimiento Humano. De hecho, se han venido utilizando desde inicios y mediados de los años 90's^{[19][20]}. Por ejemplo, por medio de dispositivos de medición de aceleración portátiles, se pudo recolectar datos de marcha y se diseñaron dos RN para reconocer cada patrón y estimar la velocidad y la inclinación de la marcha. Las RN primero fueron "entrenadas" por medio de patrones conocidos de caminata en banda sin fin y luego se estimaron las inclinaciones, las velocidades y la distancia recorrida durante la caminata sobre el suelo realizadas en un circuito al aire libre. Los investigadores encontraron una buena concordancia entre las variables reales y las previstas^[20].

El funcionamiento de las RN puede basarse en números o en imágenes (e.g., fotografías, videos)^{[436][549]}. Recientemente se utilizó una RN gráfica para estudiar cambios físicos ocasionados por el deporte en estudiantes y el impacto en su salud causado por alguna lesión. Los resultados muestran que la tecnología de reconocimiento de imágenes puede analizar con precisión el proceso de lesión de la columna cervical y lesión deportiva en las actividades del aula realizadas por los estudiantes^[436]. En otro estudio se utilizó una RN, la cual fue "entrenada" en 400 cuadros de video de deslizamiento en dos dimensiones de 8 nadadores de élite^[549]. Se encontró que la RN digitalizó los puntos de referencia del cuerpo de los nadadores 233 veces más rápido que un operador manual, con un error de digitalización de ~4-5 mm. También se reportó una alta precisión y confiabilidad, por lo que los investigadores concluyeron que las RN se pueden aplicar para reducir el tiempo de análisis cinemático en los deportes y facilitar una retroalimentación rápida a los atletas^[549].

En el ámbito del entrenamiento físico para la competición, el aprendizaje profundo de las RN brinda información técnica para mejorar el desempeño de los atletas, en donde se ha encontrado que se puede mejorar la eficiencia del entrenamiento en un 20%^[429]. El proceso de análisis requiere de conocimientos de estadística, informática y del uso de lenguajes de programación (e.g., JavaScript, Python, Java, TypeScript, C#, PHP, C++, C). Los lenguajes de programación le permiten al humano comunicarse con la computadora por medio de algoritmos, como por ejemplo el de Optimización de Enjambre de Partículas Adaptativas (APSO, por sus siglas en inglés) la función de base radial (RBF, por sus siglas en inglés), el de Growing y Pruning-RBF (GAP-RBF, por sus siglas en inglés), y el de mínimos cuadrados ortogonales (OLS, por sus siglas en inglés)^[301].

Los algoritmos son instrucciones que permiten estudiar relaciones entre las variables, y se desea que sean lo más precisos posibles al clasificar y ponderar las variables de interés. Con algoritmos como APSO se logró determinar que las variables relevantes para el desempeño físico de 100 atletas de China fueron la habilidad, calidad atlética, capacidad psicológica y expresión artística^[301]. La interacción de estas variables colabora para potenciar el rendimiento deportivo de los atletas, por lo que se deben estudiar modelos autóctonos para poder comprender mejor las variables que determinan el rendimiento deportivo en poblaciones particulares.

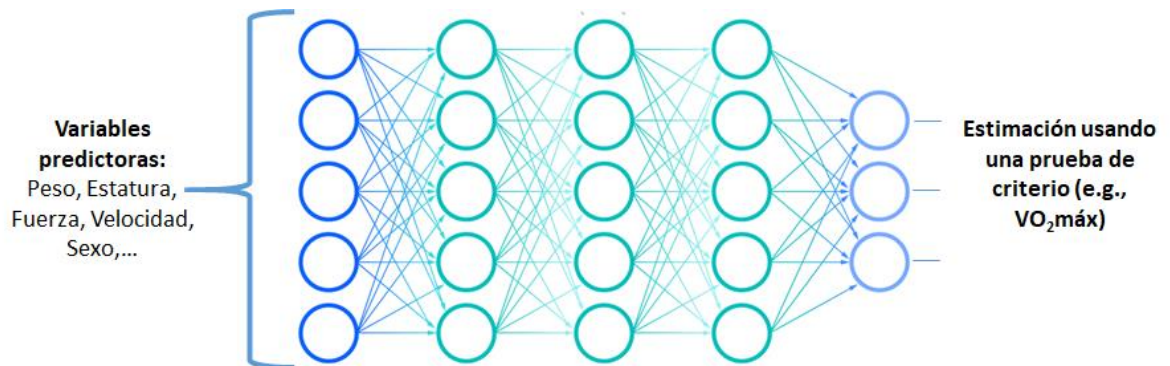
Otra manera de utilizar las RN es cuando se desea validar el juicio de expertos en algún tema, por ejemplo, la determinación del primer y segundo umbral ventilatorio (i.e., $VT_{(1)}$ y $VT_{(2)}$) en las pruebas de laboratorio de fisiología del ejercicio^{[780][781]}. Generalmente, $VT_{(1)}$ y $VT_{(2)}$ se determinan visualmente; para lo cual se debe tener conocimiento y experiencia para calcularlos con precisión. Estos indicadores permiten conocer los límites fisiológicos entre el ejercicio de moderada a alta intensidad con el propósito de asociarlos con alguna variable de interés práctica para la prescripción del ejercicio, como, por ejemplo, la frecuencia cardíaca. Así, cuando se tiene el juicio

de expertos, entonces se podrían utilizar las RN para crear modelos robustos para la predicción automática de $VT_{(1)}$ y $VT_{(2)}$, tal y como lo hizo un grupo de investigadores, quienes usaron clasificadores sencillos^[781] y convolucionales (i.e., operación matemática que transforma dos funciones en una tercera función) de RN^[780] y los aplicaron a una base de entrenamiento de 228 pruebas de esfuerzo y las compararon con el juicio de 7 expertos, quienes juzgaron 25 pruebas de esfuerzo. En el estudio del 2019, se encontró que los algoritmos de las RN permiten detectar los umbrales ventilatorios con una alta precisión, lo cual le permitió al grupo de investigación poder desarrollar una aplicación web para obtenerlos automáticamente^[780]. Sin embargo, estas RN están en evolución y obviamente arrojan resultados confiables y precisos dependiendo de la calidad y cantidad de la información que se les proporcione; por lo que los hallazgos reportados siempre deben interpretarse en el contexto de cada estudio particular y no pueden ser generalizados.

Las RN también se han utilizado como una herramienta para monitorear la información de la carga del entrenamiento en atletas. Así, se han utilizado algoritmos de optimización paralela de RN basados en big data para estimar la carga y la intensidad del movimiento de acuerdo con el modo de movimiento y los datos de aceleración, a fin de realizar el seguimiento en tiempo real del entrenamiento físico^[585]. En el estudio, se obtuvieron 176 muestras de datos de seguimiento de seleccionados de tenis de China. Del total de muestras, se tomaron 160 para "entrenamiento" de la RN y 16 como muestras de prueba. Los investigadores encontraron que el algoritmo de clasificación bayesiano puede evitar eficazmente las lesiones deportivas causadas por el sobreentrenamiento al advertir la fatiga del atleta^[585].

Las RN se utilizan para la predicción en modelos animales y humanos^[773]. Por ejemplo, en Corea, se utilizó una RN para estimar el VO_2 máx basado en una prueba de carrera de ida y vuelta de 10 m de varias etapas en 118 adultos sanos entre 20 y 50 años^[550] (Figura 2).

Figura 2. Ejemplo de modelo de predicción utilizando redes neuronales.



Para construir los modelos de predicción de VO_2 máx se utilizó una RN de retroalimentación totalmente conectada compuesta por tres tipos diferentes de capas: a) de entrada ("input"), b) la oculta, y c) de salida ("output"). La capa de entrada estaba conformada de 14 nodos o variables independientes (e.g., sexo, edad, peso, talla, etc.) y recibió predictores para estimar el VO_2 máx. Los datos de entrada fueron pasados a las capas ocultas, que estaban conformadas de tres capas con 21 nodos. La estimación final de VO_2 máx se hizo con una prueba gradual de esfuerzo y fue incluida en la capa de salida. A ese arreglo se le conoce como un modelo de predicción de múltiples entradas y una sola salida. El "entrenamiento" de la RN se realizó con una tasa de aprendizaje de 0.001 mediante descenso de gradiente estocástico hasta que el número de epochs llegó a 1000. El modelo de RN se implementó utilizando el lenguaje Python utilizando el software Tensorflow y Keras, y los resultados indicaron que el modelo de predicción fue adecuado; sin embargo, los autores concluyen que el modelo necesita de una mayor cantidad de datos para brindar resultados más robustos^[550]. De esta forma puede observarse que los profesionales en Ciencias del Movimiento Humano necesitarían de colaboración de otros especialistas para responder a preguntas de investigación propias del área, o alternativamente, estudiar y capacitarse en programación de RN.

También existen las RN convolucionales, que permiten ser utilizadas, por ejemplo, en el desarrollo de aplicaciones (APP) móviles de entrenamiento físico usando capturas de movimiento basadas en el aprendizaje automático. En un estudio en el que participaron 104 personas se evaluó una APP cuya base fueron las RN y se comparó con 72 participantes que realizaron un programa de ejercicios transmitidos por video^[551]. Las variables de interés fueron la salud, medida por medio del formulario SF-36 y el dolor lumbar. Los hallazgos reportados indicaron que la APP fue mejor que el video de ejercicios para reducir el dolor lumbar y para aumentar la percepción de salud, por lo que los investigadores concluyeron que las APPS basadas en RN podrían ser herramientas prometedoras para lograr mejores resultados médicos y una mayor satisfacción con el tratamiento^[551]. Además, podrían ser fuentes novedosas de ingresos para emprendedores de las Ciencias del Movimiento Humano.

En síntesis, las RN son herramientas informáticas y estadísticas de aplicación directa en las Ciencias del Movimiento Humano. Dentro de poco tiempo será indispensable poseer fundamentos de la programación en lenguajes de computación modernos, así como comunicarse eficazmente con profesionales provenientes de las áreas de la ingeniería y la estadística para resolver problemas relacionados con las especialidades de la actividad física y la salud, así como del rendimiento físico y deportivo. El desarrollo del área tendrá implicaciones en la renovación de planes de estudio universitarios, que necesariamente deberán incorporar temas como los analizados en esta sección.

2.8.1.3. Aprendizaje de máquina

Machine learning (ML), aprendizaje de máquinas o aprendizaje automático, es el subconjunto de la IA más exitoso que usa algoritmos computacionales para analizar grandes conjuntos de datos para clasificar, predecir u obtener inferencias útiles sin instrucciones explícitas^[592]. Se han descrito más de 15 enfoques de ML, los cuales se pueden subdividir en supervisado, no supervisado y reforzado^[465]:

- a. El aprendizaje supervisado constituye la mayor parte de las aplicaciones, y es donde la computadora observa ejemplos de pares de entrada-salida y aprende una función de mapeo que explica el cambio de entrada a salida. El aprendizaje supervisado se puede utilizar para predecir una salida valorada como continua a través de la regresión o una salida valorada como discreta a través de la clasificación. Así, el aprendizaje supervisado abarca un proceso de prueba y error mediante el cual el algoritmo compara predicciones con salidas etiquetadas correctamente en el conjunto de entrenamiento.
- b. El aprendizaje no supervisado depende del programa de computación para encontrar la estructura dentro de los datos, en función de características particulares. Los algoritmos no supervisados buscan patrones de forma autónoma sin requerir salidas etiquetadas correctamente en el conjunto de entrenamiento. El sistema aprende patrones en la entrada, aunque no se proporcione retroalimentación explícita. El tipo más común de aprendizaje no supervisado es el agrupamiento o "clustering".
- c. En el aprendizaje reforzado, el programa de computación debe aprender a comportarse con éxito dentro de un entorno particular, en función de la retroalimentación de éxitos y fracasos.

Algunas de las técnicas de ML incluyen máquinas de vectores de soporte (SVM), redes neuronales de regresión generalizada (GRNN), red de función de base radial (RBFN), el K-Vecinos más cercanos (K-NN), y bosques de árboles de decisión (DTF) (Tabla 1).

ML es un proceso exploratorio iterativo en el que no existe un modelo que se ajuste a todos los problemas o conjuntos de datos. Es decir, no se conoce algún modelo que logre la mayor precisión para un dominio específico, tipo de problema o conjunto de datos. De hecho, existen críticas fundamentadas con respecto al uso de ML ya que se promocionan como eficientes, valiosos y objetivos, ocasionando una confianza ciega en la tecnología (i.e., sesgo de automatización). Es decir, la manera intrínsecamente opaca, compleja e invisible en la que funcionan los sistemas genera cierta incomodidad en los usuarios al dejar que las decisiones las tome una "caja negra"^[465]. Aun así, el modelo de ML con mejor rendimiento varía de un resultado a otro según las características de las variables y la observación^[11]. A pesar de esto, independientemente del modelo de ML utilizado, se estudian métricas comunes como el error estándar de estimación (SEE), el coeficiente de Gini, el coeficiente de correlación múltiple (R), el error cuadrático medio (RMSE), el área bajo la curva característica operativa del receptor (AUC), la sensibilidad y la especificidad para determinar el desempeño de los modelos y poder decidir cuál es el que mejor modelo de ML que se ajusta a los datos analizados^{[161][379]}.

Tabla 1. Características de algunas técnicas de machine learning (ML) utilizadas para la reducción y clasificación de datos.

Técnica de ML	Características
SVM	El algoritmo SVM es una técnica de clasificación supervisada cuyo objetivo es encontrar un hiperplano que separe de la mejor manera posible dos clases diferentes de puntos de datos. El algoritmo SVM estándar está formulado para problemas de clasificación binaria, en donde los problemas multiclase normalmente se pueden reducir a una serie de problemas binarios. Esta técnica de ML pertenece a los llamados métodos kernel.
GRNN	GRNN es una variación de las redes neuronales de base radial y se puede usar para regresión, predicción y clasificación. También se le considera como una técnica mejorada en las redes neuronales basada en la regresión no paramétrica. Utiliza una función de activación gaussiana en la capa oculta y está compuesta de capas de entrada, ocultas, de suma y de división.
RBFN	Las RBFN son un tipo de redes de neuronas artificiales que calculan la salida de la función en función de la distancia a un punto denominado centro. Al igual que con los perceptrones multicapa, sirven como aproximadores universales.
K-NN	El algoritmo K-NN es una técnica de clasificación supervisada en la que el clasificador utiliza el conjunto de datos de entrenamiento para entrenar el modelo sobre cómo se ven los elementos de la clase "Sí" y de la clase "No". La predicción se realiza clasificando el nuevo punto en la base de datos de acuerdo con la clase de los puntos más similares o "más cercanos" en el conjunto de datos. Es un método de aprendizaje no paramétrico que no hace suposiciones acerca de la distribución de los datos.
K-means	K-promedios es un algoritmo iterativo que intenta dividir el conjunto de datos en subgrupos distintos no superpuestos (i.e., clústeres), en donde cada punto de datos pertenece a un solo grupo. Intenta hacer que los puntos de datos dentro del clúster sean lo más similares posible y, al mismo tiempo, mantiene los clústeres lo más diferentes o lejanos posible. Asigna puntos de datos a un grupo de modo que la suma de la distancia al cuadrado entre los puntos de datos y el centroide del grupo (i.e., media aritmética de todos los puntos de datos que pertenecen a ese grupo) sea el mínimo. Cuanta menos variación exista dentro de los conglomerados, más homogéneos o similares serán los puntos de datos dentro del mismo conglomerado.
DTF	DTF es un tipo de aprendizaje automático supervisado que se utiliza para categorizar o hacer predicciones en función de cómo se respondió un conjunto anterior de preguntas. El modelo es una forma de aprendizaje supervisado, lo que significa que el modelo se entrena y prueba en un conjunto de datos que contiene la categorización deseada.

Nota: SVM: vectores de soporte; GRNN: redes neuronales de regresión generalizada; RBFN: red de función de base radial; K-NN: K-Vecinos más cercanos; K-means: K-promedios; DTF: bosques de árboles de decisión.

Las técnicas de ML han encontrado aplicación directa en las Ciencias del Movimiento Humano, en donde diariamente se utilizan para apoyar la toma de decisiones en el deporte profesional, estudiar diferencias en el rendimiento entre hombres y mujeres dentro del mismo deporte^[283] e incluso predecir futuras lesiones en el fútbol^[614]. Estas metodologías inteligentes han mostrado resultados prometedores en los dominios de clasificación y predicción^[120]. Algunas de las aplicaciones incluyen la estrategia, táctica y análisis del partido, la compra de jugadores, valoración de jugadores y gasto del equipo, el estudio del régimen de entrenamiento, la predicción y prevención de lesiones, la gestión y predicción del rendimiento, estudiar el resultado del partido y la predicción de la tabla de posiciones en un torneo, el diseño y programación de torneos, y el cálculo de probabilidades de apuestas deportivas^{[108][210]}.

Algunas de las técnicas de ML se han usado para predecir el rendimiento físico a partir de indicadores antropométricos. Por ejemplo, se deseaba conocer cuál era la contribución de indicadores antropométricos de los jugadores de fútbol para predecir su rendimiento físico^[92]. Para ello, se reclutaron 16 jugadores de una academia profesional de fútbol juvenil y se midieron variables antropométricas (e.g., área muscular del brazo, circunferencia del músculo del brazo, circunferencia suprapatelar derecha e izquierda). El rendimiento físico se evaluó con pruebas de cambio de dirección, sprints, salto vertical y el Yo-Yo test. Los investigadores utilizaron modelos de regresión de árbol de decisión (DTR), regresión de bosque aleatorio (RFR), regresión de árbol adicional (ETR), regresión de AdaBoost (ABR) y aumento de gradiente para regresión (GBR) para predecir cuál de las características antropométricas era de importancia para cada rendimiento físico. Se encontró que las características antropométricas permitieron predecir con precisión el rendimiento físico de los jugadores y que esas características antropométricas de la parte superior e inferior del cuerpo están estrictamente relacionadas con el rendimiento físico aeróbico y de velocidad en el fútbol juvenil de élite^[92].

Las aplicaciones de ML también alcanzan el área de la salud. En algunos casos, la aplicación de ML al estudio de la actividad física no ha generado los resultados deseados^[272], aunque en otras áreas sí lo ha logrado^[629]. Por ejemplo, ya que la aptitud cardiorrespiratoria es un factor determinante de la salud cardiovascular, actualmente se utilizan modelos de ML para predecir la mortalidad en cohortes de individuos que realizaron pruebas de esfuerzo para determinar el $VO_{2m\acute{a}x}$ ^[11]. Un análisis de 98 personas, a las cuales se le describieron variables como sexo, edad, estatura, peso, frecuencia cardíaca máxima, velocidad y tiempo de ejercicio permitió desarrollar 15 modelos diferentes de predicción del $VO_{2m\acute{a}x}$ usando dichas técnicas de ML^[9]. Utilizando otros algoritmos, también se ha logrado identificar adolescentes con un estado futuro desfavorable en una prueba de aptitud cardiorrespiratoria en función de más de 14 características iniciales^[379].

En síntesis, las técnicas de minería de datos y ML han llegado para quedarse. Su comprensión, uso inteligente y conocimiento de las fortalezas y debilidades podrá colaborar con el avance de las Ciencias del Movimiento Humano, al facilitar el análisis de millones de datos que asistirán en la toma de decisiones. Se espera que dentro de muy poco tiempo se pueda contar con mucha mayor claridad acerca de los algoritmos que hacen posible los análisis, de manera que exista mucha mayor transparencia y se eviten sesgos en la interpretación de los resultados. Adicionalmente, se espera que los profesionales de las Ciencias del Movimiento Humano se preocupen por comprender el idioma de las Ciencias de la Computación, la Informática y la Estadística, entre otras, para mejorar la generación de preguntas de investigación pertinentes que solamente pueden ser respondidas interdisciplinariamente.

2.8.2. Meta regresión

Una forma simple de definir la técnica de meta regresión, es mencionar que se trata de una regresión lineal ajustada a los procedimientos de un meta análisis. En otras palabras, consiste en aplicar la técnica de regresión lineal a un meta análisis, como un análisis estadístico complementario^{[418][431]} o como el análisis principal^[405]. Por lo tanto, se necesita revisar brevemente el concepto de regresión lineal, antes de proseguir.

La regresión es un análisis predictivo que se basa en la correlación. Este último análisis examina la relación o asociación estadística entre dos variables, lo que se conoce como una correlación bivariada, que es la más común. También, se puede extender la técnica para valorar la relación entre más variables, lo cual se conoce como una correlación múltiple o multivariada.

El principio de la asociación o correlación entre dos variables es simple: si las variables "A" y "B" están correlacionadas, entonces puede suceder que a mayor magnitud en la variable "A", mayor tienda a ser la magnitud en la variable "B" y viceversa. En este caso, la correlación entre ambas sería positiva, directa o estrictamente creciente. O puede ocurrir que a mayor magnitud en "A", menor tienda a ser la magnitud en "B" y viceversa, siendo en este caso una correlación negativa, inversa o estrictamente decreciente.

Cuando existe correlación entre dos o más variables, si una de ellas cambia por alguna razón, esa modificación se relaciona de algún modo con cambios en la(s) variable(s) con las que tenga asociación. Un ejemplo muy simple es el del porcentaje de masa muscular y su relación con el tiempo sedentario (i.e., el tiempo que se dedica a actividades hipocinéticas), o sea, de muy baja activación neuromuscular. En ese ejemplo se podría observar una tendencia a que las personas con mayor tiempo sedentario semanalmente, también muestren un menor porcentaje de masa muscular, mientras que las personas con menor tiempo sedentario, y, por ende, más activas físicamente, tenderían a mostrar mayor porcentaje de masa muscular. Este sería un ejemplo de dos variables correlacionadas de forma inversa. Claro que no sería correcto concluir que el tiempo sedentario sea la causa de la magnitud de porcentaje de masa muscular, pues existen otros factores que podrían afectar a la musculatura. Es decir, la existencia de una correlación estadística entre dos o más variables, no necesariamente es evidencia de causalidad.

En el caso del análisis de regresión, cuando existe correlación entre dos variables, conociendo los valores de una de ellas, se podría estimar valores de la otra, dada su relación estadística. Esa es la base de la regresión y su finalidad: predecir valores de una variable (\hat{Y}) en función de los valores de una (X_1) o más variables ($X_2, X_3...X_n$) denominadas como predictoras. Por lo tanto, la expresión matemática de una regresión lineal múltiple sería: $\hat{Y} = a + \beta_1 (X_1) + \beta_2 (X_2) + \beta_3 (X_3)... + \beta_n (X_n) \pm \text{error}$, en donde a es un valor constante, β son los coeficientes de regresión para la variable predictor X , y el error representa la precisión de la predicción^[501].

En los inicios de la historia de la técnica meta analítica existía la regresión, y de hecho se empleó para comprender las variables que moderaban los efectos de los estudios meta analizados^[698]. En aquel momento, en la década de los años 70's cuando surgió la técnica del meta análisis como se le conoce contemporáneamente, no se le denominaba como meta regresión. Es decir, al principio estuvo la meta regresión (i.e., regresión) pero luego, fue cayendo en desuso, para ser rescatada entre finales del s. XX y principios del s. XXI, sin llegar a ser altamente popular,

pese a su importancia. Para una mayor profundización en el tema de la regresión y meta regresión, se recomienda consultar autores que ofrecen definiciones más amplias sobre la regresión lineal^[694] o con respecto a la meta regresión^{[246][445]} y con más ejemplos aplicados a las Ciencias del Movimiento Humano.

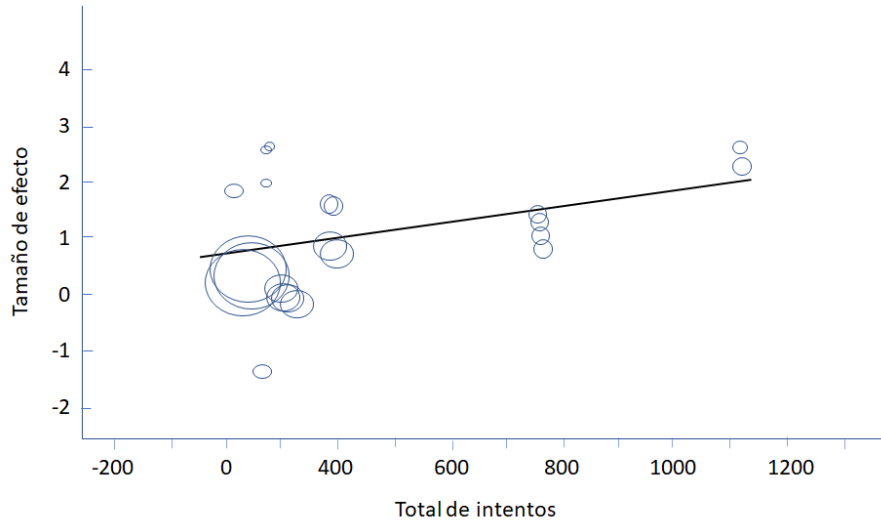
Varios autores han desarrollado artículos sobre la metodología de la meta regresión^{[469][695]}, los cuales son muy útiles para profundizar en este interesante método y conocer la evolución que sigue teniendo^[714]. El principal propósito de aplicar una meta regresión es para comprender mejor y así poder explicar, la heterogeneidad (i.e., variabilidad) que puede presentarse en los hallazgos de un meta análisis^[50]. ¿A qué nos referimos con ese concepto de heterogeneidad en un meta análisis? Brevemente, se puede decir que en un meta análisis se reúnen resultados de múltiples estudios; cada uno de éstos con sus características particulares (e.g., ciertas diferencias entre los participantes de cada estudio, algunas diferencias en el ambiente en que se aplicó la investigación, entre otras). Así, los estudios que se incluyen en un meta análisis no son idénticos, aunque deben tener ciertas similitudes para que se les pueda meta analizar; como por ejemplo, medir la misma variable dependiente (aunque podría medirse con distintos métodos o instrumentos) y aplicar un tipo similar de tratamiento.

Entonces, las diferencias que tienen los estudios, se van a reflejar de algún modo en sus resultados. En un meta análisis se procede a calcular los tamaños de efecto (TE) para convertir los resultados de los distintos estudios en un estadístico o métrica común o estandarizada. Luego, los TE de cada estudio se combinan para obtener un TE global que sería una especie de promedio ponderado de todos los estudios. Por consiguiente, ese TE global podría verse afectado por grandes diferencias entre los estudios a partir de los cuales fue calculado. En esto consiste la heterogeneidad, en el grado de diferencia que puede existir entre los TE individuales; es decir, de cada estudio. La heterogeneidad tiene que estudiarse y reportarse, y se detecta mediante pruebas estadísticas como la Q de Cochran. Adicionalmente, la magnitud de la heterogeneidad se cuantifica mediante estadísticos como el I^2 ^[392].

El problema puede ser importante cuando la heterogeneidad es de una magnitud elevada, por lo que en esos casos se debe proceder a explicar las razones de dicha heterogeneidad. Esto implica hacer análisis de posibles variables moderadoras (i.e., las características que podrían ser responsables de que los estudios tengan diferencias importantes entre sí). La meta regresión cumple con el objetivo de determinar la relación entre una o más variables moderadoras que pueden expresarse en un nivel de medición métrico (i.e., mediante números reales) o no métrico (e.g., ordinales). También es posible transformar variables categóricas (e.g., grupos) en ordinales o incluso en métricas, para introducirlas en la meta regresión.

En la figura 1 se ilustra un ejemplo de una meta regresión de la cantidad de intentos realizados para poder adquirir una destreza motora. Esta sería la variable moderadora. Nótese en la figura cómo se relaciona la variable moderadora con los TE. En la figura 1, cada círculo es un par ordenado compuesto por el TE individual y su correspondiente número de intentos y el tamaño del círculo indica el peso o la importancia que se le brinda al estudio.

Figura 1. Visualización gráfica de una meta regresión en la que se muestra la cantidad de intentos realizados para adquirir una destreza motora.



En la tabla 1 se muestra un ejemplo de resultados de una meta regresión que cuenta con dos variables moderadoras: a) la cantidad de intentos para adquirir una destreza motora, y b) la edad de los participantes de los estudios; pero cada una de ellas se examina por separado.

Tabla 1. Ejemplo de estadísticas de una meta regresión con dos variables moderadoras.

Modelos	Fase de adquisición			Fase de retención		
	β	IC95%	$p \leq$	β	IC95%	$p \leq$
Edad			0.49			0.68
Intercepto	1.07	0.54, 1.59	0.01	-0.37	-1.05, 0.32	0.29
Edad promedio	-0.006	-0.02, 0.01	0.49	-0.005	-0.03, 0.02	0.68
Cantidad total			0.01			0.63
Intercepto	0.65	0.34, 0.96	0.01	-1.23	-2.06, -0.41	0.003
Total de intentos	0.001	0.01, 0.02	0.014	0.01	-0.01, 0.01	0.63

En la tabla 2, se muestra un ejemplo de resultados de una meta regresión en la cual se examinan distintas variables moderadoras en un mismo modelo. Así, la meta regresión puede llevarse a cabo como una regresión simple; es decir, con una variable moderadora a la vez, o múltiple, con dos o más moderadoras en un mismo modelo. En el primer caso, solo habría una variable predictora en el modelo, mientras que en el segundo se tendrían dos o más variables predictoras.

Tabla 2. Ejemplo de tabla resumen de una meta regresión con múltiples variables moderadoras.

Modelos	Competencia motriz general		
	Tamaño de efecto – z de Fisher		
	Coefficiente β	IC95%	p =
Modelos de efectos aleatorios conglomerados			
Edad central promedio	0.011	-0.01, 0.03	0.141
Sexo: sólo hombres (Ref.)	0		
Solamente mujeres	-0.092	-0.25, 0.07	0.217
Hombres y mujeres	-0.152	-0.28, -0.02	0.026
Desarrollo: típico (Ref.)	0		
Atípico	0.125	-0.01, 0.26	0.067
Típico y atípico	-0.025	-0.30, 0.25	0.594
Nivel de alineamiento: Optimo (Ref.)	0		
Mediocre	0.132	-0.07, 0.33	0.169
Intercepto	0.247	-0.01, 0.51	0.059

En el ejemplo, los coeficientes β representan la magnitud y la dirección de la asociación entre cada variable moderadora y el tamaño del efecto calculado. Además, se proporcionan los intervalos de confianza al 95% que acompañan los coeficientes β , de manera que se brinda una clara idea acerca de la precisión del estimador. Finalmente, se presenta la significancia estadística del hallazgo.

La interpretación de la tabla 2 es sencilla; por ejemplo, tomemos la *edad*. Su coeficiente β es de signo positivo, indicando una relación directa o positiva con la variable dependiente, que en este caso sería la competencia motriz general. Esto significa que a mayor edad, mayor es la magnitud del TE de la competencia motriz. Pero si se observan los IC95%, uno de ellos es negativo y el otro positivo. Por lo tanto, de negativo a positivo se debe pasar por cero y cuando eso sucede, como en este caso, se concluiría que la variable (i.e., la edad) tendría un β igual a cero y por lo tanto esta variable no tendría efecto moderador, con un 95% de confianza.

Ahora, para la categoría de la variable *sexo*, se tienen dos subcategorías de *hombres* y *mujeres*. El coeficiente β es de signo negativo, indicando una relación inversa con el TE de competencia motriz. Además, los dos intervalos de confianza tienen el mismo signo, negativo, en este caso. Por lo tanto, el valor cero no estaría contenido entre ambos intervalos. Entonces, la categoría de *sexo: hombres y mujeres*, sí tendría un efecto moderador. ¿Y eso qué significa? Esta variable se refiere a grupos conformados por una mezcla de hombres y mujeres. Como es una variable categórica, para incluirla en la meta regresión, se tuvo que convertir en variable métrica ¿Cómo se hizo eso? Se procedió a transformarla en una variable binaria, es decir con dos valores: 0 y 1. Entonces se asigna 1 a los grupos que combinaban *hombres y mujeres* y 0 a los demás grupos. Así, se interpretaría el efecto moderador de esta variable de esta forma: cuando los grupos son conformados por hombres y mujeres, el TE de competencia motriz general tiende a disminuir en su magnitud.

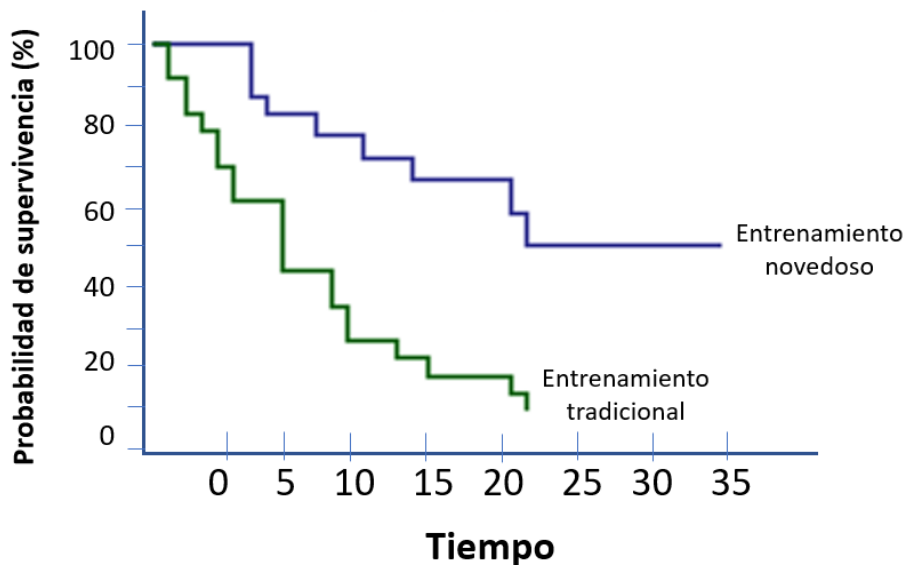
En las Ciencias del Movimiento Humano se utilizarán más meta regresiones en los análisis de variables moderadoras de los meta análisis, ya que existen programas estadísticos de acceso gratuito que lo permiten, como por ejemplo el *OpenMEE*. La interpretación es muy sencilla; sin embargo, construir el modelo requiere de conocimiento teórico de las variables que se desean incluir como potencialmente moderadoras de los efectos, por lo que se deben seleccionar cuidadosamente.

2.8.3. Análisis de supervivencia

En las Ciencias del Movimiento Humano, existen ocasiones en las que interesa conocer la distribución de respuestas en un período entre dos eventos; como por ejemplo, los efectos crónicos del ejercicio, en el que se estudia el tiempo transcurrido entre la aplicación del entrenamiento, su suspensión y la consecuente pérdida de fuerza muscular. Sin embargo, este tipo de datos incluye generalmente algunos casos censurados; es decir, casos de personas para los que no se registra el segundo evento (e.g., las personas que todavía no han perdido fuerza muscular al final del estudio). Para analizar datos de ese tipo, se utiliza un procedimiento de análisis de supervivencia llamado Kaplan-Meier, nombrado así por sus inventores, Edward L. Kaplan y Paul Meier^[386]. Es un método no paramétrico de estimación de modelos hasta el evento en presencia de casos censurados. Así, se estiman las condiciones probables en cada punto temporal cuando tiene lugar un evento y toma el límite del producto de esas probabilidades para estimar la tasa de supervivencia en cada punto temporal.

Por ejemplo, se podría estudiar si un nuevo tipo de entrenamiento físico posee algún beneficio clínico sobre la muerte por infarto al miocardio. Entonces se podría proponer un programa de ejercicios utilizando dos grupos de personas con antecedentes de enfermedad cardíaca. Por una parte, un grupo recibe un programa de ejercicios "tradicional", y por otra, un grupo recibe el programa de ejercicios experimental o "novedoso". Al construir un modelo de Kaplan-Meier a partir de los datos, se pueden comparar las tasas de supervivencia global entre los dos grupos, para determinar si el tratamiento experimental o novedoso representa una mejora con respecto a la terapia tradicional (Figura 1). Kaplan-Meier generalmente utiliza un método de cálculo de tablas de mortalidad que estiman la función de riesgo o supervivencia para el tiempo en que tiene lugar cada evento.

Figura 1. Curva de Kaplan-Meier de supervivencia. En un ejemplo hipotético, se estudia la probabilidad de supervivencia en un tiempo determinado (e.g., años) en personas que completan un determinado tipo de entrenamiento físico.



En las Ciencias del Movimiento Humano se han publicado estudios epidemiológicos de ese tipo. Por ejemplo, se realizó un estudio cuyo objetivo era cuantificar la longevidad de los atletas olímpicos masculinos y femeninos de EE.UU. y los años de vida perdidos o ganados debido a múltiples causas de muerte en comparación con la población general de ese país^[30]. Para ello, se analizaron exatletas estadounidenses que participaron en los Juegos Olímpicos de verano o invierno al menos una vez entre 1912 y 2012, de los que se anotaron la fecha de nacimiento, la muerte y las causas. De los 8124 atletas se observaron 2309 muertes (Mujeres = 225 vs. Hombres = 2084). Los atletas ganaron 2.2 años para adquirir enfermedades cardiovasculares y 1.5 años para cáncer en comparación con la población general. Las tasas de mortalidad debidas a trastornos del sistema nervioso (e.g., Alzheimer, Parkinson) y a deterioro cognitivo (e.g., demencia, esquizofrenia) no fueron diferentes a las de la población general.

Los análisis de supervivencia se pueden utilizar administrativamente para conocer la tasa de deserción de clientes o para conocer la adherencia a un régimen de entrenamiento como la natación^{[516][658]}. Por ejemplo, se investigó la participación a largo plazo y las tasas de abandono en la natación juvenil australiana, para lo cual se realizó un estudio de cohorte retrospectivo con datos de rendimiento deportivo^[516]. Se utilizaron análisis de supervivencia de Kaplan-Meier junto con la regresión de Cox para determinar las tasas de abandono y sus factores influyentes, y se encontraron tasas medianas de participación sostenida de 4 años (IC95% = 3.93, 4.06 años) y un 15.9% mantuvo la participación durante 10 años. Las regresiones de Cox identificaron que el grupo de edad se asoció con la deserción, con una tasa de riesgo 184.9% más alta para los niños de 10 años que para los de 15 años. También, la proximidad residencial a las principales ciudades se asoció con la deserción y los nadadores urbanos mostraron una tasa de deserción 24.8% más alta que los nadadores de zonas rurales. El sexo y la edad relativa no se asociaron con la deserción. Los investigadores concluyeron que el grupo de edad, el nivel de competencia y la proximidad de la ciudad se asociaron con mayores tasas de abandono de la natación.

Es muy posible que los análisis de supervivencia se puedan aplicar a nuevos contextos dentro de las Ciencias del Movimiento Humano. Los programas informáticos y la facilidad para la graficación serán herramientas indispensables para utilizar esta técnica, que además es complementada por diversos tipos de análisis de regresión.

2.8.4. Alometría

En las Ciencias del Movimiento Humano se miden variables de todo tipo (e.g., fisiológicas, de aptitud física, motriz, psicológicas, biomecánicas, etc.) y a menudo esos valores se comparan con datos provenientes de normas para emitir un juicio acerca de las características de la persona. Sin embargo, las normas son difíciles de crear porque requieren de un tamaño de muestra grande y representativo de la población, lo cual pocas veces es alcanzable.

Por otra parte, las normas con referencia de criterio se obtienen generalmente utilizando criterios subjetivos que varían entre investigadores. Estos problemas pueden generar que la clasificación o el juicio emitido acerca de una persona esté sesgado y sea injusto. Incluso, cuando se utilizan percentiles en pruebas similares, las diferencias pueden ser de 2 o 3 min en pruebas de resistencia cardiovascular^[504], lo cual crea un reto cuando se pretende comparar pruebas que miden el mismo constructo.

Debido a estos inconvenientes, se ha sugerido usar el escalamiento alométrico. La alometría se refiere a los cambios de dimensión relativa de las partes corporales que ocurren en la infancia y en la adultez^[530]. Por ejemplo, el esqueleto se vuelve mucho más fuerte y robusto en relación con el tamaño del cuerpo a medida que aumenta el tamaño del cuerpo. El enfoque alométrico provee una metodología para interpretar las diferencias en el desempeño de niños y adolescentes que están asociadas con cambios en el tamaño de su cuerpo ocasionados por el crecimiento y el desarrollo.

La alometría expresa matemáticamente el grado en que una variable (e.g., fisiológica, anatómica o temporal) se relaciona con una unidad de tamaño corporal, a medida que aumenta el tamaño^[529]. Por lo tanto, el modelado alométrico es una forma particularmente relevante para resolver este problema dada su base teórica apoyada biológicamente y por su versatilidad matemática^[220].

En términos simples, se puede encontrar que personas de la misma edad y sexo que ejecutan la misma prueba (e.g., correr 1609 m) pueden variar en los tiempos alcanzados solamente porque su tamaño es distinto. En este caso, el tamaño es una fuente de error que podría afectar el juicio de valor que emitamos sobre esa persona. Por ello, el escalamiento alométrico o análisis dimensional asume que la relación verdadera entre el resultado de una prueba y una variable antropométrica es curvilínea, pasa por el origen, y alcanza la forma: $y = bx^a$; en donde "y" es el resultado en la prueba, "x" es la variable a corregir (escala) o la variable antropométrica, "b" es una constante multiplicativa, y "a" es el exponente^[718].

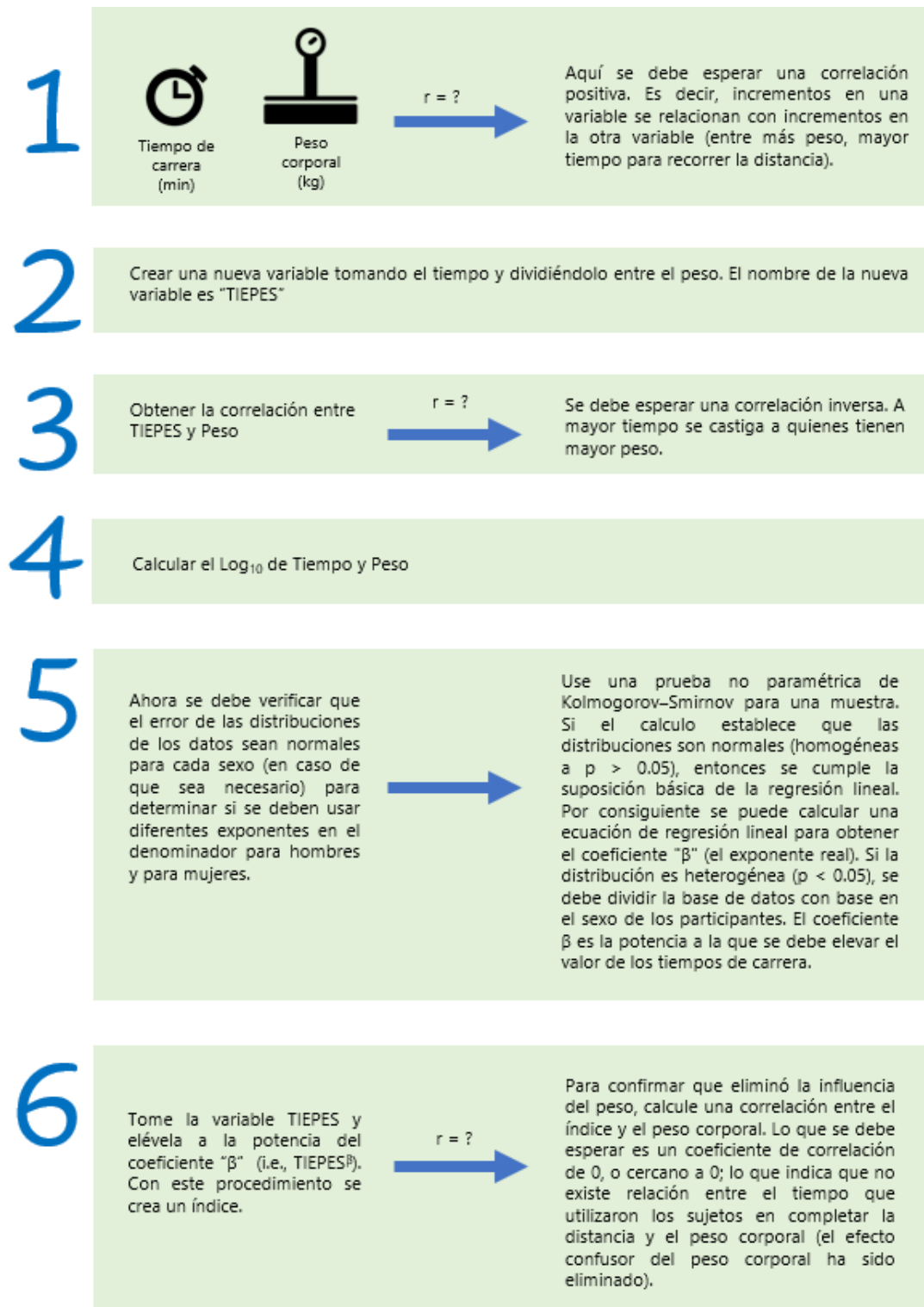
Un ejemplo clásico en donde se debería utilizar el escalamiento alométrico es con los valores de $VO_2\text{máx}$ obtenidos por una persona. Para efectos de comparación entre personas, frecuentemente el $VO_2\text{máx}$ se expresa por unidad de masa corporal en $\text{ml O}_2\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$; pero los individuos con mayor masa corporal tienden a ser penalizados debido a la desproporcionada variación en la relación entre el $VO_2\text{máx}$ y la masa corporal. Así, un individuo con mayor peso no necesariamente refleja a una persona con mayor cantidad de grasa corporal o menor desarrollo muscular, por lo que se ha sugerido que el $VO_2\text{máx}$ se debería expresar en $\text{ml O}_2\cdot\text{kg}^{-0.7}\cdot\text{min}^{-1}$, de manera que se pueda considerar y corregir el efecto de la masa corporal y así finalmente poder comparar diferentes individuos^[71].

En un estudio en el que se compararon niños de EE.UU. y de Costa Rica, quienes corrieron la prueba de 1609 m (1 milla) y a quienes se les midió el peso corporal, se encontró que si se deseaba corregir el efecto del peso (i.e., masa corporal) sobre el tiempo de carrera alcanzado, los participantes de la muestra de Costa Rica debían dividir el tiempo de carrera (T) entre la masa corporal (MC) elevada al exponente $a = 0.438$ (i.e., $T \cdot MC^{0.438}$), mientras que los niños de EE.UU. debían elevarla al exponente $a = 0.220$ (i.e., $T \cdot MC^{0.220}$)^[504].

En otros estudios se han utilizado exponentes para la fuerza máxima en la prueba de cuarto de sentadilla ($N \cdot kg^{-0.61}$) en el gasto energético al estimar la economía de carrera ($kJ \cdot kg^{-0.75} \cdot km^{-1}$) y en el VO_2 máx ($ml \cdot kg^{-0.67} \cdot min^{-1}$)^{[80][81][82][83]}.

Los pasos básicos para calcular un índice alométrico se realizan por medio de correlaciones bivariadas, aunque se pueden utilizar análisis de covarianza (ANCOVA) y otras técnicas (Figura 1).

Figura 1. Ejemplo de pasos para obtener un índice alométrico a partir de los datos de una prueba de carrera y el peso de los participantes.



2.8.5. Análisis composicional y de sustitución isotemporal

Tanto el análisis composicional como el modelado de sustitución isotemporal son enfoques novedosos y complementarios en el área de las Ciencias del Movimiento Humano. Gran parte de su relevancia radica en que permiten cuantificar la codependencia entre variables relacionadas (e.g., conductas saludables) y las implicaciones de modificar alguna de estas variables en función de las demás, sobre distintos efectos de interés (e.g., indicadores de riesgo cardiovascular y mortalidad)^[666]. Teóricamente, el abordaje del análisis composicional permite el manejo de datos que representan o describen las partes de un todo y que, como tal, se reportan comúnmente como vectores de proporciones, concentraciones, porcentajes, o frecuencias^{[275][553]}. A su vez, el análisis de sustitución isotemporal se basa en análisis de regresión lineal y permite estimar, por ejemplo, el efecto de sustituir un tipo de actividad por otra de la misma duración^{[235][236][487][616]}, de ahí que también se le denomine modelado de redistribución isotemporal^[74].

La utilización de ambos enfoques se origina en otras disciplinas, cuya naturaleza de los datos implica interrelaciones y proporciones multivariadas. Por ejemplo, los modelos de sustitución isotemporal se originaron en epidemiología nutricional para comprender el riesgo de enfermedad coronaria al sustituir ciertos componentes de alimentos en la dieta (e.g., grasa saturada por grasa poliinsaturada) mientras se mantiene constante la ingesta calórica total^{[487][751]}. De la misma manera, desde la referente publicación de Aitchison^[8], el análisis de datos composicionales se aplica en campos tan diversos como la nutrición, las ciencias políticas (e.g., datos electorales de múltiples partidos) o la geoquímica (e.g., datos sobre la cantidad de sedimentos y compuestos químicos en rocas y aguas subterráneas)^[158], hasta que fue adaptado para su utilización en otras áreas del conocimiento como las Ciencias del Movimiento Humano.

Así, ambos tipos de análisis han sido considerados como métodos analíticos óptimos para estudiar nuevos paradigmas de actividad física, tal como el ciclo de actividad de 24 h (24-HAC, por sus siglas en inglés), cuyo interés ha crecido exponencialmente en los últimos años^{[74][616]}. En síntesis, el 24-HAC se basa en la premisa de que el tiempo diario disponible (i.e., 24 h) es finito y constante, por lo que el incremento en el tiempo dedicado a una actividad inevitablemente disminuye o desplaza el tiempo dedicado a otra actividad, lo cual implica una co-existencia entre comportamientos. Ambos enfoques han sido también acogidos para analizar simultáneamente la actividad específica que se realiza y la actividad que se desplaza mientras se iguala el tiempo intercambiado entre actividades y se mantiene el tiempo total constante^{[74][298][485][486][657]} o para determinar efectos combinados del tiempo que se dedica a distintas conductas saludables^[158]. De tal manera que, permiten examinar si el sustituir conductas sedentarias por alternativas más saludables, como dormir o realizar actividad física de intensidad moderada a vigorosa, influye en el valor de múltiples variables de interés, por ejemplo, ciertos indicadores de riesgo de enfermedad cardiovascular, como la sensibilidad a la insulina o el perfil lipídico^[119], o sobre indicadores relacionados con la cognición y el desarrollo socioemocional en la niñez, como los son las funciones ejecutivas^[74]; lo que representa un interés sustancial para la salud pública.

En este tipo de modelado estadístico a partir de los datos observados, las inferencias resultantes de los cambios o asociaciones estimadas son de carácter teórico. Es decir, se puede cuantificar cuál podría ser la influencia de la redistribución del tiempo dedicado a ciertas actividades sobre las variables de interés y se brindan sugerencias con base en estas predicciones, por lo que, la muestra inicial del estudio no necesariamente modificó sus comportamientos como

parte del estudio. Por ejemplo, el estudio de Bezerra et al.^[74] observó que se predecía un incremento significativo en las funciones ejecutivas (e.g., memoria de trabajo, control inhibitorio) cuando 5, 10, 15, ó 20 min de actividad física de intensidad leve se reemplazaban por actividad física de intensidad moderada a vigorosa, mientras que la sustitución de 15 ó 20 min de sueño por tiempo sedentario se asoció con una reducción significativa en los resultados obtenidos en las pruebas de función ejecutiva en niños de edad preescolar. Esto a partir de datos obtenidos por medio de acelerómetros, cuestionarios, entrevistas a padres de familia y pruebas de función ejecutiva, con base en un diseño transversal. De manera que, las conclusiones a partir de estudios que implementen análisis de datos composicionales o de sustitución isotemporal podrían eventualmente informar futuras intervenciones que efectivamente manipulen o modifiquen las variables de interés en poblaciones específicas, lo que llevaría a que se puedan establecer relaciones causales entre las variables.

Cabe destacar que, el estudio de Bezerra et al.^[74] detalla en su metodología que el análisis composicional se utilizó específicamente para determinar la asociación entre la composición del ciclo de movimiento de 24 h (i.e., actividades físicas de intensidad leve, moderada y vigorosa, conducta sedentaria y tiempo de sueño) y las funciones ejecutivas (i.e., memoria de trabajo visoespacial y fonológica, control inhibitorio, desarrollo de lenguaje y cambio y desplazamiento); mientras que el análisis de sustitución isotemporal se utilizó específicamente para predecir posibles cambios en las funciones ejecutivas, cuando se redistribuía el tiempo dedicado a las distintas actividades mencionadas. Por lo que se evidencia el potencial de aplicación complementaria de ambos análisis.

Tales paradigmas estadísticos también pueden ser utilizados para analizar conductas saludables y de actividad física a partir de bases de datos ya existentes y provenientes de estudios epidemiológicos a gran escala, realizados a nivel nacional e internacional, tal como se reporta en los estudios de Buman et al.^[119] y Mekary et al.^[486], basados en una muestra de 2185 adultos participantes de la *US National Health and Nutrition Examination Survey* y 32900 mujeres participantes del *Nurses' Health Study*, respectivamente. Asimismo, podrían analizarse submuestras de estas bases de datos, dependiendo de la población de interés, como el estudio de Hamer et al.^[316] delimitado a personas adultas mayores participantes de *Whitehall II Epidemiological Cohort*.

Otros ejemplos de estudios que podrían implementar este tipo de análisis serían aquellos en lo que se compara si las asociaciones entre la actividad física o la conducta sedentaria con múltiples variables físicas, psicosociales y cognitivas varían si la participación en estas actividades se mide por métodos subjetivos u objetivos. De esta manera, se podría también aprovechar información proveniente de dispositivos de acceso generalizado que registran/rastrear conductas relacionados con la salud (e.g., relojes inteligentes, teléfonos inteligentes, *Apps*), los cuales típicamente brindan datos como la cantidad de pasos diarios, minutos de actividad física semanal, horas de sueño, nivel de estrés, entre otros.

A diferencia del análisis composicional, los modelos de sustitución isotemporal asumen que las asociaciones entre las variables son lineales, por lo que las mismas no se transforman en variables compuestas^[616]. Recientemente, se debate si el modelado de datos utilizando sustitución isotemporal es, en efecto, el "estándar de oro", para los estudios de investigación epidemiológica de la actividad física^[485]. La discusión radica en que el análisis de sustitución isotemporal considera los datos de uso del tiempo como valores absolutos, mientras que otros enfoques, como el enfoque de análisis composicional, los considera como valores relativos^{[75][76][485]}.

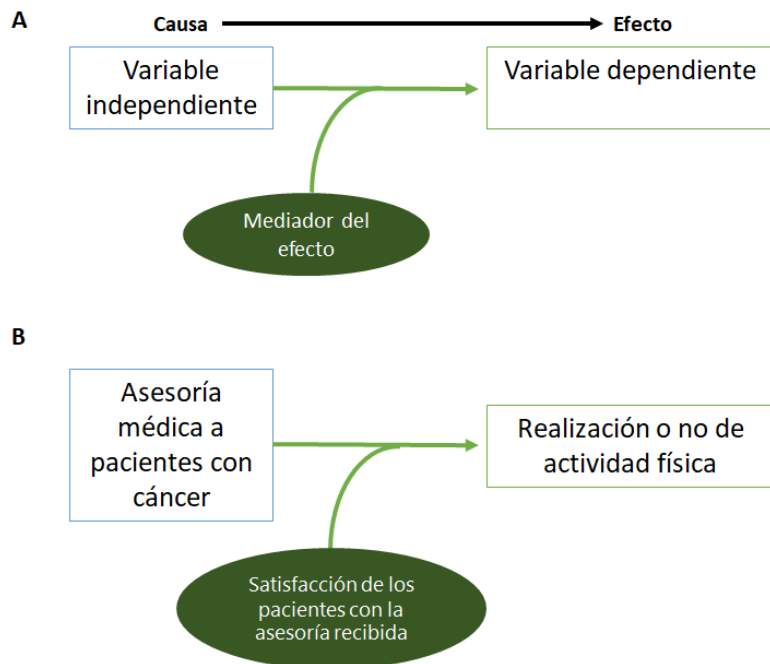
La literatura consultada también sugiere que aún está pendiente de explorar la aplicación de análisis composicional y de sustitución isotemporal en variables orientadas al desarrollo motor, la composición corporal y el rendimiento deportivo, o que consideren más componentes del espectro del comportamiento sedentario, por ejemplo, cuando se está en posición sentada o reclinada mientras se trabaja, se socializa, o se ingieren alimentos. Además, si bien los estudios de Mekary et al.^[486] y Tully et al.^[705] que utilizan análisis de sustitución isotemporal resultan novedosos, el primero al utilizar datos prospectivos explorando los efectos de la conducta de actividad física y el riesgo de depresión como una variable dicotómica, y el segundo al enfocarse en la salud mental de personas adultas mayores; en general, la aplicación de estas metodologías en el estudio de las Ciencias del Movimiento Humano y diversas variables psicosociales resulta también escasa.

2.8.6. Análisis de mediación

Cuando se planifica una investigación, es materialmente imposible controlar todas las posibles variables que pueden afectar a los participantes, y consecuentemente, los resultados provenientes de los datos recolectados. Por lo tanto, se tienen explicaciones parciales que se fundamentan principalmente en la variable independiente manipulada o controlada y en la calidad de los instrumentos con los cuales se midió la variable dependiente de interés. Sin embargo, en la literatura también se describen algunos términos que permiten hipotetizar la influencia directa o indirecta de factores sobre una variable dependiente. A estos factores se les llama variables mediadoras o moderadoras.

Por ejemplo, se realizó un estudio con el propósito de examinar la asociación entre el asesoramiento que brindan los médicos para que los pacientes con cáncer realicen ejercicio físico, y en el que se intentó conocer si esa asociación estaba “mediada” por la satisfacción de los pacientes con esa asesoría y si difería de acuerdo a su diagnóstico inicial de actividad física^[205] (Figura 1). En el estudio se encontró que la “mediación”, a través de la satisfacción percibida, fue significativa entre las personas previamente activas cuando se analizó la probabilidad de realizar cualquier actividad física de moderada a vigorosa y solo entre el grupo con el mayor diagnóstico inicial de actividad física de moderada a vigorosa al observar los minutos de actividad física de moderada a vigorosa posteriores al diagnóstico^[205].

Figura 1. Esquema para comprender el papel de variables influyentes sobre una variable dependiente. El modelo conceptual establece que el efecto de una variable independiente sobre una variable dependiente puede estar mediado por otra variable (A). La satisfacción de un grupo de participantes podría ser un mediador del efecto de la asesoría médica en pacientes con cáncer para que realicen actividad física (B).



Tratar de explicar un fenómeno complejo a través de una sola variable es algo común en la investigación en las Ciencias del Movimiento Humano. Esto se logra algunas veces porque los recursos son limitados o los diseños para realizar los estudios son básicos. Sin embargo, sería preferible tratar de explicar los fenómenos utilizando más de una variable, es decir, con múltiples variables. Es por ello que los diseños multivariados proporcionan mucha mayor información que los modelos univariados, aunque también es más complejo llevarlos a cabo.

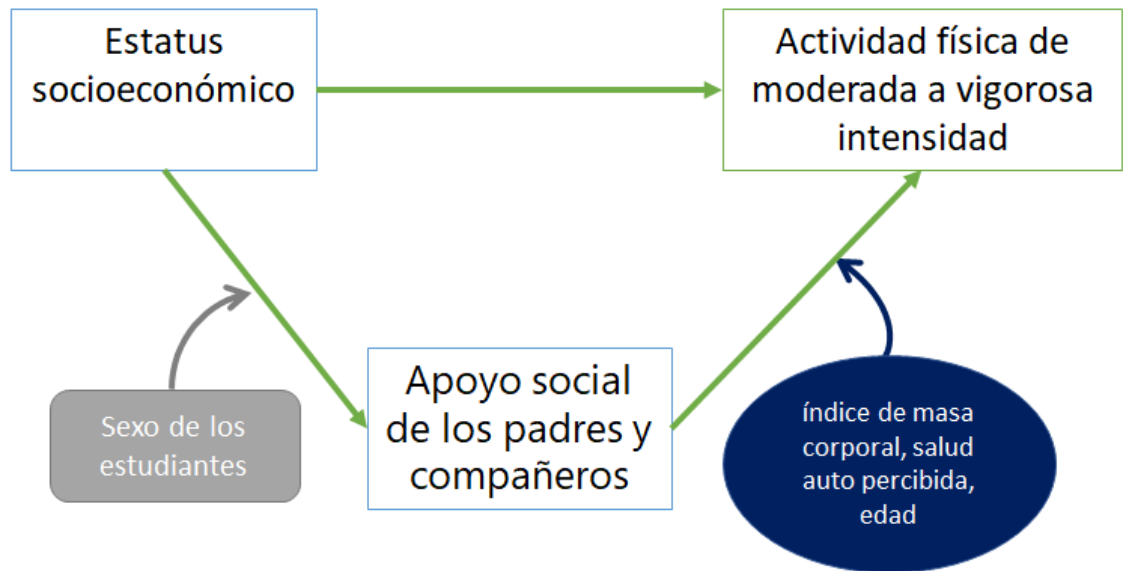
Cuando se pueden medir múltiples variables, se puede conocer la forma en que éstas interactúan, y desde un punto de vista de la investigación, esto permite comprender mucho mejor el fenómeno de estudio. Por ejemplo, adquirir y mantener el hábito de la realización de actividad física es un fenómeno muy complejo, pues esta conducta está sujeta a la interacción, en mayor o en menor grado, de múltiples variables. Comprender si la actividad física, la aptitud física y las habilidades motoras se asocian indirectamente con el rendimiento matemático también podría ser un escenario en donde se requiera analizar múltiples variables^[680]. De manera similar, cuando se desea examinar las asociaciones entre el clima de implicación en la tarea creado por el entrenador y las intenciones de los atletas de continuar practicando deporte, será necesario aplicar varios instrumentos a los atletas y analizar los datos a través de técnicas específicas, como por ejemplo, un análisis de mediación en serie^[690].

La realización de actividad física o mantenerse físicamente activo, es un comportamiento que está explicado por la edad, el sexo, el ingreso socioeconómico, lesiones o discapacidades, escolaridad, cantidad de miembros en el núcleo familiar, experiencias previas y el apoyo social, entre otras variables^[284]. Es decir, algunas de esas variables son potenciales moderadoras del efecto de alguna intervención para incrementar la actividad física de las personas. Por ejemplo, se realizó un estudio para evaluar si la relación entre el nivel socioeconómico y la actividad física de moderada a vigorosa intensidad está mediada por el apoyo social de los compañeros y los padres de niños y adolescentes de 11 a 18 años^[284] (Figura 2).

En ese estudio se incluyó el sexo de los participantes como una variable moderadora, y se incluyó el IMC, la salud auto percibida y la edad como variables confusoras del efecto. Una variable confusora significa que obstruye o crea ruido (i.e., aumenta la varianza de error) en el análisis de los datos principales, y por lo tanto, debe ser controlada o al menos tomada en consideración. Esto significa, que en la estrategia de análisis de los datos se les considera dentro del análisis estadístico y se logra de esa manera tener indicadores estadísticos “libres” del efecto de esas variables que crean ruido. Es decir, se obtienen promedios y coeficientes de correlación o de regresión “más puros” o ajustados, lo cual permite maximizar la varianza verdadera.

En el estudio, se encontró que parte de la asociación entre el nivel socioeconómico y la actividad física de moderada a vigorosa intensidad en niños y adolescentes está mediada por el apoyo de los padres y los compañeros^[284]. También se encontró que los efectos indirectos condicionales mostraron que tanto el apoyo de los padres como el de los compañeros eran mediadores significativos entre los niños, mientras que el apoyo de los padres era un mediador significativo entre las niñas. Un aspecto relevante de los modelos en los que se analizan en conjunto variables mediadoras, moderadoras y confusoras, es que se debe fundamentar teóricamente cada constructo, y a partir de allí, crear hipótesis que luego serán estudiadas estadísticamente^{[452][467]}.

Figura 2. Ejemplo para comprender el papel de variables mediadoras, moderadoras y confusoras. En el modelo, se estudia el efecto del estatus socioeconómico sobre los niveles de actividad física de moderada a vigorosa intensidad, efecto mediado por el apoyo social de los padres y compañeros. A ese modelo se le añade el posible efecto moderador del sexo de los participantes (rectángulo gris) y se trata de eliminar el efecto confusor (“ruido”) que pueden causar las variables índice de masa corporal, la salud auto percibida y la edad (círculo azul).



Por ejemplo, en el área de la psicología del deporte, se ha estudiado el tema de la relevancia del liderazgo transformacional. El liderazgo transformacional se caracteriza por la construcción de relaciones con los atletas a través de intercambios de experiencias inspiradoras, personales y emocionales. Este tipo de liderazgo motiva a los atletas a desempeñarse más allá de sus expectativas iniciales y superar los niveles normales de desempeño, lo que tiene un efecto positivo y significativo en los resultados grupales^[452].

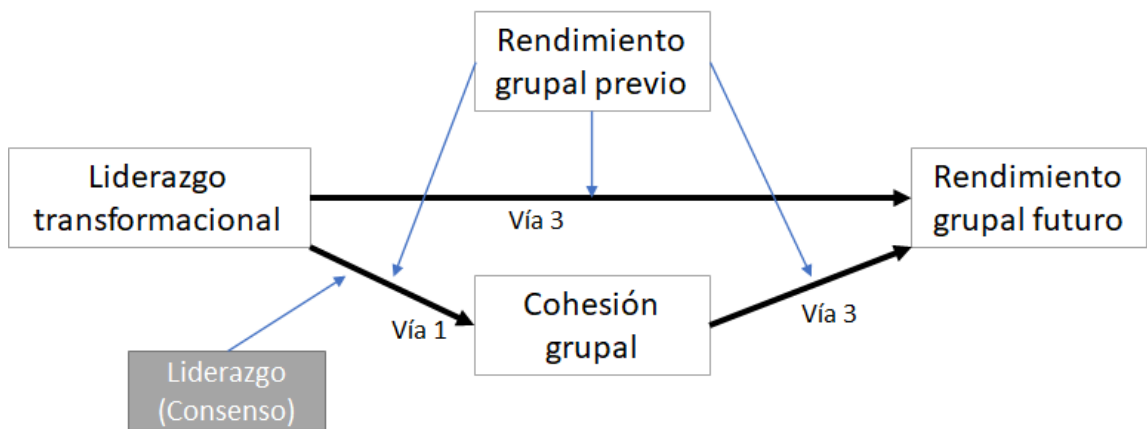
Sin embargo, el liderazgo transformacional puede estar mediado, moderado o puede estar bajo la influencia de confusores en su posible efecto sobre el rendimiento de un equipo deportivo. Así, los investigadores interesados en comprender un fenómeno complejo como el rendimiento grupal podrían establecer hipótesis y luego recolectar datos para comprender las posibles asociaciones de mediación y moderación (Tabla 1). Debido a que las asociaciones son complejas y abstractas, las hipótesis siempre se deben acompañar con un esquema que muestre dichas asociaciones (Figura 3).

Las técnicas de análisis estadístico necesarias para analizar datos de estudios de modelos de mediación suelen requerir conocimientos multivariados avanzados a nivel de maestría o doctorado (e.g., modelo jerárquico lineal longitudinal, regresión múltiple, regresión logística, “path analysis”)^[152].

Tabla 1. Ejemplo de hipótesis de las posibles relaciones entre el liderazgo transformacional y el rendimiento deportivo grupal^[452].

Hipótesis	Explicación de las posibles asociaciones de mediación y moderación
Hipótesis 1	La cohesión del equipo media la relación entre el liderazgo transformacional y el desempeño del equipo.
Hipótesis 2	El desempeño previo del equipo moderará la relación directa (Vía 3) e indirecta entre los comportamientos de liderazgo transformacional y el desempeño futuro del equipo a través de la cohesión del equipo (Vía 1 y 2). Específicamente, un desempeño previo alto (o bajo) (hipótesis 2a) fortalecerá (o debilitará) el efecto directo de los comportamientos de liderazgo transformacional sobre el desempeño futuro y (hipótesis 2b) fortalecerá (o debilitará) la influencia mediadora de la cohesión sobre el efecto del liderazgo transformacional comportamientos sobre el rendimiento.
Hipótesis 3	El consenso del liderazgo del equipo (fuerza del clima) moderará la relación entre el liderazgo transformacional y la cohesión del equipo (Vía 1), por lo que el liderazgo transformacional tendrá un efecto indirecto condicional en el desempeño futuro del equipo a través de la cohesión del equipo y este efecto indirecto será más fuerte cuando el consenso de liderazgo es alto.
Hipótesis 4	El efecto mediador de la cohesión del equipo entre los comportamientos de liderazgo transformacional y el desempeño del equipo será moderado simultáneamente por el consenso de liderazgo entre los miembros (Vía 1) y el desempeño previo del equipo. Específicamente, los efectos indirectos serán más fuertes y significativos cuando el desempeño previo y el consenso de liderazgo sean altos.

Figura 3. Ejemplo para mostrar esquemáticamente las asociaciones teóricas y las hipótesis del estudio. Modificado de Mach et al.^[452].



Por ejemplo, se realizó un estudio para proporcionar una mayor investigación empírica de las vías propuestas dentro del modelo de evitación del miedo en pacientes con dolor lumbar crónico^[467]. Los objetivos específicos del estudio fueron: a) investigar si el catastrofismo, en combinación con el miedo y la angustia psicológica, mediaba la relación entre el dolor y la discapacidad, b) investigar si el catastrofismo mediaba la relación entre el dolor y el miedo, y c) examinar si el compromiso con la actividad física estructurada regular moderó el efecto indirecto del catastrofismo, el miedo y la angustia psicológica en la relación entre el dolor y la discapacidad.

Para responder a esos objetivos, los investigadores realizaron un análisis de mediación múltiple sugerido por Hayes^[323] para examinar si la relación entre el dolor y la discapacidad se explicaba por el miedo, el catastrofismo, la depresión y la ansiedad. Estos análisis usualmente se realizan con la ayuda de programas especializados y macros (i.e., rutinas de instrucciones), como por ejemplo, los del programa PROCESS (www.processmacro.org), una herramienta de modelado de análisis de ruta ("path analysis") de regresión logística y mínimos cuadrados ordinarios de variable observada que se puede usar con paquetes estadísticos como SPSS, SAS y R^[324].

Para obtener estimadores sólidos también se recurre a la técnica del remuestreo o "bootstrapping", un método estadístico que consiste en extraer muestras repetidas de los datos con reemplazo para obtener múltiples estimaciones del efecto indirecto atribuido a las posibles variables mediadoras^{[580][681]}. Los investigadores concluyeron que el miedo, la depresión y el catastrofismo median la relación entre el dolor y la discapacidad en personas con dolor lumbar crónico. El efecto mediador del catastrofismo, pero no el miedo o la depresión, estuvo condicionado a que los participantes reportaran el desempeño semanal de sesiones estructuradas de actividad física. Por lo tanto, los pacientes con dolor de espalda crónico que realizan actividad física regular pueden necesitar intervención psicológica y apoyo para las percepciones negativas del dolor. El efecto del miedo y la depresión sobre la discapacidad relacionada con el dolor no se relacionó con la actividad física regular, lo que sugiere que las intervenciones psicológicas son probablemente la mejor opción de tratamiento para estos factores.

En síntesis, los análisis de mediación utilizados en las Ciencias del Movimiento Humano permiten estudiar los efectos directos o indirectos de variables asociadas con una respuesta particular, lo cual enriquece la interpretación de la información y la integración de múltiples variables para explicar el fenómeno de estudio. Por lo tanto, son análisis multivariados que permiten acercarse cada vez más a la realidad; es decir, su aplicación asemeja cada vez más el mundo real donde ocurren los eventos.

2.8.7. Análisis de varianza multivariante permutacional

La prueba paramétrica de análisis de varianza (ANOVA) es poderosa para el análisis de datos univariados; sin embargo, las pruebas multivariantes o multivariadas son muy estrictas en sus supuestos; por ello, se prefiere utilizar métodos no paramétricos basados en pruebas de permutación^[22]. El ANOVA multivariante permutacional (PERMANOVA) es una partición geométrica de la varianza a través de una nube de datos multivariante, definida explícitamente en el espacio de una medida de disimilitud elegida en respuesta a uno o más factores en un diseño tradicional de ANOVA^[24].

Con PERMANOVA, las inferencias estadísticas se realizan en un entorno sin distribución utilizando algoritmos permutacionales para acomodar efectos aleatorios, modelos jerárquicos, modelos mixtos, covariables cuantitativas, medidas repetidas, diseños desequilibrados o asimétricos y, más recientemente, dispersiones heterogéneas entre grupos^[24]. La estadística de la prueba de hipótesis es un análogo multivariante de la F de Fisher y se calcula directamente a partir de cualquier matriz de disimilitud o distancia simétrica. Los valores de significancia (i.e., p) se obtienen posteriormente usando permutaciones^[23]. Los gráficos que acompañan los modelos de PERMANOVA incluyen ordenaciones de matrices de distancia ajustadas o residualizadas, incluidos análogos multivariados de los efectos principales y gráficos de interacción, para visualizar los resultados^[24].

El cálculo de PERMANOVA requiere de software especializado que permita analizar una amplia cantidad de datos^[23]. En las Ciencias del Movimiento Humano el uso de PERMANOVA es apenas incipiente, en donde se ha aplicado especialmente en el área de la salud intestinal^{[16][216][514][683]}. Por ejemplo, se sabe que el ejercicio físico está asociado a una composición microbiana intestinal alterada en los seres humanos, pero no se sabe si la microbiota intestinal y los metabolitos asociados están modulados por el entrenamiento físico. Un grupo de investigadores determinaron el efecto de 6 semanas de ejercicio aeróbico en la composición, la capacidad funcional y el rendimiento metabólico de la microbiota intestinal en adultos delgados y obesos con controles dietéticos de varios días antes de la recopilación de las variables de estudio^[16]. Para ello, midieron a 18 personas delgadas y 14 personas obesas, quienes fueron sometidas a ejercicio aeróbico 3 días/semana durante 30 a 60 min y a una intensidad de entre 60% y 75% de la frecuencia cardíaca de reserva. A los participantes se les solicitaron muestras fecales antes y después de las 6 semanas de ejercicio. Los datos de la evaluación de taxonomía y β -diversidad de la microbiota intestinal se analizaron con muestras uniformes de 14038 secuencias por muestra mediante PERMANOVA, en donde se encontró que las alteraciones de la microbiota intestinal inducidas por el ejercicio dependían del grado de obesidad.

Se espera que la técnica de PERMANOVA se utilice más frecuentemente en las Ciencias del Movimiento Humano para resolver problemas de investigación de naturaleza multivariada, como, por ejemplo, la obesidad, el rendimiento deportivo y el sedentarismo, entre otros. El avance informático permitirá procesar millones de datos rápidamente para responder a preguntas que no tienen una solución univariada y que requieren de la integración de varias fuentes de información.

2.8.8. Análisis de Rasch

La teoría clásica de las pruebas (TCP) se ha utilizado para desarrollar y evaluar pruebas, incluyendo una amplia lista de instrumentos comúnmente utilizados en las Ciencias del Movimiento Humano (e.g., pruebas motoras y físicas). Los métodos que sustentan los análisis de las pruebas TCP para obtener la validez son el análisis factorial y las correlaciones. Existen críticas en el ámbito motor^[55], en donde se afirma que el enfoque TCP está enfocado en los ítems, lo que implica que los ítems forman la variable latente, lo que a la vez indica la dependencia de la muestra y el ítem de los resultados de la prueba. Otra limitación de este enfoque es el requisito de variables escaladas por intervalos para calcular las correlaciones debido a la falta de estándares de interpretación relativos y absolutos.

Como alternativa a las pruebas TCP, en las Ciencias del Movimiento Humano desde los años 90's se ha utilizado el modelo Rasch, llamado así en honor al matemático de origen danés Georg William Rasch^{[775][778][779]}. Rasch es un modelo psicométrico para analizar datos categóricos (e.g., respuestas a preguntas en una evaluación de lectura o respuestas a cuestionarios), aunque también se ha utilizado en áreas como el aprendizaje motor en pruebas de habilidad. La teoría matemática que la subyace es un caso especial de la teoría de respuesta al ítem (TRI). A diferencia de la TRI, que ajusta un modelo a los datos, un modelo de Rasch se crea a partir de los datos mismos, y se le conoce como modelo logístico de un parámetro. Los pasos básicos para realizar un análisis de Rasch incluyen seleccionar el modelo de medición más apropiado, estimar los parámetros del ítem y la habilidad utilizando alguna herramienta informática, realizar un análisis de bondad de ajuste, e interpretar los parámetros estimados si los datos y el modelo propuesto se ajustan^[775].

Con el modelo Rasch cuando la habilidad de la persona que está siendo evaluada es mayor que la dificultad que le impone el ítem o prueba, entonces aumenta la probabilidad de que la persona sea más exitosa, y viceversa. Una de las ventajas del análisis de Rasch aplicado a una escala continua es que los parámetros de dificultad de la prueba y la habilidad de la persona que está siendo evaluada son determinadas en una escala "log-odds", que posee las mismas propiedades que una escala continua^[775]. A esas unidades resultantes se les llama "logits", y se les puede interpretar como probabilidades.

Otra de las ventajas de utilizar el modelo Rasch es que se puede estimar la "invariancia"; es decir, las estimaciones de parámetros, como la dificultad de los elementos y la capacidad del examinado, son estadísticamente independientes de los elementos de prueba y las muestras de los examinados empleados. También se utiliza una métrica común para los parámetros del ítem y habilidad, lo que proporciona una gran conveniencia para determinar la idoneidad de los ítems usados. Otra ventaja es que las puntuaciones brutas son estadísticas suficientes para los parámetros de los ítems y los examinados, lo que significa que las puntuaciones brutas totales se pueden utilizar en el análisis de Rasch. Por último, la precisión de la medición se puede determinar en cualquier nivel de habilidad del participante^[775].

El modelo Rasch ha sido utilizado para apoyar el proceso de validación de instrumentos en Educación Física y en otras áreas de las Ciencias del Movimiento Humano^{[439][673][691][778]}. En Canadá, se hizo una evaluación de las habilidades de movimiento fundamentales, combinadas y complejas requeridas para apoyar la alfabetización física infantil, a la cual se le calculó además la evidencia

de validez utilizando el modelo de Rasch^[439]. La prueba, llamada CAMSA, utilizó un sistema de puntuación de 20 puntos para la evaluación de habilidades, y para mejorar la viabilidad de los procedimientos de puntuación de la evaluación, se ajustó un modelo de Rasch a la puntuación inicial de 20 puntos. Ese análisis de Rasch compara los datos recopilados con lo que se esperaría en función de un modelo teórico. Luego, se asignó a cada elemento una calificación de dificultad, en donde las categorías fueron: fácil (< -0.7 logits), media (-0.7 a 0.7 logits) o difícil (>0.7 logits). Los ítems con puntajes logit similares y baja singularidad se eliminaron o combinaron para mantener un rango similar de dificultad del ítem dentro de un puntaje de habilidad de 14 puntos que se esperaba que fuera más fácil de administrar. Los investigadores concluyeron que la prueba CAMSA es una evaluación factible de una muestra de habilidades de movimiento fundamentales, complejas y combinadas para niños entre 8 y 12 años.

En otro contexto, se desarrolló un instrumento utilizando el modelo de Rasch de múltiples facetas a través del juicio de expertos^[778]. El instrumento consistió de un inventario de orientación de valor de 150 elementos-2 (IOV-2) que evalúa el valor de los objetivos del plan de estudios de Educación Física. El instrumento fue desarrollado y evaluado por 128 educadores universitarios y 103 educadores físicos que laboraban en educación primaria. Se pidió a los expertos que calificaran la consistencia de cada elemento para representar una parte de los objetivos generales del plan de estudios utilizando una escala de calificación de 5 puntos. El modelo Rasch se usó para calibrar las puntuaciones de calificación y se definieron 6 facetas: a) género, b) grupo étnico, c) tipo de empleo, d) evaluador, e) área de contenido, y f) ítem. Los resultados sugieren que la mayoría de los ítems del IOV-2 eran representativos del dominio de contenido y que los evaluadores eran verdaderamente expertos, por lo que se concluye que el modelo multifacético de Rasch demuestra ser una técnica psicométricamente apropiada para aplicar el juicio experto en el desarrollo de pruebas.

Al desarrollar nuevos instrumentos asociados a percepciones y cogniciones en el área de la psicología del deporte, también se puede recurrir al modelaje de Rasch. Por ejemplo, se estudió un instrumento para evaluar la fortaleza mental en deportistas y no deportistas, en el que 261 participantes completaron un cuestionario de 45 ítems y se examinaron las dificultades de los elementos calibrados por medio de Rasch. Se encontró que 43 de los 45 ítems (i.e., 96%) tenían un buen ajuste de los datos del modelo con estadísticas de ajuste aceptables, por lo que se confirmó la validez del instrumento^[454].

A pesar de que los análisis de Rasch contienen complejas fórmulas estadísticas, se ha descrito que es un modelo de fácil comprensión^[775], por lo que se espera que en los próximos años se utilice más este modelo robusto para el análisis de datos. De hecho, lo primero que se debe hacer es "calibrar" el instrumento con el modelo de Rasch. En el contexto de las Ciencias del Movimiento Humano, calibración significa el esfuerzo conjunto de expertos en la materia, especialistas en medición y otros profesionales, quienes brindan un juicio acerca de la medición en particular. Se debe recordar que el modelo de Rasch tiene como característica que las puntuaciones brutas totales proporcionan estadísticas suficientes de las puntuaciones de habilidad, lo que significa que las habilidades de los examinados se pueden estimar a partir de sus puntuaciones totales después de haber calibrado la escala de calificación. Es por ello que Zhu^[775] indica que el usuario final no necesita saber nada sobre el modelo de Rasch en sí, excepto saber cómo convertir puntajes totales a logits a través de una tabla de conversión o simplemente una curva de conversión proporcionada por los desarrolladores de escalas.

2.8.9. Análisis discriminante espectral

Existen entornos de investigación en los cuales se recolectan datos continuos que posteriormente son categorizados. Por ejemplo, por medio de acelerometría se puede determinar la cantidad de pasos que una persona alcanza durante un período de tiempo. Posteriormente, esa cantidad de pasos se puede categorizar, por ejemplo, en pocos pasos, moderados pasos o muchos pasos, dependiendo de algún criterio técnico que definan los investigadores. También se podrían hacer muchas más categorías, y tener un espectro de posibilidades; sin embargo, ese enfoque presenta el problema de que las categorías son altamente colineales^[1]. En ese contexto, al tratar de utilizar estadísticas tradicionales para comparar grupos o predecir resultados (e.g., ANOVA, regresión lineal múltiple), se presenta el problema del manejo de la colinealidad^[270].

Actualmente, se han propuesto técnicas de análisis estadístico multivariado. Una de estas técnicas es la regresión de mínimos cuadrados parciales (PLS, por sus siglas en inglés), que permite analizar un espectro de categorías. La técnica sustituye el tradicional análisis de regresión al convertirlo en multivariado, y se sabe que PLS descompone la información de las variables predictoras o independientes (X_n) en una serie de variables latentes (\hat{X}_n), de manera que permite maximizar la covarianza con la variable de respuesta, dependiente o criterio (\hat{Y})^[270]. La técnica PLS también se puede utilizar para analizar diferencias grupales mediante análisis discriminante^[57].

En la literatura de las Ciencias del Movimiento Humano no se encuentran abundantes estudios en los que se aplique la técnica de análisis discriminante espectral. Recientemente, se aplicó para el análisis de datos de actividad física^[270]. En el estudio se analizaron dos grupos de niños; uno que recibía clases adicionales de educación física, y otro que siguió el horario habitual. Los datos de actividad física se obtuvieron por medio de acelerómetros y se categorizaron de acuerdo a la intensidad. Posteriormente, se aplicó la PLS multivariante para análisis discriminante (PLS-DA) para comparar los grupos del espectro de intensidad de actividad física y se comparó con el análisis estadístico univariante tradicional. Se encontró que el análisis tradicional de las categorías crudas de intensidad no mostró diferencias entre los grupos, aunque el análisis PLS-DA sí lo hizo.

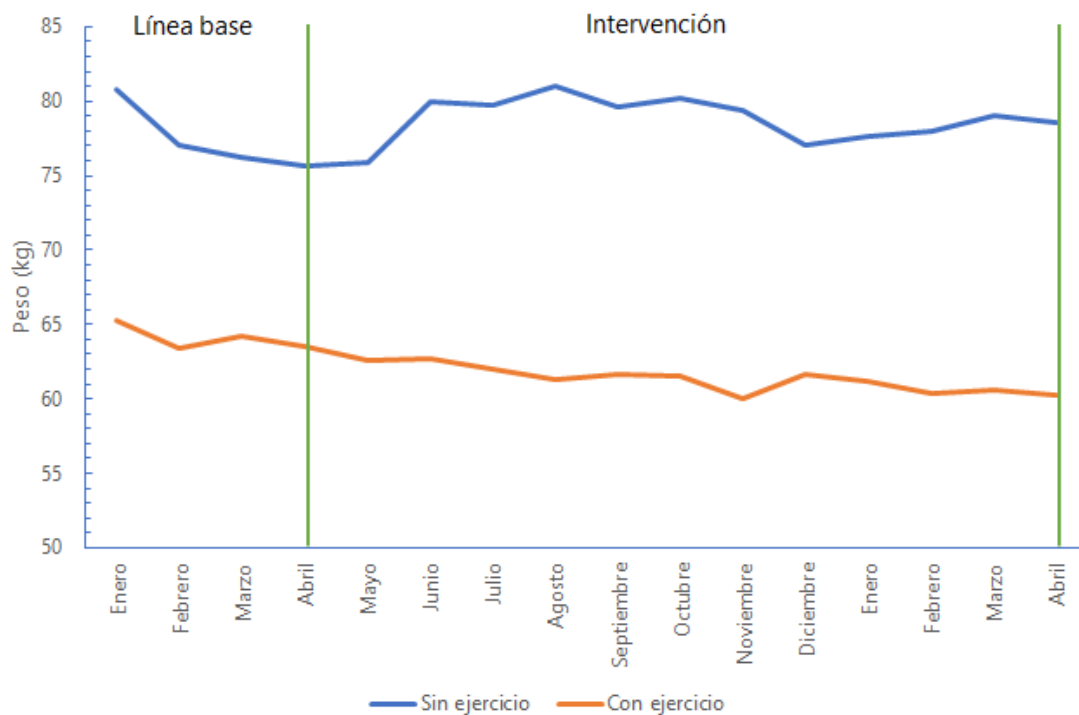
Todavía queda por explorar nuevas formas de analizar datos continuos provenientes de tecnología inercial y de acelerómetros triaxiales, así como de dispositivos que permiten recolectar grandes cantidades de datos. Los análisis PLS-DA servirán para obtener conclusiones más robustas que las que actualmente se obtienen por medio de técnicas tradicionales de análisis estadísticos. Esto significa que se debe consultar a profesionales en estadística para decidir la mejor ruta de análisis de datos.

2.8.10. Series de tiempo

Existen diseños de investigación que involucran recolectar datos durante algún tiempo determinado, para posteriormente estudiar los datos y así conocer la respuesta de la variable de interés a través de ese período. Para lograr esos análisis, se utilizan diversas técnicas, entre las cuales se encuentra el análisis de series temporales (AST) (en inglés, time-series analysis). Para estimar un AST los investigadores anotan puntos de datos a intervalos constantes durante un período de tiempo determinado en lugar de simplemente registrar los puntos de datos de forma intermitente o aleatoria^[188].

Así, una serie temporal es una sucesión de datos medidos en momentos determinados y ordenados cronológicamente (Figura 1). En las Ciencias del Movimiento Humano, los datos pueden estar espaciados a intervalos iguales (e.g., mediciones automáticas de presión arterial cada 30 min, datos de acelerómetros para medir el sueño durante la noche)^[143], o desiguales (e.g., % de grasa de una persona durante un año). El AST utiliza métodos para interpretar los datos y conocer las posibles relaciones subyacentes entre los datos de la serie.

Figura 1. Una serie temporal hipotética de un grupo de personas que participaron en una investigación para conocer el efecto de una intervención de ejercicio físico sobre el peso corporal. El protocolo incluía varios meses de mediciones de control y 12 meses de valoraciones periódicas.



Generalmente, el AST se utiliza para predecir el comportamiento de la serie en momentos en que la variable no fue medida. En ese caso, si se desea predecir en el futuro se le llama extrapolación pronóstica^[699], si es para explicar datos del pasado, se le llama extrapolación retrógrada^[524], y si es para momentos intermedios, se le denomina interpolación.

Por ejemplo, se diseñó un estudio para conocer el posible efecto de la temperatura ambiental en la atención de lesiones deportivas en hospitales de España, tratando de explicarlas por los cambios en las estaciones del año^[699]. Los investigadores calcularon la proporción de visitas a las salas de emergencia atribuidas a los cambios en las estaciones del año antes y después de ajustar por la influencia de la temperatura ambiental diaria. Para analizar los datos recolectados, utilizaron el AST por medio de modelos de regresión cuasi-Poisson. En general, se encontró que la proporción de visitas a las salas de emergencias atribuibles a los cambios estacionales fue de 24.1% y disminuyó al 7.6% después de ajustar por la influencia de la temperatura. Así los investigadores lograron concluir que existe un alto riesgo de visitas de emergencia asociadas con temperaturas frías y calientes, mientras que el riesgo fue mayor para el calor.

En síntesis, el AST es una opción robusta de análisis de datos poco explorada en las Ciencias del Movimiento Humano; sin embargo, se espera que pronto se considere su utilización para el análisis de datos recolectados en serie, ya sea por medios manuales o electrónicos. Estos análisis pueden complementar los tradicionales análisis de mediciones repetidas como los que se hacen con ANOVA, lo cual fortalecerá el reporte e interpretación de los datos. Incluso, para quienes ya han recolectado datos durante años, podrían utilizar las extrapolaciones retrógradas y obtener información valiosa.

2.8.11. Análisis del perfil latente

En algunas áreas de las Ciencias del Movimiento Humano a menudo se recolecta información con el propósito de identificar patrones de agrupamiento de los datos. Por ejemplo, en el ámbito de la psicología del ejercicio, se pueden recolectar datos para saber si existen diferentes perfiles de motivación hacia las clases de Educación Física en el ámbito escolar. Uno de los métodos de agrupamiento o de identificación de esos "clusters" o conglomerados es el análisis de perfil latente (APL, del inglés "Latent Profile Analysis"). El APL es un enfoque de variable latente categórica que identifica subpoblaciones latentes (i.e., patrones de agrupamiento) dentro de una población en función de un determinado conjunto de variables^[664].

El APL es un tipo de análisis que supone que las personas pueden clasificarse de acuerdo con diversos grados de probabilidad en categorías que tienen diferentes perfiles de configuración de atributos personales^[664]. El APL se orienta en patrones de variables llamados "indicadores" e identifica los perfiles de personas que comparten patrones similares de variables, los cuales a la vez se comparan con otros perfiles en términos de cómo se combinan las variables para formar los perfiles y de cómo esas combinaciones se relacionan diferencialmente con predictores y resultados^[735].

En las Ciencias del Movimiento, el APL se ha utilizado en diferentes ámbitos. En las investigaciones se comienza midiendo variables de interés que poseen un sustento teórico, y lo que se pretende es identificar categorías de pertenencia a partir de las mismas. Por ejemplo, combinaciones de variables se analizan con APL en temas como la motivación en golfistas^[635], comportamiento sedentario y riesgo de enfermedades metabólicas y muerte^{[242][730]}, creencias irracionales, motivación y salud en deportistas^[708], fortaleza mental en atletas^[771], conmociones cerebrales en atletas^[493], características de corredores^[696], motivación y ejercicio físico en niños y adultos^{[250][438]}, y planteamiento de objetivos de actividad física y deporte en adolescentes y adultos jóvenes^[303] (Tabla 1). Se espera una creciente utilización de APL para resolver asociaciones y agrupaciones de personas en grupos con características similares. Esta técnica de análisis ha demostrado ser robusta y se requiere la recolección de cantidades importantes de datos para que las conclusiones sean sólidas.

Tabla 1. Ejemplos acerca de la utilización de análisis de perfil latente (APL) en las Ciencias del Movimiento Humano.

Referencia	Propósito(s)	Hallazgos principales
Ekblom-Bak et al. ^[242]	<p>a. Identificar y describir las características de los patrones naturales de ejercicio, sentarse en el tiempo libre y en el trabajo y la aptitud cardiorrespiratoria.</p> <p>b. Determinar la asociación de los perfiles con los factores de riesgo metabólicos, la salud percibida y los síntomas percibidos.</p>	<ul style="list-style-type: none"> • El APL identificó y etiquetó 6 perfiles basados en el ejercicio, sentarse en el tiempo libre y en el trabajo y la aptitud cardiorrespiratoria: <ul style="list-style-type: none"> ○ Perfil 1: Inactivo, baja condición física y tiempo promedio sentado en actividades de ocio, con menos tiempo sentado en el trabajo. ○ Perfil 2: Inactivo, baja condición física y sedentario. ○ Perfil 3: Activo y con condición física promedio, con menos tiempo sentado en el trabajo. ○ Perfil 4: Activo y con condición física promedio y sedentario en el tiempo de ocio, con un trabajo sedentario. ○ Perfil 5: Activo y con buena condición física y con un trabajo sedentario. ○ Perfil 6: Activo y con buena condición física y con menos tiempo sentado en el trabajo. • Los perfiles 1 y 2 demostraron una salud metabólica y percibida más adversa. • El perfil 4 tuvo una salud más beneficiosa que el perfil 3. • El perfil 6 tuvo una salud más beneficiosa que el perfil 5.
Emm-Collison et al. ^[250]	<p>a. Identificar perfiles motivacionales para el ejercicio, utilizando la Teoría de la Autodeterminación en padres de niños de escuela primaria del Reino Unido.</p> <p>b. Explorar el movimiento entre perfiles motivacionales durante un período de cinco años.</p> <p>c. Examinar las diferencias entre los perfiles motivacionales en términos de género, actividad física e IMC.</p>	<ul style="list-style-type: none"> • El APL identificó 6 perfiles con diferentes combinaciones de tipos de motivación. • Entre cada punto de tiempo era más probable moverse entre perfiles que permanecer en el mismo. • Era poco probable que las personas con un perfil más autónomo en un punto de tiempo anterior pasaran a perfiles más controlados o desmotivados. • En los tres puntos de tiempo, los perfiles más autónomos se asociaron con niveles más altos de actividad física de intensidad moderada a vigorosa y un IMC más bajo.
		Continúa...

Tabla 1. Continuación.

Referencia	Propósito(s)	Hallazgos principales
Gut et al. ^[303]	Determinar los objetivos en el deporte y el ejercicio, la motivación autodeterminada, el comportamiento deportivo y de ejercicio y el género en adolescentes y adultos jóvenes.	<ul style="list-style-type: none"> • El APL identificó 6 perfiles de objetivos para ambos grupos de edad. • En adolescentes, las metas como el contacto y la percepción del desafío fueron más relevantes. • En adultos jóvenes, la salud, la figura/apariencia física y la distracción/catarsis eran predominantes. • Ambos perfiles difieren en la motivación autodeterminada, el comportamiento deportivo y de ejercicio y el género.
Lohbeck et al. ^[438]	<p>a. Examinar los perfiles de motivación de niños pequeños en los deportes y las actividades físicas.</p> <p>b. Determinar las relaciones de los perfiles con varios predictores y resultados de rendimiento.</p>	<ul style="list-style-type: none"> • El APL encontró una solución de tres perfiles de motivación: a) desmotivado, b) no autodeterminado, y c) autodeterminado. • Los niños mayores y los niños con un autoconcepto físico más bajo tenían más probabilidades de ser miembros del perfil desmotivado en relación con los otros perfiles que los niños más pequeños y los niños con un autoconcepto físico más alto. • Los niños del perfil autodeterminado demostraron el mejor rendimiento físico en diversas pruebas de motricidad.
Milroy et al. ^[493]	Categorizar y describir a los atletas en función de los factores relacionados con la revelación de conmociones cerebrales.	<ul style="list-style-type: none"> • El APL identificó 5 perfiles de los estudiantes-atletas (desde el más arriesgado hasta el menos arriesgado). • Las mujeres tenían más probabilidades de estar en el perfil menos arriesgado. • Quienes participaban en deportes de colisión tenían más probabilidades de estar entre los 2 perfiles más riesgosos. • Los atletas de deportes de contacto y de temporada tenían menos probabilidades de estar en el perfil menos riesgoso.
		Continúa...

Tabla 1. Continuación.

Referencia	Propósito(s)	Hallazgos principales
Ponnusamy et al. ^[573]	a. Identificar el número y tipo de perfiles de uso de habilidades y técnicas psicológicas (HTP) por parte de los atletas de élite. b. Examinar las diferencias entre los perfiles de HTP en términos de la fortaleza mental autoreportada por los atletas.	<ul style="list-style-type: none"> • El APL reveló tres perfiles de uso de HTP: a) bajo, b) moderado, y c) alto. • Un mayor uso de HTP se asoció con niveles más altos de fortaleza mental.
Schaefer et al. ^[635]	Examinar si el perfil de motivación personal estaba asociado con la ansiedad competitiva a través de la fortaleza mental en golfistas competitivos.	<ul style="list-style-type: none"> • El APL identificó tres perfiles de motivación distintos: a) alta, b) moderada, y c) baja. • Los golfistas que reportaron niveles bajos o moderados de motivación tenían niveles más altos de ansiedad competitiva. • Las asociaciones estaban parcialmente mediadas por niveles más bajos de fortaleza mental.
von Rosen et al. ^[730]	a. Identificar perfiles de actividad física en función de las características de actividad derivadas de acelerometría. b. Determinar si los perfiles se asocian con la mortalidad por todas las causas.	<ul style="list-style-type: none"> • El APL identificó tres perfiles de actividad: a) baja, b) media y c) alta. • La baja actividad se asoció positivamente con la mortalidad por todas las causas. • No se encontraron diferencias en el riesgo de mortalidad entre actividad alta y media.
Thuany et al. ^[696]	Identificar y describir el agrupamiento de características relacionadas con la carrera entre corredores brasileños y verificar si existe un perfil asociado con un mejor rendimiento.	<ul style="list-style-type: none"> • El APL identificó dos clases de corredores: a) aficionados, y b) recreativos. • Las variables latentes que destacaron las diferencias de las clases fueron: <ul style="list-style-type: none"> ○ Volumen de entrenamiento/semana. ○ Frecuencia de entrenamiento/semana. ○ Motivación para la práctica. ○ IMC. • Hombres jóvenes y aficionados fueron más propensos a tener un mayor rendimiento.
		Continúa...

Tabla 1. Continuación.

Referencia	Propósito(s)	Hallazgos principales
Turner et al. ^[708]	Identificar subgrupos a partir de datos relacionados con creencias irracionales, motivación y salud en deportistas y estudiantes atletas.	<ul style="list-style-type: none">• El APL identificó el perfil de dos clases: a) altas creencias irracionales, motivación autodeterminada baja y resultados de salud deficientes, y b) bajas creencias irracionales, alta motivación autodeterminada y mejores resultados de salud.
Zeiger & Zeiger ^[771]	<ol style="list-style-type: none">a. Determinar los perfiles de fortaleza mental (FM) en atletas de resistencia.b. Determinar si existen asociaciones entre los perfiles latentes de FM y las características demográficas y deportivas.	<ul style="list-style-type: none">• El APL identificó tres clases de FM: a) alta, b) moderada, y c) baja.• Hubo mayores probabilidades para los hombres de estar en la clase de alta FM en comparación con la clase de baja FM.• Hubo mayores probabilidades para los atletas que tenían más de 55 años en comparación con los que tenían de 18 a 34 años de estar en la clase alta FM en comparación con la clase baja FM.• Hubo mayores probabilidades para quienes tenían una alta satisfacción deportiva de estar en la clase de alta FM en comparación con la clase de baja FM.• Hubo mayores probabilidades para quienes se ubicaban en divisiones de competencia altas de estar en la clase de alta FM en comparación con la clase de baja FM.

2.9. Administración de la ciencia

2.9.1. ¿Cómo comunicar los hallazgos de la investigación?

El planeamiento, la ejecución y el reporte de hallazgos científicos es de vital importancia para el avance de las ciencias, incluyendo las Ciencias del Movimiento Humano. La preparación de un manuscrito para enviar a una revista científica es una tarea de alta complejidad que requiere de un esfuerzo para escribir con adecuada gramática y ortografía^[639], y de un nivel de tolerancia que permita recibir las críticas de los pares académicos^[612]. Para atenuar las críticas fundamentadas e infundadas de los pares académicos, es indispensable tener una efectiva estrategia de comunicación de los hallazgos que permita transmitir el mensaje deseado de la forma más clara posible.

Figura 1. Ejemplo de infografía que resume el hallazgo principal de un estudio y que se utiliza como apoyo visual para el artículo publicado en una revista científica.



En algunas ocasiones los científicos carecen de herramientas, creatividad o preparación para comunicar los hallazgos de una manera comprensible y clara. La presentación de los hallazgos representa una combinación entre números resumidos en tablas, en gráficos y más recientemente, en formas creativas como las infografías (Figura 1), video-revistas y el uso de las múltiples redes sociales^{[226][233][406][721]}. Con respecto a este último medio de difusión, se ha argumentado que, a pesar de su amplia cobertura, las redes sociales no cuentan con un sistema de control de calidad, por lo que la información que se “consume” debe ser inteligente y críticamente escudriñada para poder diferenciarla de la información sesgada, engañosa o falsa (e.g., “fake news”)^{[406][679]}.

Se ha indicado que los datos a menudo se presentan mejor gráficamente que en tablas o como números esparcidos por todo el documento^[639]. Así, los gráficos que muestran valores individuales son mucho más recomendables que los tradicionales gráficos de barras sólidas que indican un valor promedio. A menos que el número de datos sea grande, se puede usar un gráfico de caja (i.e., box plot), ya que reduce el énfasis en la tendencia central y porque muestra la verdadera distribución de los datos^[226]. Las revistas generalmente especifican en las instrucciones para los autores, el formato (i.e., tamaño de letra, colores, patrones, tamaño y densidad de las figuras) y la calidad de las tablas, figuras, gráficos y reporte de estadísticas (e.g., tamaño de efecto, IC95%) para mejorar la difusión de los hallazgos^[583].

En esta sección se presentan alternativas para el reporte de datos provenientes de las investigaciones en las Ciencias del Movimiento Humano. Las recomendaciones se basan en la evidencia publicada recientemente y que a menudo se observa en las revistas de alta calidad. A la vez, se proporcionan recursos gratuitos para que puedan ser diseminados entre estudiantes e investigadores.

2.9.2. Graficación de los hallazgos de la investigación

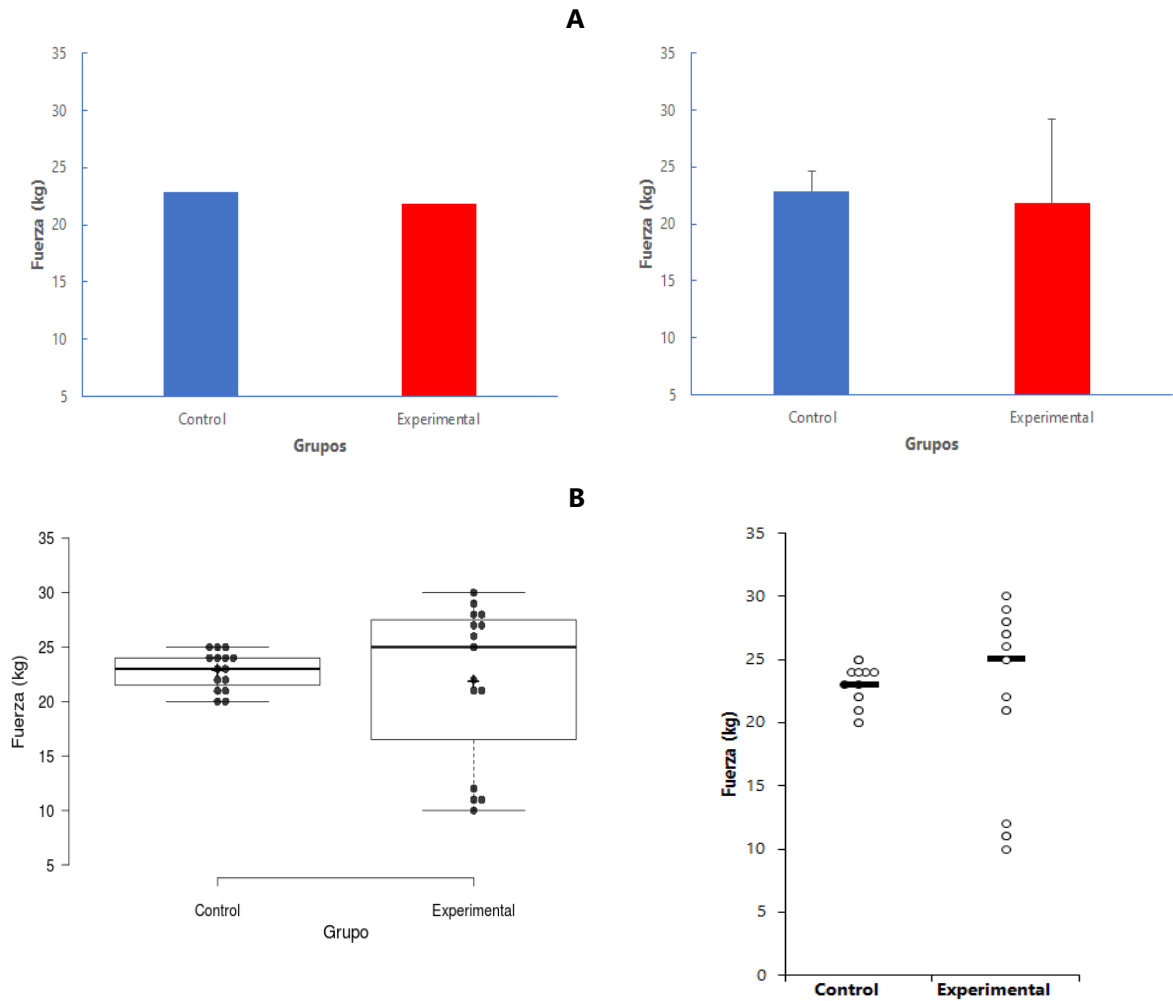
El adagio popular “Una imagen vale más que mil palabras” es sin lugar a dudas muy cierto en el reporte de datos complejos, pues lo que se desea es transmitir un mensaje complejo de la forma gráfica o visual más simple posible. En las Ciencias del Movimiento Humano existen múltiples formas para el reporte visual creativo para resumir datos estadísticos, o incluso, para representar esquemas mentales o mapas conceptuales. La creatividad es necesaria para generar nuevas representaciones gráficas, pues ha existido una crítica asociada a la alta dependencia de los gráficos de barras y de líneas simples que generalmente resumen resultados, ya que no siempre representan claramente los hallazgos grupales o individuales, lo cual puede conducir a un sesgo de interpretación^{[337][746]}.

Recientemente, se ha indicado que el reporte de los hallazgos de las investigaciones debería permitir una mayor transparencia, pues algunas veces parece que se desea ocultar información; por lo tanto, el mensaje actual es “muestre los datos, no los esconda”^[744]. Por ejemplo, en la actualidad se recomienda mostrar en cada gráfico, todos los datos individuales junto a alguna medida de variabilidad (Figura 1), así como mostrar las barras de error solamente si el tamaño de la muestra es grande, ya que el error estándar depende mucho del tamaño de la muestra^[745].

También, se recomienda crear gráficos de puntos si el tamaño de la muestra es muy pequeño y mostrar la mediana de los datos^[744]. Para una serie de ejemplos concretos en el área de la medicina, incluyendo la del deporte, se recomienda consultar los trabajos de Hertel^[337] y Weissgerber et al.^[744]. También se recomienda visitar el sitio web de CTSpedia para obtener ejemplos de plantillas de Microsoft Excel para la graficación de datos provenientes de grupos independientes y de medidas repetidas, así como ejemplos de graficación para el programa IBSM SPSS^[745].

En la actualidad, debido a la existencia de interfaces computacionales y potentes algoritmos, se pueden resumir miles o millones de datos en un simple gráfico. Existen lenguajes de programación que permiten ese procesamiento de datos o señales (e.g., MATLAB, Python), así como programas gratuitos y por suscripción en los cuales simplemente se ingresan los datos para generar la imagen (e.g., Hotjar, Smartlook, Excel). Por ejemplo, los mapas de calor (“heat maps”) permiten mostrar los datos de dos dimensiones (i.e., ejes X, Y) en colores, en los cuales el tono o densidad del color varía en función de la magnitud de la probabilidad estadística, de la diferencia, y de la asociación o correlación estadística^{[520][667]} (Figura 2). Así, en estudios en que se utiliza una intervención de ejercicio físico, se puede discernir gráficamente, el efecto de la intervención o la relación entre las diferentes variables estudiadas.

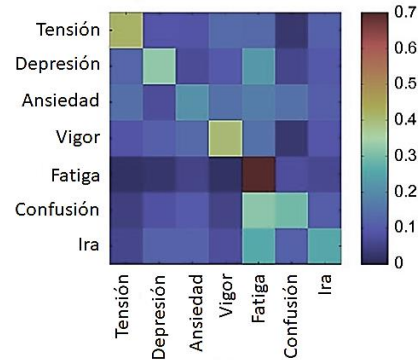
Figura 1. Ejemplo de graficación. Actualmente no se recomienda mostrar únicamente los gráficos con barras (Panel A), sino que se prefiere mostrar todos los datos obtenidos y sus indicadores de variabilidad (Panel B). En este ejemplo, todos los gráficos fueron construidos a partir de los mismos datos; sin embargo, la información transmitida visualmente no es la misma. Los gráficos del panel B son más informativos acerca de la distribución de los puntajes y de su variabilidad. Los datos del grupo control son más homogéneos que los del grupo experimental.



En las Ciencias del Movimiento Humano, la visualización por medio de mapas de calor se utiliza con mayor frecuencia para resumir datos de estudios con múltiples variables o con una cantidad de datos grande, como, por ejemplo, cuando se miden miles de células del sistema inmune, bacterias de la microbiota intestinal, y datos provenientes de variables biomecánicas en donde la frecuencia de muestreo es alta y se graban cientos o miles de datos por segundo^[94,546,607]. También se utilizan mapas de calor para representar los cambios en variables como la temperatura corporal en diferentes áreas anatómicas^[655] (Figura 3).

Figura 2. Ejemplo de mapas de calor en el que se grafican las correlaciones entre las variables psicológicas (Panel A) y de cualquier otro tipo (Panel B). En la escala de la derecha de la figura del panel A se muestra el valor de correlación correspondiente a la gradiente de color del mapa de calor. En el mapa del panel B se agrupa un dendrograma que reúne agrupaciones de variables (i.e., clusters) y a la vez grafica la densidad de las asociaciones de múltiples variables.

A



B

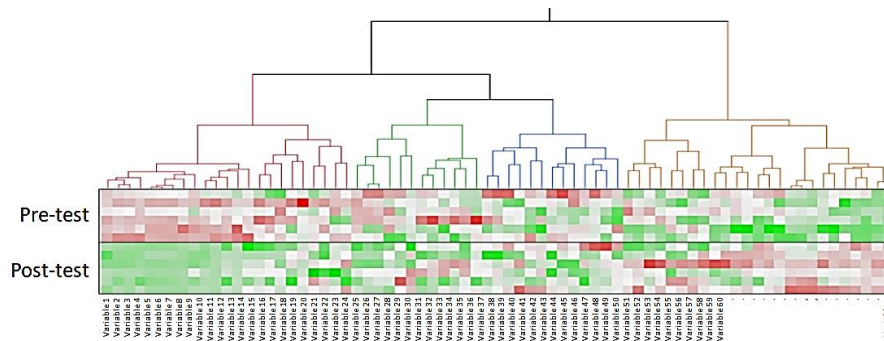
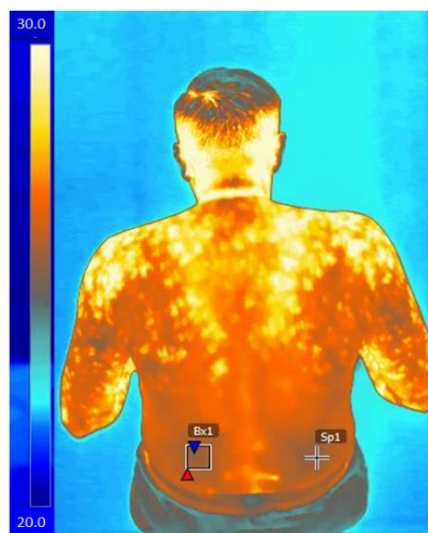


Figura 3. Mapa de calor en el que se representa la temperatura corporal localizada del torso de una persona. A la izquierda de la imagen se muestra la escala de temperatura en °C, en donde se puede observar la degradación de los colores; a medida que el color se aclara, la temperatura es mayor, y viceversa.



Los gráficos de violín permiten observar no solamente la mediana de los datos, sino también la forma de la distribución de los puntajes (Figura 4). Tal y como se recomienda en la actualidad también pueden mostrar los puntajes individuales^[228], lo que permite observar la verdadera tendencia de los mismos, lográndose observar puntajes influyentes, también conocidos como "outliers"^{[459][636]}. Los programas *BoxPlot Generator* e *Interactive Repeated Experiments Dotplot* descritos más adelante en este libro, permiten construir gráficos de violín (Figura 5).

Figura 4. Gráfico de violín. La denominación lo de "violín" permite observar la forma de la distribución de los puntajes.

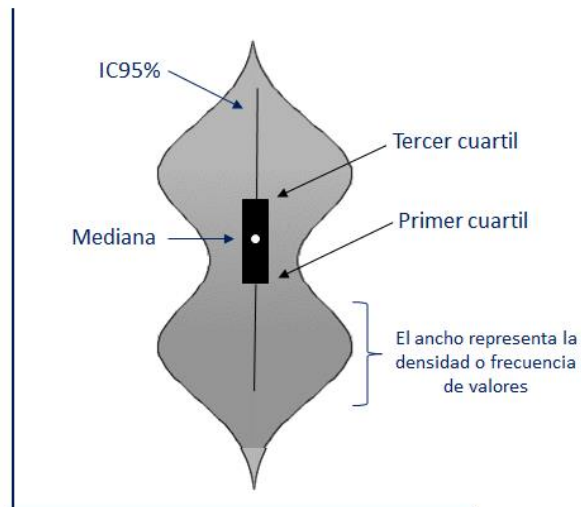
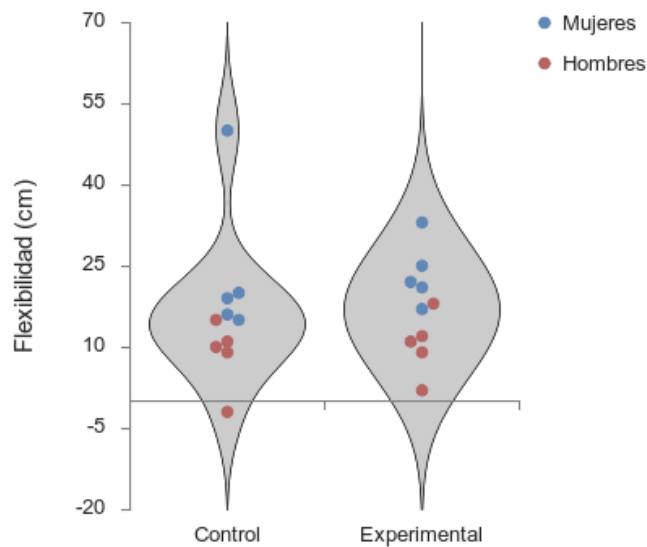


Figura 5. Gráfico de violín en donde se muestra forma de la distribución de los puntajes y posibles puntajes influyentes. En el violín del grupo control se observa que una de las mujeres (i.e., círculo color azul) muestra un comportamiento atípico (i.e., un puntaje alto en la flexibilidad) con respecto a los demás participantes.



Otra forma contemporánea para la presentación y especialmente la difusión de los hallazgos de investigación con las llamadas infografías, las cuales permiten, a través de un proceso creativo y artístico, resumir de forma gráfica todo un proceso de investigación y no solamente los datos estadísticos^[644]. Las infografías son composiciones de información, que contienen texto e imágenes, y que proporcionan un mensaje central. Todas las figuras se deben crear tomando en consideración aspectos artísticos y funcionales, como, por ejemplo, que las imágenes puedan ser vistas y comprendidas por personas con diferente nivel educativo, que los colores permitan la comprensión del mensaje incluso para personas con condiciones especiales (e.g., daltonismo)^[744]. En las Ciencias del Movimiento Humano, cada vez son más comunes las infografías; por ejemplo, debido a la emergencia por Covid-19, múltiples organizaciones y grupos de investigación realizaron esfuerzos para brindar información gráfica y gratuita acerca de los métodos para la detección del virus en los atletas (e.g., ^[594]).

En síntesis, las nuevas recomendaciones para el resumen y presentación de los hallazgos de las investigaciones en las Ciencias del Movimiento Humano requieren de un proceso creativo que permita mostrar con claridad, transparencia y precisión los datos individuales y grupales que se examinan. Existen diversos recursos electrónicos gratuitos para generar gráficos acordes al nivel de medición de la variable de interés y su distribución estadística. Gráficos de mayor calidad proporcionarán al lector experto y novato con la información adecuada para la toma de decisiones, y fomentará un reporte de investigación con menor sesgo, lo cual es valorado positivamente por los editores y revisores expertos de revistas especializadas.

2.9.3. Curva ROC

La curva "Característica Operativa del Receptor", o del inglés "Receiver Operator Characteristic" (ROC), representa gráficamente la sensibilidad frente a la especificidad^[450]. La curva ROC se compone de la proporción de casos (i.e., personas) clasificados como verdaderos positivos (eje "y") y de casos considerados como falsos positivos (eje "x"). Estos casos se pueden identificar en la llamada matriz de confusión, que es un sistema de clasificación (Tabla 1).

Tabla 1. Matriz de confusión mediante la cual se clasifica cada caso o persona de acuerdo a una característica observada.

		Valor real observado en los datos recolectados	
		Positivos	Negativos
Valor Predicho	Positivos	Verdaderos positivos (VP)	Falsos positivos (FP)
	Negativos	Falsos negativos (FN)	Verdaderos negativos (VN)

Los casos verdaderos positivos (VP) son aquellos valores predichos como positivos en el modelo de clasificación y que efectivamente corresponden a un VP en los datos recolectados. Los casos verdaderos negativos (VN) son aquellos valores predichos como negativos en el modelo y que corresponden a un valor negativo en los datos recolectados. Los casos falsos negativos (FN) son aquellos valores predichos en el modelo como negativos de forma incorrecta, ya que en los datos recolectados corresponden a un valor positivo. Finalmente, los valores falsos positivos (FP) son aquellos predichos como positivos en el modelo, pero que en los datos recolectados corresponden a valores negativos^[257].

A partir de los valores de matriz de confusión se calcula la precisión ($[(VP+VN)/(VP+FP+FN+VN)]$), la sensibilidad ($VP/(VP+FN)$), y la especificidad ($VN/(VN+FP)$). Los verdaderos positivos son equivalentes a la "sensibilidad" y los falsos positivos representan el valor "1-especificidad"^[257] (Figura 1).

La curva ROC se puede usar para generar estadísticos que resumen el rendimiento del clasificador, como, por ejemplo, el área bajo la curva (AUC, por sus siglas en inglés). Los valores recomendados para interpretar la curva ROC son de 0.5 (equivalente a lanzar una moneda), de 0.5 a 0.6 como una prueba mala, de 0.6 a 0.75 como una prueba regular, de 0.75 a 0.9 como una buena prueba, de 0.9 a 0.97 como una muy buena prueba, de 0.97 a 1 como una prueba excelente.

En la figura 2 se presenta un ejemplo hipotético de una curva ROC por medio de la cual se estudia el valor diagnóstico de la prueba de consumo de oxígeno (VO_2) para detectar trastornos del ritmo cardíaco. A las personas que realizaron la prueba de VO_2 se les observó durante un período de tiempo y se determinó si desarrollaban algún trastorno en el ritmo cardíaco. Así, la curva ROC de color verde indica que una alta cantidad de casos fueron clasificados como verdaderos positivos; es decir, hubo una excelente clasificación, lo que indica que el valor pronóstico de la prueba es excelente. Si los resultados hubieran indicado una cantidad moderada de casos correctamente clasificados, se observaría en color rojo en la curva ROC; sin embargo, si la clasificación hubiera sido mala, se vería en color azul en la curva ROC, lo cual indicaría que habría muchos casos clasificados incorrectamente, aumentando la cantidad de falsos positivos.

Figura 1. Curva Característica Operativa del Receptor (ROC, por sus siglas en inglés). En la matriz de confusión –utilizada para detectar verdaderos positivos y falsos positivos–, los valores sobre la diagonal representan una mejor clasificación de los datos y los valores bajo la diagonal representan una peor clasificación, que daría lugar a más casos de falsos positivos.

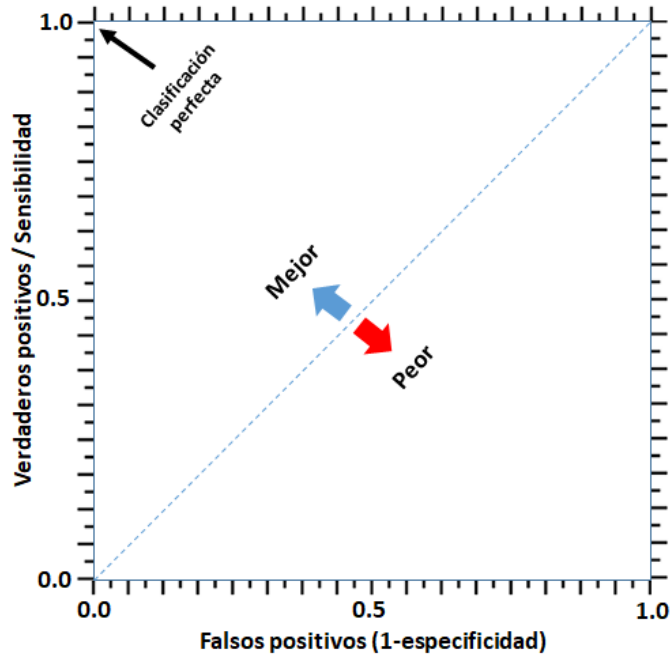
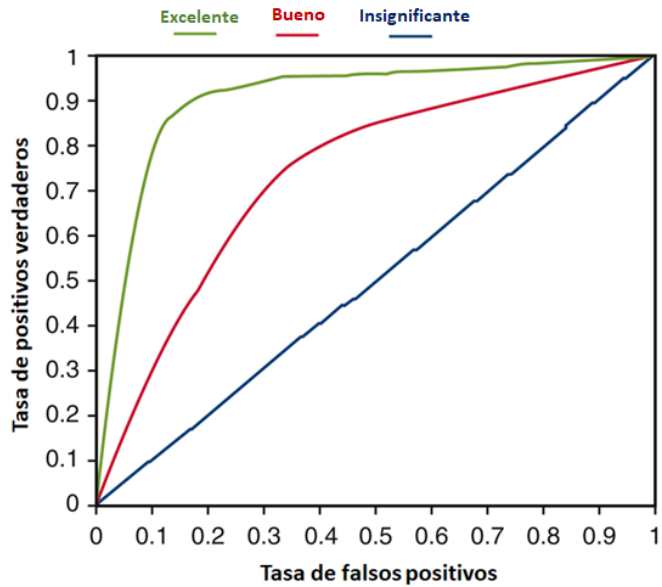


Figura 2. Curva Característica Operativa del Receptor (ROC, por sus siglas en inglés). En la matriz de confusión –utilizada para detectar verdaderos positivos y falsos positivos–, los valores sobre la diagonal representan una mejor clasificación de los datos y los valores bajo la diagonal representan una peor clasificación, que daría lugar a más casos de falsos positivos.



Los gráficos ROC se utilizan en las Ciencias del Movimiento Humano^{[11][269]}. Por ejemplo, se sabe que la capacidad aeróbica predice fuertemente la salud cardiovascular. Por ello, se realizó un estudio para comparar la predicción a 10 años de la mortalidad por todas las causas utilizando un enfoque de regresión logística y de machine learning en una cohorte de pacientes (n = 34212) que se sometieron a una prueba de esfuerzo^[11]. Del total de participantes, 3921 fallecieron a los 10 años. La sensibilidad y la especificidad para predecir la mortalidad con la regresión logística fue de 44.9% y 93.4%, respectivamente; mientras que con aprendizaje de máquinas fue de 87.4% y 97.2%, respectivamente. Estas clasificaciones se realizan con la matriz de confusión y se pueden graficar como curvas ROC.

En síntesis, las curvas ROC permiten visualizar gráficamente la clasificación de personas (i.e., casos) y obtener indicadores cuantitativos de sensibilidad y especificidad. Las curvas ROC pueden crearse utilizando hojas de cálculo como Excel y por medio de paquetes estadísticos como *IBM SPSS*, *Jamovi*, *R* y *OpenEpi*, entre otros. A la vez, se encuentran tutoriales gratuitos en YouTube que pueden ser de gran ayuda.

2.10. Herramientas para la investigación

2.10.1. Programas informáticos gratuitos

Un equipo de computación está compuesto principalmente por dos elementos; el "hardware" o componentes físicos (e.g., tarjeta madre, pantalla, procesadores, teclado), y el "software" o programas que lo hacen funcionar. Los programas informáticos se diseñan con instrucciones codificadas en algún lenguaje de computación que le permiten al usuario reducir el tiempo invertido en los procesos manuales de cálculo, así como a disminuir el error humano. En estadística, sin embargo, algunos programas de cómputo poseen un elevado costo, lo cual los hace inaccesibles para muchos estudiantes e investigadores.

Afortunadamente, el avance científico y tecnológico, así como el altruismo de muchos individuos y organizaciones dedicadas a la diseminación de la "ciencia abierta", ha permitido el diseño y acceso a múltiples programas gratuitos. Entre la amplia variedad de programas se encuentran los que realizan análisis estadísticos y también permiten la graficación de los resultados. En esta sección, se presentan algunos de los programas gratuitos que pueden servir como herramientas para el análisis de datos e información con fines de investigación (Tabla 1). La lista no es exhaustiva, y los autores de la presente obra han utilizado algunos de ellos, por lo cual los recomiendan como una opción en los procesos de investigación. Cabe destacar que ninguno de los autores posee intereses comerciales o reciben pagos o regalías de las empresas desarrolladoras de los programas mencionados en esta sección, por lo que su evaluación es independiente, libre de sesgo, y exclusivamente con fines pedagógicos.

Existen sitios web y programas para análisis estadístico de datos gratuitos. Algunos pueden ser de cuestionable calidad; mientras que otros son sugeridos y ampliamente utilizados por la comunidad internacional. Por ejemplo, uno de los programas gratuitos más difundidos y utilizados actualmente es *R*^[586]. Este es un programa de código libre, lo que significa que cualquier programador puede compartir sus códigos, "scripts", librerías o sintaxis, las cuales contienen las instrucciones para realizar los análisis estadísticos y la graficación requerida. El programa *R* posee tutoriales en YouTube y existen múltiples foros de discusión y evacuación de dudas. El programa *JASP*^[375] es otro software completo para análisis de datos frecuentistas y bayesianos. Posee un diseño de hoja de cálculo y una interfaz intuitiva de arrastrar y soltar. Una característica útil es su compatibilidad con el formato de la American Psychological Association (APA), ya que permite copiar y pegar gráficos y tablas directamente en procesadores de texto como Microsoft Word®. El programa *Jamovi*^[527], es conocido como una hoja de cálculo estadística de "tercera generación", que se basa en el lenguaje estadístico *R* y se complementa con un libro de texto también gratuito. El libro abarca temas de estadística descriptiva, presentación gráfica, teoría de probabilidad, muestreo, estimación y prueba de hipótesis nula. También cubre los temas de análisis de tablas de contingencia, correlación, pruebas t-student, regresión, ANOVA, análisis factorial y estadísticas bayesianas. Otro programa de uso general es llamado *Zelig*^[168], el cual es de código abierto y se basa en el lenguaje de programación *R*.

Tabla 1. Programas de acceso gratuito para análisis estadístico general y específico y graficación de resultados.

Nombre del software	Dirección Web	Características principales
BoxPlot Generator	http://shiny.chemgrid.org/boxplotr/	<ul style="list-style-type: none"> • Elabora gráficos de cajas ("box-plots") para datos individuales y grupales. • Se pueden cortar y pegar datos de una hoja de Excel.
CiteSpace	http://cluster.cis.drexel.edu/~cchen/citespace/	<ul style="list-style-type: none"> • Elabora gráficos para la visualización de patrones y tendencias en la literatura científica.
CitNetExplorer	www.citnetexplorer.nl	<ul style="list-style-type: none"> • Software para visualizar y analizar redes de citas de publicaciones científicas. • La herramienta permite importar redes de citas directamente desde la base de datos de Web of Science. • Las redes de citas se pueden explorar de forma interactiva, por ejemplo, profundizando en una red e identificando grupos de publicaciones estrechamente relacionadas.
Develve	https://develve.net/	<ul style="list-style-type: none"> • Develve es un software diseñado para ayudar a los usuarios a generar análisis estadísticos simples. • El programa cuenta con múltiples recursos estándar, como pruebas Chi², ANOVA unidireccional y modelos de regresión lineal. • Cuenta con una licencia gratuita de un año, período después del cual se puede comprar una licencia a un bajo precio.
Free Statistics Software	www.wessa.net	<ul style="list-style-type: none"> • Página con un grupo de diversas calculadoras estadísticas. • Incluye análisis estadísticos descriptivos, exploratorios, inferenciales univariados, multivariados y graficación.
		Continúa...

Tabla 1. Continuación.

Nombre del software	Dirección Web	Características principales
<i>G*Power</i>	www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower	<ul style="list-style-type: none"> • Software para calcular la potencia estadística. • Calcular la potencia para varias pruebas estadísticas (e.g., t-student, ANOVA, χ^2). • El usuario debe saber 4 de 5 variables: a) número de grupos, b) número de observaciones, c) tamaño del efecto, d) nivel de significancia (α) y e) potencia ($1-\beta$). • El programa posee una herramienta integrada para determinar el tamaño del efecto si no se puede estimar a partir de la literatura previa o si no se puede calcular fácilmente.
<i>Interactive Repeated Experiments Dotplot</i>	http://statistika.mfub.bg.ac.rs/interactive-repeated-experiments-dotplot	<ul style="list-style-type: none"> • Elabora gráficos de cajas ("box-plots"), de puntos y de violín. • Se pueden crear los gráficos añadiendo los datos en la página o importándolos desde una hoja CSV. • Muestra las estadísticas asociadas a los gráficos y se pueden construir subgrupos con los datos.
<i>GNU PSPP</i>	www.gnu.org/software/pspp/get.html	<ul style="list-style-type: none"> • Es una versión gratuita de software para análisis estadístico. • Es una aplicación estable y confiable que permite realizar análisis estadísticos descriptivos, pruebas t-student, ANOVA, regresión lineal y logística, medidas de asociación, análisis de conglomerados, análisis factorial y de confiabilidad y pruebas no paramétricas y más. • El programa posee manuales en línea que facilitan su uso e interpretación de los datos.
<i>Jamovi</i>	www.jamovi.org	<ul style="list-style-type: none"> • Es una hoja de cálculo estadística de "tercera generación". • Se basa en el lenguaje estadístico R.
		Continúa...

Tabla 1. Continuación.

Nombre del software	Dirección Web	Características principales
JASP	https://jasp-stats.org/	<ul style="list-style-type: none"> • Paquete estadístico completo para estadística básica y avanzada. • Ofrece procedimientos de análisis estándar en su forma clásica y bayesiana. • El programa puede leer el formato de archivos provenientes de otros paquetes estadísticos: .sav, .txt, .csv, .ods, .tsv, .dta, .por, .sas7bdat, .sas7bcat.
MetaXL	www.epigear.com/index_files/metaxl.html	<ul style="list-style-type: none"> • Complemento de Microsoft Excel para meta análisis. • Compatible con todos los principales métodos de meta análisis. • Incluye modelos de efectos de calidad y heterogeneidad de varianza inversa (IVHet). • Se pueden realizar meta análisis de red. • Construye gráficos Doi y calcula el índice LFK para estudiar sesgos.
OpenEpi	www.openepi.com/Menu/OE_Menu.htm	<ul style="list-style-type: none"> • Disponible en idioma español. • Produce estadísticas para casos y medidas en estudios descriptivos y analíticos, análisis estratificado con límites de confianza exactos, análisis de datos apareados y de personas-tiempo, tamaño de la muestra y cálculos de potencia, números aleatorios, sensibilidad, especificidad y otras estadísticas de evaluación, tablas F x C, χ^2 para dosis-respuesta.
OpenMEE	www.cebm.brown.edu/openmee	<ul style="list-style-type: none"> • Proporciona métodos avanzados de meta análisis y meta regresión para sintetizar datos continuos y categóricos, incluida la meta regresión con múltiples covariables y sus interacciones, análisis filogenéticos e imputación simple de datos faltantes. • Permite crear forest y funnel plots. • Recomendable como complemento de MetaXL para meta análisis de datos agregados donde se requiera realizar meta regresiones.
		Continúa...

Tabla 1. Continuación.

Nombre del software	Dirección Web	Características principales
R	www.r-project.org	<ul style="list-style-type: none"> • Pertenece a la categoría de software libre. • Entorno y lenguaje de programación con un enfoque al análisis estadístico. • Popular en las áreas de machine learning y minería de datos. • Capacidad para cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo y graficación. • Muy versátil y en múltiples sitios de internet se obtienen tutoriales y bibliotecas para análisis estadísticos específicos.
SAS® OnDemand for Academics	https://welcome.oda.sas.com/login	<ul style="list-style-type: none"> • Es una versión gratuita en línea del paquete estadístico SAS®, en la que hay que suscribirse gratuitamente para poder acceder a los programas. • Posee capacidades para programación de análisis estadísticos comunes.
Social Sciences Statistics	www.socscistatistics.com	<ul style="list-style-type: none"> • El sitio ofrece calculadoras estadísticas. • Se divide en 3 secciones: <ul style="list-style-type: none"> ○ Calculadoras estadísticas: se ocupa principalmente de pruebas de hipótesis. ○ Calculadoras de valor p: permiten derivar valores p de Z, t, χ^2 y r de Pearson. ○ Estadísticas descriptivas: promedios, varianza, calculadora de desviación estándar, y graficadores de histogramas y gráficos de barras.
		Continúa...

Tabla 1. Continuación.

Nombre del software	Dirección Web	Características principales
SOFA	www.sofastatistics.com/home.php	<ul style="list-style-type: none"> • SOFA es un programa de informes, análisis y estadísticas fácil de usar. • El programa es gratuito, con énfasis en la facilidad de uso y en aprender sobre la marcha. • Su graficación es amigable, e incluye gráficos de: barras simples (frecuencia o medias), barras agrupadas (frecuencia o media), circulares, de una y varias líneas (frecuencia o medias), de área (frecuencia o medias), histogramas, de dispersión y diagramas de cajas (box plots) y de bigotes (whisker plots). • Los análisis estadísticos incluyen porcentajes de filas y columnas, con la capacidad de anidar variables, mediana, cuartil inferior, cuartil superior, desviación estándar, χ^2 de Pearson con tablas de contingencia, prueba t de muestras independientes, prueba t de muestras pareadas, ANOVA de una vía, prueba U de Mann Whitney, prueba rangos de Wilcoxon, prueba H de Kruskal Wallis, correlación de Pearson y correlación de Spearman.
Statistical Lab	www.statistiklabor.de/en	<ul style="list-style-type: none"> • El programa informático Statistical Lab es una caja de herramientas exploratoria e interactiva para el análisis estadístico y de visualización de datos para apoyar la educación en estadística como para proporcionar una herramienta para la simulación y solución de problemas estadísticos. • La interfaz gráfica de usuario está diseñada para hacer que las relaciones estadísticas complejas sean fáciles de entender. • Conecta y muestra marcos de datos, tablas de frecuencia, números aleatorios o matrices en una hoja de trabajo estadística fácil de usar que permite a los usuarios ejecutar cálculos, realizar análisis y realizar múltiples simulaciones y manipulaciones. • El programa se basa en R y es apoyado por el Gobierno de Alemania y la Freie Universität Berlin.
		Continúa...

Tabla 1. Continuación.

Nombre del software	Dirección Web	Características principales
VOSViewer	www.vosviewer.com	<ul style="list-style-type: none"> • Herramienta de software para construir y visualizar redes bibliométricas. • Las redes pueden incluir revistas, investigadores o publicaciones individuales. • Las redes pueden construirse sobre la base de relaciones de citación, acoplamiento bibliográfico, cocitación o coautoría. • Ofrece la funcionalidad de minería de texto que se puede utilizar para construir y visualizar redes de co-ocurrencia de términos importantes extraídos de un cuerpo de literatura científica.
WebPlotDigitazer	https://automeris.io/WebPlotDigitizer	<ul style="list-style-type: none"> • Programa para extraer datos a partir de gráficos. • Puede ser útil para meta análisis de datos cuando en los artículos publicados se cuenta con gráficos, pero no con los datos con los que se construyeron. • Se trata de un proceso en tres pasos: <ul style="list-style-type: none"> ○ Importa el gráfico desde un archivo. Acepta formatos como .gif, .png, .tiff, .jpeg y .bmp. ○ Define el sistema de coordenadas ○ Se digitaliza automática o manualmente. • Los valores de los datos se pueden guardar en formato CSV o se pueden copiar y pegar directamente en cualquier otra aplicación (e.g., Excel u otra similar).
WinBUGS	www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs	<ul style="list-style-type: none"> • Software estadístico para análisis bayesiano utilizando los llamados métodos Markov Chain Monte Carlo (MCMC). • Se basa en el proyecto llamado <i>Bayesian inference Using Gibbs Sampling</i> (BUGS). • Permite la graficación geométrica de los resultados de los meta análisis en red.
		Continúa...

Tabla 1. Continuación.

Nombre del software	Dirección Web	Características principales
Zelig	https://zeligproject.org/	<ul style="list-style-type: none">• Zelig es un programa estadístico de propósito general, gratuito, de código abierto y fácil de usar para estimar, interpretar y presentar resultados de cualquier método estadístico.• Zelig se basa en el lenguaje de programación R, con miles de paquetes de código abierto, pero con sintaxis de rango libre, diversos ejemplos y documentación escrita para diferentes audiencias, en los mismos tres comandos y documentación consistente para cada método.• Incluye muchos métodos específicos, basados en teorías de inferencia de probabilidad, frecuentista, bayesiana, bayesiana robusta, no paramétrica y de población y superpoblación.

También existen programas estadísticos más especializados. Por ejemplo, el programa *G*Power*^{[255][256]} se especializa en el cálculo del tamaño de la muestra para las investigaciones. Es un programa sumamente utilizado; sin embargo, requiere que el usuario ingrese información, para lo cual se requiere de conocimientos de la temática de estudio y de probabilidad estadística básica. Los programas *MetaXL*^[251], *OpenMEE*^[732], y *WinBUGS*^[206] permiten realizar diversos cálculos y graficación para meta análisis tradicional y meta análisis en red, y además se complementan entre sí.

MetaXL es una extensión que se utiliza con la hoja de cálculo de Microsoft Excel® con la que se pueden obtener los tamaños de efecto para meta análisis utilizando los tradicionales modelos de efectos fijos, aleatorios y el más novedoso y recomendable de heterogeneidad de varianza inversa (IVhet)^[213]. Por su parte, *OpenMEE* se utiliza para calcular tamaños de efecto de meta análisis, graficación, meta regresión y pruebas para el análisis de sesgo. *WinBUGS* permite construir gráficos geométricos para los meta análisis en red. El programa *WebPlotDigitizer*^[121] funciona para obtener datos continuos a partir de un gráfico o imagen publicada en algún documento.

Finalmente, existen sitios mucho más generales, los cuales ponen a disposición varios tipos de calculadoras y técnicas de análisis estadísticos más ampliamente utilizados. Por ejemplo, el programa *OpenEpi*^[202] proporciona análisis paramétricos y no paramétricos, así como calculadoras para el cálculo del tamaño de la muestra. Otros sitios web como *Free Statistics Software*^[748] y *Social Science Statistics*^[668] también proporcionan múltiples calculadoras para análisis estadísticos.

Entre los programas y sitios web para la graficación de datos se encuentran el *BoxPlot Generator*^[709] y el *Interactive Repeated Experiments Dotplot*^[746]. Estos programas permiten construir gráficos de caja (i.e., Box Plot), de puntos o de violín y su forma de presentación puede ser grupal o con datos individuales. En las Ciencias del Movimiento Humano se publican estudios que han utilizado ese programa pues permite observar las respuestas grupales e individuales ante un tratamiento experimental^[111].

Otro tipo de programas diseñados para servir como interfase gráfica para resumir información bibliográfica incluye el *CiteSpace*^[160], *CitNetExplorer*^[712], y *VOSviewer*^[713]. Por ejemplo, la principal fuente de datos de entrada de *CiteSpace* es *Web of Science (WoS)*. El usuario busca un tema de interés y descarga los resultados de la búsqueda, incluyendo los registros completos y las referencias. El programa también se puede utilizar para generar superposiciones de mapas geográficos visibles en Google Earth en función de las ubicaciones de los autores.

En los próximos años se espera que surjan nuevos programas gratuitos que incorporen la inteligencia artificial, lo cual indudablemente ayudará a los investigadores; sin embargo, las bases adecuadas de teoría estadística deben reforzarse para tomar las mejores decisiones y buscar la mayor calidad en el reporte de los datos.

2.11. *Errare humanum est*

El proceso de investigación científica no está exento de errores. En todas las fases de un estudio ocurren inconvenientes que pueden ser detectados mientras se realiza el estudio o pueden ser indetectables hasta que se publicó el manuscrito. Incluso, cuando un manuscrito enviado a una revista enfrenta un estricto proceso de revisión de pares, pueden ocurrir errores inadvertidos que solamente algún ojo crítico finalmente puede detectar. Por lo tanto, *errare humanum est*; que traducido del latín significa "errar es humano", y como corolario "*errare humanum est, sed perseverare diabolicum*", que significa "errar es humano, pero persistir es diabólico", se puede interpretar en la investigación en el sentido de que equivocarse es parte de la naturaleza humana; sin embargo, esto no puede ser un factor atenuante para repetir un error, sino un medio para aprender de la experiencia.

En la literatura científica abundan los errores; algunos son detectables y evidentes, aunque otros no. El error puede provenir de los investigadores o de las propias revistas. En el caso de los investigadores, el error puede ser conceptual, procedimental, del análisis de los datos^{[175][507]}, de la presentación de los resultados, de naturaleza ética académica, o del propio manuscrito enviado a revisión de pares académicos. Un estudio acerca de las tendencias con respecto a los errores metodológicos de las revisiones sistemáticas y meta análisis retractados indicaron que las razones más frecuentes para retirar un artículo de una revista fueron violaciones éticas académicas, errores metodológicos y problemas de redacción. Incluso, se llegó a determinar que la cantidad de violaciones éticas académicas fueron mucho mayores que los errores metodológicos o de problemas de redacción^[159].

El error atribuible a los investigadores puede ser posible durante el proceso de revisión de galeradas (i.e., "galley proofs"); es decir, la forma editada de manuscrito que sería finalmente publicada en la revista^{[329][374]}. Por su parte, los errores propios de la revista incluyen la asignación poco idónea de pares revisores a un manuscrito (i.e., incapaces de detectar errores ya que son ajenos al tema de estudio o que carecen de la formación especializada para evaluar un manuscrito), la pobre revisión estadística, y a procesos propios del montaje editorial del manuscrito, en donde algunas veces se "pierden" caracteres especiales (e.g., letras griegas, símbolos matemáticos).

En algunos artículos se omite información o se publica información errónea, y para reparar ese error, los autores solicitan al editor de la revista, incluir una nota aclaratoria que explique el error y su respectiva corrección. Por ejemplo, en un estudio se analizó la ingesta de líquidos de 12 ultramaratonistas durante 24 h y se publicaron estadísticas descriptivas en tablas y gráficos^[419]. Sin embargo, los autores, una vez publicado el estudio, encontraron un error y solicitaron la enmienda respectiva^[420]. Algunas veces, los errores tienen consecuencias importantes dependiendo del contexto de la investigación; es decir, publicar datos erróneos puede ser más perjudicial para pacientes con cáncer que para deportistas. En todo caso, el concepto más relevante es tener una integridad científica y humildad que permita reconocer el error y la valentía y ética para corregirlo.

Uno de los errores más fatales consiste en violar algún principio ético, y lo que se recomienda es retractar o retirar el manuscrito, ya sea durante el proceso de revisión o cuando ya ha sido publicado. Por ejemplo, en la revista *Medicine & Science in Sports & Exercise*, la revista oficial del American College of Sports Medicine, que goza de una excelente reputación (i.e., generalmente en el "top 10" del ranking de las revistas científicas en las Ciencias del Movimiento Humano), se publicó un manuscrito que pasó por todos los filtros de calidad^[51]. Sin embargo, dos autores detectaron posibles irregularidades con los datos, por lo que el editor en jefe publicó que se iba a investigar el asunto^[483], y posterior a una revisión de expertos en estadística se llegó a la conclusión de que los datos eran "altamente irregulares"; por lo que el editor decidió retractar el artículo de la revista^[484]. Esa fue una dura decisión para los autores del estudio, y a la vez, una decisión acertada del editor de la revista, entre cuyas responsabilidades se encuentra vigilar la veracidad y calidad de los manuscritos publicados en la revista.

Otro de los errores fatales consiste en utilizar material previamente publicado sin el debido permiso, violando las leyes de propiedad intelectual^[114]. Existen más errores, muchos de los cuales son menos fatales, pero de considerable importancia, como no brindar el reconocimiento a los patrocinadores de la investigación^[318] o equivocarse en el orden de los autores o en sus atestados académicos^[411]. En estos casos, no es necesario retractar el artículo sino más bien publicar una fe de erratas o *corrigendum*, que se define como una corrección a algo que ya ha sido impreso.

Es importante aclarar que las discusiones académicas no representan una categoría dentro de los errores discutidos en esta sección. De hecho, las discusiones académicas son parte indispensable para el avance en la ciencia que permite tener investigadores utilizando argumentos para discutir ideas, lo cual es la esencia del pensamiento científico: cuestionar todo. Por ejemplo, se sabe que las pruebas clínicas de ejercicio cardiopulmonar pueden determinar las causas de la limitación del ejercicio. Un grupo de investigadores determinó la pendiente de la frecuencia cardíaca frente al VO_2 , llamado índice cronotrópico, el cual permite identificar el deterioro cardiovascular. Entonces, plantearon un estudio cuyo objetivo fue desarrollar una ecuación de referencia para el índice cronotrópico basada en un gran número de sujetos que se consideraba que tenían respuestas normales al ejercicio^[651]. Sin embargo, un lector del estudio realizó una crítica fundamentada^[769] y los autores del estudio original tuvieron la oportunidad de responderle; es decir, plantearon una respuesta académica^[652]. En esta situación, no hay ganadores ni perdedores, sino puntos de vista contrastantes que permiten a los lectores poseer más información para la toma de decisiones. Por supuesto, pueden existir acalorados debates pues el ego de algunos investigadores se puede ver ofendido, pero en general, la comunidad científica se beneficia de los debates académicos.

En síntesis, el error es parte de todo proceso de producción humana. Reconocer y corregir los errores forma parte de la integridad de quienes investigan y permite que la investigación sea creíble, transparente y rigurosa. En las Ciencias del Movimiento Humano, como en cualquier otra ciencia, equivocarse con dolo y no corregir los errores "involuntarios" son parte de las malas prácticas de investigación que atentan contra el conocimiento científico y que debilitan los esfuerzos para avanzar la ciencia; por lo tanto, deben ser castigadas con rigor por la comunidad internacional.

Referencias

- [1] Aadland E, et al. Multicollinear physical activity accelerometry data and associations to cardiometabolic health: Challenges, pitfalls, and potential solutions. *Int J Behav Nutr Phys Act.* 2019;16doi:10.1186/s12966-019-0836-z
- [2] Aaltonen S, et al. Motives for physical activity in older men and women: A twin study using accelerometer-measured physical activity. *Scand J Med Sci Sports.* 2020;30(8):1409-1422. doi:10.1111/sms.13673
- [3] Abt G, et al. Power, precision, and sample size estimation in sport and exercise science research. *J Sports Sci.* 2020;38(17):1933-1935. doi:10.1080/02640414.2020.1776002
- [4] Abt G, et al. Raising the bar in sports performance research. *J Sports Sci.* 2022;40(2):125-129. doi:10.1080/02640414.2021.2024334
- [5] Agnihotri K. The nocebo effect in current practice. *Can Fam Physician.* 2020;66(11):862-864.
- [6] Ahtainen JP, et al. Inter-individual variation in response to resistance training in cardiometabolic health indicators. *Scand J Med Sci Sports.* 2020;30(6):1040-1053. doi:10.1111/sms.13650
- [7] Aitchison B, et al. The experiences and perceived health benefits of individuals with a disability participating in sport: A systematic review and narrative synthesis. *Disability and Health Journal.* 2022;15(1)doi:10.1016/j.dhjo.2021.101164
- [8] Aitchison J. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological).* 1982;44(2):139-160.
- [9] Akay MF, et al. Development of Novel Maximal Oxygen Uptake Prediction Models for Turkish College Students Using Machine Learning and Exercise Data. Institute of Electrical and Electronics Engineers (IEEE); 2017:186-189.
- [10] Akhtar A. The flaws and human harms of animal experimentation. *Camb Q Healthc Ethics.* 2015;24(4):407-19. doi:10.1017/s0963180115000079
- [11] Al-Mallah MH, et al. Using Machine Learning to Define the Association between Cardiorespiratory Fitness and All-Cause Mortality (from the Henry Ford Exercise Testing Project). *American Journal of Cardiology.* 2017;120(11):2078-2084. doi:10.1016/j.amjcard.2017.08.029
- [12] Al Attar WSA, et al. The FIFA 11+ Kids Injury Prevention Program Reduces Injury Rates Among Male Children Soccer Players: A Clustered Randomized Controlled Trial. *Sports Health.* 2022;19417381221109224. doi:10.1177/19417381221109224
- [13] Alexander RJ. *Sweeteners: nutritive.* Eagan Press; 1998.
- [14] Alibhai SM, et al. A phase II RCT and economic analysis of three exercise delivery methods in men with prostate cancer on androgen deprivation therapy. *BMC Cancer.* 2015;15:312. doi:10.1186/s12885-015-1316-8
- [15] Alibhai SMH, et al. A phase II randomized controlled trial of three exercise delivery methods in men with prostate cancer on androgen deprivation therapy. *BMC Cancer.* 2019;19(1):2. doi:10.1186/s12885-018-5189-5
- [16] Allen JM, et al. Exercise Alters Gut Microbiota Composition and Function in Lean and Obese Humans. *Med Sci Sports Exerc.* 2018;50(4):747-757. doi:10.1249/mss.0000000000001495
- [17] Alliot O, et al. Do adolescents' experiences of the barriers to and facilitators of physical activity differ by socioeconomic position? A systematic review of qualitative evidence. *Obes Rev.* 2021;doi:https://doi.org/10.1111/obr.13374
- [18] Amar D, et al. Time trajectories in the transcriptomic response to exercise - a meta-analysis. *Nat Commun.* 2021;12(1):3471. doi:10.1038/s41467-021-23579-x
- [19] Aminian K, et al. Level, downhill and uphill walking identification using neural networks. *Electronics Letters. Institution of Engineering and Technology;* 1993;29(17):1563-1565. doi:10.1049/el:19931042
- [20] Aminian K, et al. Incline, speed, and distance assessment during unconstrained walking. *Med Sci Sports Exerc.* 1995;27(2):226-34.
- [21] Amrhein V, et al. Scientists rise up against statistical significance. *Nature.* 2019;567(7748):305-307. doi:10.1038/d41586-019-00857-9
- [22] Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecology.* 2001;26(1):32-46. doi:https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x
- [23] Anderson MJ. PERMANOVA. Permutational multivariate analysis of variance. A computer program. http://img2.timg.co.il/forums/1_124959686.pdf
- [24] Anderson MJ. Permutational Multivariate Analysis of Variance (PERMANOVA). *Wiley StatsRef: Statistics Reference Online.* 2017:1-15.
- [25] Anguera-Argilaga T. Observational typology. *Quality and Quantity.* 1979;13(6):449-484.
- [26] Anguera MT. La observación. In: Moreno Rosset C, ed. *Evaluación psicológica Concepto, proceso y aplicación en las áreas del desarrollo y de la inteligencia.* Sanz y Torres; 2003. vol. 271-308.
- [27] Anguera MT. Observational methods (general). In: Fernández-Ballesteros R, ed. *Encyclopedia of psychological assessment.* Sage; 2003:632-637.
- [28] Anguera MT, et al. Integración de elementos cualitativos y cuantitativos en metodología observacional. *Ámbitos Revista Internacional de Comunicación.* 2020;(49):49-70.
- [29] Anguera MT, et al. Indirect Observation in Everyday Contexts: Concepts and Methodological Guidelines within a Mixed Methods Framework. *Frontiers in psychology.* 2018;9doi:10.3389/fpsyg.2018.00013
- [30] Antero J, et al. Female and male US Olympic athletes live 5 years longer than their general population counterparts: a study of 8124 former US Olympians. *British journal of sports medicine.* 2021;55(4):206. doi:10.1136/bjsports-2019-101696
- [31] Araujo JS, et al. Effect of Training and Detraining in the Components of Physical Fitness in People Living With HIV/AIDS. *Front Physiol.* 2021;12:586753. doi:10.3389/fphys.2021.586753
- [32] Araya-Ramírez F, et al. Who is likely to benefit from phase II cardiac rehabilitation? *J Cardiopulm Rehabil Prev.* 2010;30(2):93-100. doi:10.1097/HCR.0b013e3181d0c17f
- [33] Araya-Ramírez F, et al. Improved Walk Test Performance and Blood Pressure Responses in Men and Women Completing Cardiac Rehabilitation: Implications Regarding Exercise Trainability. *American Journal of Lifestyle Medicine.* 2021;1559827621995129. doi:10.1177/1559827621995129
- [34] Arden CL, et al. Implementing the 27 PRISMA 2020 Statement items for systematic reviews in the sport and exercise medicine, musculoskeletal rehabilitation and sports science fields: the PERSiST (implementing Prisma in Exercise, Rehabilitation, Sport medicine and Sports science) guidance. *British journal of sports medicine.* 2021;doi:10.1136/bjsports-2021-103987
- [35] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology.* 2005;8(1):19-32. doi:10.1080/1364557032000119616
- [36] Aromataris E, et al. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *JBI Evidence Implementation.* 2015;13(3):132-140. doi:10.1097/xeb.0000000000000055
- [37] Arrieta-Leandro MC, et al. Effect of different high intensity interval training (HIIT) modalities on glycaemic control, aerobic resistance and body composition in people with type 2 diabetes. A meta-analysis. 2022. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021282723
- [38] Astorino TA, Schubert MM. Individual responses to completion of short-term and chronic interval training: a retrospective study. *PLoS One.* 2014;9(5):e97638. doi:10.1371/journal.pone.0097638
- [39] Atkinson G, Batterham AM. True and false interindividual differences in the physiological response to an intervention. *Exp Physiol.* 2015;100(6):577-588. doi:https://doi.org/10.1113/EP085070
- [40] Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26(4):217-38. doi:10.2165/00007256-199826040-00002

- [41] Aum S, Choe S. srBERT: automatic article classification model for systematic review using BERT. *Systematic Reviews*. 2021;10(1):285. doi:10.1186/s13643-021-01763-w
- [42] Aune D, et al. Physical activity and the risk of sudden cardiac death: a systematic review and meta-analysis of prospective studies. *BMC Cardiovasc Disord*. 2020;20(1):318. doi:10.1186/s12872-020-01531-z
- [43] Babey SH, et al. How can schools help youth increase physical activity? An economic analysis comparing school-based programs. *Prev Med*. 2014;69:S55-S60. doi:https://doi.org/10.1016/j.ypmed.2014.10.013
- [44] Bachmann JM, et al. Cardiorespiratory Fitness in Middle Age and Health Care Costs in Later Life. *Journal of the American College of Cardiology*. 2015;66(17):1876-1885. doi:https://doi.org/10.1016/j.jacc.2015.08.030
- [45] Bacon AP, et al. VO2max trainability and high intensity interval training in humans: a meta-analysis. *PLoS One*. 2013;8(9):e73182. doi:10.1371/journal.pone.0073182
- [46] Baethge C, et al. SANRA—a scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review*. 2019;4(1):5. doi:10.1186/s41073-019-0064-8
- [47] Bagg MK, et al. Comparing interventions with network meta-analysis. *J Physiother*. 2018;64(2):128-132. doi:10.1016/j.jphys.2018.02.014
- [48] Bailey J. Does the stress of laboratory life and experimentation on animals adversely affect research data? A critical review. *Altern Lab Anim*. 2018;46(5):291-305. doi:10.1177/026119291804600501
- [49] Bakeman R. Untangling streams of behavior: Sequential analysis of observation data. In: Sackett GP, ed. *Observing Behavior: Data collection and analysis methods*. University of Park Press.; 1978:63-78.
- [50] Baker WL, et al. Understanding heterogeneity in meta-analysis: the role of meta-regression. *Int J Clin Pract*. 2009;63(10):1426-34. doi:10.1111/j.1742-1241.2009.02168.x
- [51] Barbalho M, et al. Evidence for an Upper Threshold for Resistance Training Volume in Trained Women. *Med Sci Sports Exerc*. 2019;51(3):515-522. doi:10.1249/mss.0000000000001818
- [52] Barbosa A, et al. Physical Activity and Academic Achievement: An Umbrella Review. *Int J Environ Res Public Health*. 2020;17(16):doi:10.3390/ijerph17165972
- [53] Barbosa Mendes A, et al. Network Meta-Analysis for Single-Case Design Studies: An Illustration. *Evaluation & the Health Professions*. 2022;01632787211067532. doi:10.1177/01632787211067532
- [54] Barchard KA. Examining the reliability of interval level data using root mean square differences and concordance correlation coefficients. *Psychological Methods*. 2012;17(2):294-308. doi:10.1037/a0023351
- [55] Bardid F, et al. Investigating the construct of motor competence in middle childhood using the BOT-2 Short Form: An item response theory perspective. *Scand J Med Sci Sports*. 2019;29(12):1980-1987. doi:10.1111/sms.13527
- [56] Barker J, et al. *Single-case research methods in sport and exercise psychology*. Routledge; 2011.
- [57] Barker M, Rayens W. Partial least squares for discrimination. *Journal of Chemometrics*. 2003;17(3):166-173. doi:https://doi.org/10.1002/cem.785
- [58] Barnhart HX, et al. An Overview on Assessing Agreement with Continuous Measurements. *Journal of Biopharmaceutical Statistics*. 2007;17(4):529-569. doi:10.1080/10543400701376480
- [59] Barzideh A, et al. Effect of reactive balance training on physical fitness poststroke: study protocol for a randomised non-inferiority trial. *BMJ Open*. 2020;10(6):e035740. doi:10.1136/bmjopen-2019-035740
- [60] Batrakoulis A, et al. Hybrid neuromuscular training promotes musculoskeletal adaptations in inactive overweight and obese women: A training-detraining randomized controlled trial. *J Sports Sci*. 2021;39(5):503-512. doi:10.1080/02640414.2020.1830543
- [61] Batterham AM. Is the intervention as good as (or not substantially worse than) a comparator? *Exp Physiol*. 2022;107:199-200. doi:https://doi.org/10.1113/EP090321
- [62] Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. *Int J Sports Physiol Perform*. 2006;1(1):50-7.
- [63] Bazeley P. Integrative Analysis Strategies for Mixed Data Sources. *American Behavioral Scientist*. 2012;56(6):814-828. doi:10.1177/0002764211426330
- [64] Bazeley P. Mixed or merged? Integration as the real challenge for mixed methods. *Qualitative Research in Organizations and Management: An International Journal*. 2016;11(3):189-194. doi:10.1108/QROM-04-2016-1373
- [65] Beato M, et al. Validity and reliability of a flywheel squat test in sport. *J Sports Sci*. 2021;39(5):482-488. doi:10.1080/02640414.2020.1827530
- [66] Beedie C, et al. Incorporating methods and findings from neuroscience to better understand placebo and nocebo effects in sport. *Eur J Sport Sci*. 2020;20(3):313-325. doi:10.1080/17461391.2019.1675765
- [67] Bell L, et al. Overreaching and overtraining in strength sports and resistance training: A scoping review. *J Sports Sci*. 2020;38(16):1897-1912. doi:10.1080/02640414.2020.1763077
- [68] Benham JL, et al. Significant Dose-Response between Exercise Adherence and Hemoglobin A1c Change. *Med Sci Sports Exerc*. 2020;doi:10.1249/MSS.0000000000002339
- [69] Benjamin DJ, Berger JO. Three Recommendations for Improving the Use of p-Values. *The American Statistician*. 2019;73(sup1):186-191. doi:10.1080/00031305.2018.1543135
- [70] Benson LC, et al. Evaluating Methods for Imputing Missing Data from Longitudinal Monitoring of Athlete Workload. *Journal of Sports Science and Medicine*. 2021;20(2):187-195. doi:https://doi.org/10.52082/jssm.2021.188
- [71] Bergh U. The influence of body mass in cross-country skiing. *Med Sci Sports Exerc*. 1987;19(4):324-331.
- [72] Berkel AEM, et al. Effects of Community-based Exercise Prehabilitation for Patients Scheduled for Colorectal Surgery With High Risk for Postoperative Complications: Results of a Randomized Clinical Trial. *Annals of Surgery*. 2022;275(2):e299-e306. doi:10.1097/sla.0000000000004702
- [73] Beynon C, et al. Risk factors for gambling and problem gambling: a protocol for a rapid umbrella review of systematic reviews and meta-analyses. *Systematic Reviews*. 2020;9(1):198. doi:10.1186/s13643-020-01455-x
- [74] Bezerra TA, et al. 24-hour movement behaviour and executive function in preschoolers: A compositional and isotemporal reallocation analysis. *Eur J Sport Sci*. 2021;21(7):1064-1072. doi:10.1080/17461391.2020.1795274
- [75] Biddle GJH, et al. Associations of Physical Behaviours and Behavioural Reallocations with Markers of Metabolic Health: A Compositional Data Analysis. *Int J Environ Res Public Health*. 2018;15(10):2280. doi:10.3390/ijerph15102280
- [76] Biddle GJH, et al. Reply to Mekary, R.A.; Ding, E.L. Isotemporal Substitution as the Gold Standard Model for Physical Activity Epidemiology: Why It Is the Most Appropriate for Activity Time Research. *Int. J. Environ. Res. Public Health* 2019, 16, 797. *Int J Environ Res Public Health*. 2019;16(16):2885. doi:10.3390/ijerph16162885
- [77] Biduski GM, et al. Effects of 6-weeks of detraining on functional capacity and rapid torque production in older women. *J Bodyw Mov Ther*. 2022;29:167-173. doi:10.1016/j.jbmt.2021.09.033
- [78] Bishop D. An Applied Research Model for the Sport Sciences. *Sports Med*. 2008;38(3):253-263. doi:10.2165/00007256-200838303-00005
- [79] Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ (Clinical research ed)*. 1996;312(7040):1215-1218. doi:10.1136/bmj.312.7040.1215
- [80] Blagrove RC, et al. Inter-limb strength asymmetry in adolescent distance runners: Test-retest reliability and relationships with performance and running economy. *J Sports Sci*. 2021;39(3):312-321. doi:10.1080/02640414.2020.1820183
- [81] Blagrove RC, et al. Test-retest reliability of physiological parameters in elite junior distance runners following allometric scaling. *Eur J Sport Sci*. 2017;17(10):1231-1240. doi:10.1080/17461391.2017.1364301
- [82] Blagrove RC, et al. Quantification of aerobic determinants of performance in post-pubertal adolescent middle-distance runners. *European Journal of Applied Physiology*. 2019;119(8):1865-1874. doi:10.1007/s00421-019-04175-w

- [83] Blagrove RC, et al. Effects of Strength Training on Postpubertal Adolescent Distance Runners. *Med Sci Sports Exerc.* 2018;50(6):1224-1232. doi:10.1249/mss.0000000000001543
- [84] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet.* 1986;327(8476):307-310.
- [85] Bland JM, Altman DG. Statistics Notes: Transforming data. *BMJ (Clinical research ed).* 1996;312(7033):770. doi:10.1136/bmj.312.7033.770
- [86] Bleakley C, Smoliga JM. Validating new discoveries in sports medicine: we need FAIR play beyond p values. *British journal of sports medicine.* 2020;54(21):1239-1240. doi:10.1136/bjsports-2019-101797
- [87] Blume JD, et al. An Introduction to Second-Generation p-Values. *The American Statistician.* 2019;73(sup1):157-167. doi:10.1080/00031305.2018.1537893
- [88] Boccomino HL, et al. Log-Transformed Electromyography Amplitude Power Output Relationship: Nondominant vs. Dominant Limb. *Journal of strength and conditioning research.* 2022;36(3):851-856. doi:10.1519/jsc.0000000000004222
- [89] Bonafiglia JT, et al. An appraisal of the SD_R as an estimate of true individual differences in training responsiveness in parallel-arm exercise randomized controlled trials. *Physiological reports.* 2019;7(14):e14163. doi:10.14814/phy2.14163
- [90] Bonafiglia JT, et al. Moving beyond threshold-based dichotomous classification to improve the accuracy in classifying non-responders. *Physiological reports.* 2018;6(22):e13928. doi:10.14814/phy2.13928
- [91] Bonafiglia JT, et al. Inter-Individual Variability in the Adaptive Responses to Endurance and Sprint Interval Training: A Randomized Crossover Study. *PLoS One.* 2016;11(12):e0167790. doi:10.1371/journal.pone.0167790
- [92] Bongiovanni T, et al. Importance of anthropometric features to predict physical performance in elite youth soccer: a machine learning approach. *Research in Sports Medicine.* 2021;29(3):213-224. doi:10.1080/15438627.2020.1809410
- [93] Bonilla DA, et al. Effects of Ashwagandha (*Withania somnifera*) on Physical Performance: Systematic Review and Bayesian Meta-Analysis. *Journal of Functional Morphology and Kinesiology.* 2021;6(1):20.
- [94] Bonnette S, et al. A Technical Report on the Development of a Real-Time Visual Biofeedback System to Optimize Motor Learning and Movement Deficit Correction. *Journal of Sports Science and Medicine.* 2020;19:84-94.
- [95] Borenstein M, et al. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods.* 2010;1(2):97-111. doi:10.1002/jrsm.12
- [96] Bossuyt PM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med.* 2003;138(1):40-4. doi:10.7326/0003-4819-138-1-200301070-00010
- [97] Bothe TL, et al. The Art of Planning-How many samples are enough? *Acta Physiologica.* 2021:e13746. doi:10.1111/apha.13746
- [98] Bothe TL, Patzak A. Significant significance? *Acta Physiologica.* 2021;232(4):e13665. doi:https://doi.org/10.1111/apha.13665
- [99] Botonis PG, et al. The impact of daytime napping on athletic performance – A narrative review. *Scand J Med Sci Sports.* 2021;31(12):2164-2177. doi:https://doi.org/10.1111/sms.14060
- [100] Bouchard C. Genomic predictors of trainability. *Exp Physiol.* 2012;97(3):347-52. doi:10.1113/expphysiol.2011.058735
- [101] Bouchard C. The study of human variability became a passion. *Eur J Clin Nutr.* 2022;76(5):631-636. doi:10.1038/s41430-021-00871-z
- [102] Bouchard C, et al. Familial aggregation of VO₂(max) response to exercise training: results from the HERITAGE Family Study. *Journal of applied physiology (Bethesda, Md : 1985).* 1999;87(3):1003-8. doi:10.1152/jappl.1999.87.3.1003
- [103] Bouchard C, et al. Personalized preventive medicine: genetics and the response to regular exercise in preventive interventions. *Prog Cardiovasc Dis.* 2015;57(4):337-46. doi:10.1016/j.pcad.2014.08.005
- [104] Bouchard C, et al. Adverse metabolic response to regular exercise: is it a rare or common occurrence? *PLoS One.* 2012;7(5):e37887. doi:10.1371/journal.pone.0037887
- [105] Bouchard C, et al. The HERITAGE family study. Aims, design, and measurement protocol. *Med Sci Sports Exerc.* 1995;27(5):721-9. doi:https://doi.org/10.1249/00005768-199505000-00015
- [106] Bradshaw E, et al. Reliability and variability of day-to-day vault training measures in artistic gymnastics. *Sports Biomechanics.* 2010;9(2):79-97. doi:10.1080/14763141.2010.488298
- [107] Bray MS, et al. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. *Med Sci Sports Exerc.* 2009;41(1):35-73. doi:10.1249/mss.0b013e3181844179
- [108] Brefeld U, et al., eds. *Machine Learning and Data Mining for Sports Analytics.* Springer Nature; 2018. Goebel R, et al., eds. *Lecture Notes in Artificial Intelligence.*
- [109] Brennan MC, et al. Barriers and facilitators of physical activity participation in adults living with type 1 diabetes: A systematic scoping review. *Applied Physiology, Nutrition & Metabolism.* 2020;doi:10.1139/apnm-2020-0461
- [110] Brett BL, et al. Cumulative Concussion and Odds of Stroke in Former National Football League Players. *Stroke.* 2022;53(1):e5-e8. doi:10.1161/strokeaha.121.035607
- [111] Briceño-Torres JM, et al. Low-Intensity Resistance Training Improves Flow-Mediated Dilation in Young Hispanic Adults. *Journal of strength and conditioning research.* 2022;10.1519/JSC.0000000000004273. doi:10.1519/jsc.0000000000004273
- [112] Briceño-Torres JM, Moncada-Jiménez J. The Effects of Acute Exercise on Arterial Endothelial Diameter: A Meta-Analysis. *American Journal of Medicine and Medical Sciences.* 2018;8(10):251-258. doi:10.5923/j.ajmms.20180810.01
- [113] Bridgeman LA, et al. Test-Retest Reliability of a Novel Isokinetic Squat Device With Strength-Trained Athletes. *Journal of strength and conditioning research.* 2016;30(11):3261-3265. doi:10.1519/jsc.0000000000001410
- [114] British Journal of Sports Medicine. Retraction: The time lords – measurement and performance in sprinting. *British journal of sports medicine.* 2022;bjsports-39-11-785ret. doi:10.1136/bjsports-39-11-785ret
- [115] Brown WA. The placebo effect. *Sci Am.* 1998;278(1):90-5. doi:10.1038/scientificamerican0198-90
- [116] Brylka A, et al. Physical Activity, Mental Health, and Well-Being in Very Pre-Term and Term Born Adolescents: An Individual Participant Data Meta-Analysis of Two Accelerometry Studies. *Int J Environ Res Public Health.* 2021;18(4)doi:10.3390/ijerph18041735
- [117] Bubnjević KV, et al. Effect of aerobic exercise on frequency of vaginal birth: A metaanalysis. *Vojnosanitetski preglad.* 2022;79(1):55-61.
- [118] Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? *Frontiers in Physiology.* 2014;5:73. doi:10.3389/fphys.2014.00073
- [119] Buman MP, et al. Reallocating time to sleep, sedentary behaviors, or active behaviors: associations with cardiovascular disease risk biomarkers, NHANES 2005-2006. *American journal of epidemiology.* 2014;179(3):323-34. doi:10.1093/aje/kwt292
- [120] Bunker RP, Thabtah F. A machine learning framework for sport result prediction. *Applied Computing and Informatics.* 2019;15(1):27-33. doi:https://doi.org/10.1016/j.aci.2017.09.005
- [121] Burda BU, et al. Estimating data from figures with a Web-based program: Considerations for a systematic review. *Res Synth Methods.* 2017;8(3):258-262. doi:https://doi.org/10.1002/jrsm.1232
- [122] Burden RJ, et al. Elite female athlete research: stop searching for the 'magic P'. *Exp Physiol.* 2021;106(10):2029-2030. doi:https://doi.org/10.1113/EP089884
- [123] Burigo RL, et al. Concentric and eccentric isokinetic hamstring injury risk among 582 professional elite soccer players: a 10-years retrospective cohort study. *BMJ Open Sport & Exercise Medicine.* 2020;6(1):e000868. doi:10.1136/bmjsem-2020-000868
- [124] Burke ER. Nutritional ergogenic aids. In: Berning JR, Nelson-Steen S, eds. *Nutrition for sports and exercise.* Aspen Publishers; 1998:119-142:chap 7.
- [125] Burke LM, Hawley JA. Swifter, higher, stronger: What's on the menu? *Science.* 2018;362(6416):781-787. doi:10.1126/science.aau2093

- [126] Burke LM, Peeling P. Methodologies for Investigating Performance Changes With Supplement Use. *Int J Sport Nutr Exerc Metab.* 2018;28(2):159-169. doi:10.1123/ijsnem.2017-0325
- [127] Burland JP, et al. Reliability of wearable sensors to assess impact metrics during sport-specific tasks. *J Sports Sci.* 2021;39(4):406-411. doi:10.1080/02640414.2020.1823131
- [128] Büttner F, et al. Concussed athletes walk slower than non-concussed athletes during cognitive-motor dual-task assessments but not during single-task assessments 2 months after sports concussion: a systematic review and meta-analysis using individual participant data. *British journal of sports medicine.* 2020;54(2):94-101. doi:10.1136/bjsports-2018-100164
- [129] Büttner F, et al. Are questionable research practices facilitating new discoveries in sport and exercise medicine? The proportion of supported hypotheses is implausibly high. *British journal of sports medicine.* 2020;doi:10.1136/bjsports-2019-101863
- [130] Büttner F, et al. Identifying the 'incredible'! Part 1: assessing the risk of bias in outcomes included in systematic reviews. *British journal of sports medicine.* 2020;54(13):798-800. doi:10.1136/bjsports-2019-100806
- [131] Büttner F, et al. Identifying the 'incredible'! Part 2: Spot the difference - a rigorous risk of bias assessment can alter the main findings of a systematic review. *British journal of sports medicine.* 2020;54(13):801-808. doi:10.1136/bjsports-2019-101675
- [132] Butts J, et al. Creatine Use in Sports. *Sports Health.* 2018;10(1):31-34. doi:10.1177/1941738117737248
- [133] Bytowski JR. Fueling for Performance. *Sports Health.* 2018;10(1):47-53. doi:10.1177/1941738117743913
- [134] Cahusac P. Data as evidence. *Exp Physiol.* 2020;105(7):1071-1080. doi:10.1113/EP088664
- [135] Callaghan MW, Müller-Hansen F. Statistical stopping criteria for automated screening in systematic reviews. *Systematic Reviews.* 2020;9(1):273. doi:10.1186/s13643-020-01521-4
- [136] Callahan MJ, et al. Skeletal Muscle Adaptive Responses to Different Types of Short-Term Exercise Training and Detraining in Middle-Age Men. *Med Sci Sports Exerc.* 2021;53(10):2023-2036. doi:10.1249/mss.0000000000002684
- [137] Calmettes G, et al. Making do with what we have: use your bootstraps. *J Physiol.* 2012;590(15):3403-6. doi:10.1113/jphysiol.2012.239376
- [138] Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research.* Houghton Mifflin Company; 2010.
- [139] Campbell MJ, et al. Methods matter: pilot and feasibility studies in sports medicine. *British journal of sports medicine.* 2020;54(22):1309-1310. doi:10.1136/bjsports-2020-102631
- [140] Campos-Salazar C, et al. Normative values for anthropometric, adiposity, and handgrip strength in Costa Rican adults aged 53 to 110 yrs. old. *Revista Costarricense de Salud Pública.* 2019;28(2):239-248.
- [141] Carazo-Vargas P, et al. Metaanálisis sobre el efecto del ejercicio en el funcionamiento cognitivo en adultos mayores. *Revista iberoamericana de psicología del ejercicio y el deporte.* 2006;1(2):89-104.
- [142] Carazo-Vargas P, Moncada-Jiménez J. A meta-analysis on the effects of exercise training on the VO₂max in children and adolescents. *Retos.* 2015;27:184-187.
- [143] Carazo-Vargas P, et al. Lack of Agreement between Polysomnography and Accelerometry Devices for Measuring Sleep Efficiency and Movement in University Students. *Measurement in Physical Education and Exercise Science.* 2020;24(4):291-302. doi:10.1080/1091367X.2020.1819814
- [144] Carlson SA, et al. Inadequate Physical Activity and Health Care Expenditures in the United States. *Prog Cardiovasc Dis.* 2015;57(4):315-323. doi:https://doi.org/10.1016/j.pcad.2014.08.002
- [145] Carmichael MA, et al. The Impact of Menstrual Cycle Phase on Athletes' Performance: A Narrative Review. *Int J Environ Res Public Health.* 2021;18(4):1667.
- [146] Carmody S, et al. Health conditions among retired professional footballers: a scoping review. *BMJ Open Sport & Exercise Medicine.* 2022;8(2):e001196. doi:10.1136/bmjsem-2021-001196
- [147] Carpio-Rivera E, et al. Acute effects of exercise on blood pressure: A meta-analytic investigation. *Arquivos Brasileiros de Cardiologia.* 2016;106(5):422-433. doi:10.5935/abc.20160064
- [148] Carrard J, et al. Diagnosing Overtraining Syndrome: A Scoping Review. *Sports Health.* 2021;a:19417381211044739. doi:10.1177/19417381211044739
- [149] Carter EC, et al. Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods. *Advances in Methods and Practices in Psychological Science.* 2019;2(2):115-144. doi:10.1177/2515245919847196
- [150] Catellier DJ, et al. Imputation of missing data when measuring physical activity by accelerometry. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S555-S562. doi:10.1249/01.mss.0000185651.59486.4e
- [151] Ceballos-Laita L, et al. Does the Addition of Pain Neurophysiology Education to a Therapeutic Exercise Program Improve Physical Function in Women with Fibromyalgia Syndrome? Secondary Analysis of a Randomized Controlled Trial. *J Clin Med.* 2021;10(11)doi:10.3390/jcm10112518
- [152] Cece V, et al. Self-determined motivation, emotional process and subjective performance among young elite athletes: A longitudinal hierarchical linear modelling approach. *Eur J Sport Sci.* 2020;20(9):1255-1267. doi:10.1080/17461391.2019.1709562
- [153] Chacón-Moscoso S, et al. Methodological quality checklist for studies based on observational methodology (MQCOM). *Psicothema.* 2019;31(4):458-464. doi:10.7334/psicothema2019.116
- [154] Chalal LF, et al. Statistical rigor and kappa considerations: which, when and clinical context matters. *Pediatric Research.* 2020;88(1):5. doi:10.1038/s41390-020-0890-x
- [155] Challoumas D, et al. Management of patellar tendinopathy: a systematic review and network meta-analysis of randomised studies. *BMJ Open Sport & Exercise Medicine.* 2021;7(4):e001110. doi:10.1136/bmjsem-2021-001110
- [156] Chan AW, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med.* 2013;158(3):200-7. doi:10.7326/0003-4819-158-3-201302050-00583
- [157] Charan J, Biswas T. How to calculate sample size for different study designs in medical research? *Indian journal of psychological medicine.* 2013;35(2):121-6. doi:10.4103/0253-7176.116232
- [158] Chastin SFM, et al. Combined Effects of Time Spent in Physical Activity, Sedentary Behaviors and Sleep on Obesity and Cardio-Metabolic Health Markers: A Novel Compositional Data Analysis Approach. *PLoS One.* 2015;10(10):e0139984. doi:10.1371/journal.pone.0139984
- [159] Chen C-Y, et al. Increasing retractions of meta-analyses publications for methodological flaw. *Systematic Reviews.* 2021;10(1):267. doi:10.1186/s13643-021-01822-2
- [160] Chen C, et al. The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for Information Science and Technology.* 2010;61(7):1386-1409. doi:https://doi.org/10.1002/asi.21309
- [161] Chen G, et al. A Hierarchical Learning Framework for Chinese Kids Physical Exercise Prescription. Institute of Electrical and Electronics Engineers (IEEE); 2017:279-284.
- [162] Chen Y-T, et al. Two weeks of detraining reduces cardiopulmonary function and muscular fitness in endurance athletes. *Eur J Sport Sci.* 2022;22(3):399-406. doi:10.1080/17461391.2021.1880647
- [163] Chen Y, et al. Effect of Exercise on Fatigue in Multiple Sclerosis Patients: A Network Meta-analysis. *Int J Sports Med.* 2021;42(14):1250-1259. doi:10.1055/a-1524-1935
- [164] Chen YM, Wang XQ. Bibliometric Analysis of Exercise and Neuropathic Pain Research. *Journal of pain research.* 2020;13:1533-1545. doi:10.2147/jpr.s258696
- [165] Cheng C, et al. The association of cardiorespiratory fitness and the risk of hypertension: a systematic review and dose-response meta-analysis. *Journal of Human Hypertension.* 2021;doi:10.1038/s41371-021-00567-8
- [166] Chia KS, et al. The benefit of exercise training in pulmonary hypertension: a clinical review. *Intern Med J.* 2017;47(4):361-369. doi:10.1111/imj.13159
- [167] Chinchilla-Fonseca P, et al. Eficacia de la actividad física en combinación con la terapia cognitivo conductual sobre la depresión: metaanálisis. *Revista de Psicología Aplicada al Deporte y el Ejercicio Físico.* 2022;7(e3):1-28.
- [168] Chohat C, et al. Zelig: Everyone's Statistical Software. 2002.

- [169] Cicchetti D, et al. Rating Scales, Scales of Measurement, Issues of Reliability: Resolving Some Critical Issues for Clinicians and Researchers. *The Journal of Nervous and Mental Disease*. 2006;194(8):557-564. doi:10.1097/01.nmd.0000230392.83607.c5
- [170] Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994;6(4):284-290. doi:10.1037/1040-3590.6.4.284
- [171] Cillekens B, et al. How does occupational physical activity influence health? An umbrella review of 23 health outcomes across 158 observational studies. *British journal of sports medicine*. 2020;54(24):1474. doi:10.1136/bjsports-2020-102587
- [172] Clauw DJ. Fibromyalgia: a clinical review. *Jama*. 2014;311(15):1547-55. doi:10.1001/jama.2014.3266
- [173] Clemente FM, et al. Small-sided games: An umbrella review of systematic reviews and meta-analyses. *PLoS One*. 2021;16(2):e0247067. doi:10.1371/journal.pone.0247067
- [174] Clifford BK, et al. Barriers and facilitators of exercise experienced by cancer survivors: a mixed methods systematic review. *Support Care Cancer*. 2018;26(3):685-700. doi:10.1007/s00520-017-3964-5
- [175] Clinical Rehabilitation Editor's Office. Retraction notice: The blood flow restriction training effect in knee osteoarthritis people: A systematic review and meta-analysis. *Clinical rehabilitation*. 2021;269215521990590. doi:10.1177/0269215521990590
- [176] Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. 1960;20(1):37-46. doi:10.1177/001316446002000104
- [177] Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70(4):213-220. doi:https://doi.org/10.1037/h0026256
- [178] Cohen J. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates; 1988.
- [179] Cohen JF, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799. doi:10.1136/bmjopen-2016-012799
- [180] Cooper H, Patall EA. The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods*. 2009;14(2):165-176. doi:10.1037/a0015565
- [181] Cooper KB, et al. The impact of sleep on mental toughness: Evidence from observational and N-of-1 manipulation studies in athletes. *Sport, Exercise, and Performance Psychology*. 2020;9(3):308-321. doi:http://dx.doi.org/10.1037/spy0000174
- [182] Copay AG, et al. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine Journal*. 2007;7(5):541-6. doi:10.1016/j.spinee.2007.01.008
- [183] Corbett S, et al. Pilot aerobic exercise intervention for youth at-risk for serious mental illness. *Early Intervention in Psychiatry*. 2021;15(3):547-553. doi:https://doi.org/10.1111/eip.12977
- [184] Cormack SJ, et al. Reliability of measures obtained during single and repeated countermovement jumps. *Int J Sports Physiol Perform*. 2008;3(2):131-44. doi:10.1123/ijpspp.3.2.131
- [185] Coughlan D, et al. Leisure time physical activity throughout adulthood is associated with lower Medicare costs: evidence from the linked NIH-AARP diet and health study cohort. *BMJ Open Sport & Exercise Medicine*. 2021;7(1):e001038. doi:10.1136/bmjsem-2021-001038
- [186] Critical Appraisal Skills Programme. CASP qualitative checklist. Accessed Nov 1st, 2021. <https://casp-uk.net/casp-tools-checklists/>
- [187] Cruz Rivera S, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nature Medicine*. 2020;26(9):1351-1363. doi:10.1038/s41591-020-1037-7
- [188] Cryer JD, Chan KS. *Time series analysis*. 2nd ed. vol 286. Springer; 2008.
- [189] Curran-Everett D. Explorations in statistics: the log transformation. *Advances in Physiology Education*. 2018;42(2):343-347. doi:10.1152/advan.00018.2018
- [190] Currell K, Jeukendrup AE. Validity, Reliability and Sensitivity of Measures of Sporting Performance. *Sports Med*. 2008;38(4):297-316.
- [191] Cuschieri S. The CONSORT statement. *Saudi Journal of Anaesthesia*. 2019;13(5):27-30. doi:10.4103/sja.SJA_559_18
- [192] D'Amuri A, et al. Effectiveness of high-intensity interval training for weight loss in adults with obesity: a randomised controlled non-inferiority trial. *BMJ Open Sport & Exercise Medicine*. 2021;7(3):e001021. doi:10.1136/bmjsem-2020-001021
- [193] Dakic JG, et al. Effect of Pelvic Floor Symptoms on Women's Participation in Exercise: A Mixed-Methods Systematic Review With Meta-analysis. *J Orthop Sports Phys Ther*. 2021;51(7):345-361. doi:10.2519/jospt.2021.10200
- [194] Daly RM, et al. Feasibility, Usability, and Enjoyment of a Home-Based Exercise Program Delivered via an Exercise App for Musculoskeletal Health in Community-Dwelling Older Adults: Short-term Prospective Pilot Study. *JMIR Mhealth Uhealth*. 2021;9(1):e21094. doi:10.2196/21094
- [195] Dankel SJ, Loenneke JP. A Method to Stop Analyzing Random Error and Start Analyzing Differential Responders to Exercise. *Sports Med*. 2020;50(2):231-238. doi:10.1007/s40279-019-01147-0
- [196] Das P. Econometrics in theory and practice. *Springer*. 2019;10:978-981.
- [197] Davis AJ, et al. You don't need to administer a placebo to elicit a placebo effect: Social factors trigger neurobiological pathways to enhance sports performance. *Eur J Sport Sci*. 2020;20(3):302-312. doi:10.1080/17461391.2019.1635212
- [198] De Bruijn AGM, et al. Effects of aerobic and cognitively-engaging physical activity on academic skills: A cluster randomized controlled trial. *J Sports Sci*. 2020;38(15):1806-1817. doi:10.1080/02640414.2020.1756680
- [199] de Craen AJ, et al. Effect of colour of drugs: systematic review of perceived effect of drugs and of their effectiveness. *BMJ (Clinical research ed)*. 1996;313(7072):1624-6. doi:10.1136/bmj.313.7072.1624
- [200] de Jesus Leite MAF, et al. Exercise training and detraining effects on body composition, muscle strength and lipid, inflammatory and oxidative markers in breast cancer survivors under tamoxifen treatment. *Life Sci*. 2021;284:119924. doi:10.1016/j.lfs.2021.119924
- [201] Deakin ND, et al. Research Evaluating Sports Concussion Events—Rapid Assessment of Concussion and Evidence for Return (RESCUE-RACER): a two-year longitudinal observational study of concussion in motorsport. *BMJ Open Sport & Exercise Medicine*. 2021;7(1):e000879. doi:10.1136/bmjsem-2020-000879
- [202] *OpenEpi: Open-Source Epidemiologic Statistics for Public Health*. 2013.
- [203] Demiris G, et al. *Behavioral Intervention Research in Hospice and Palliative Care: Building an Evidence Base*. Academic Press; 2019.
- [204] DeNysschen CA, et al. Exercise with Nutrition Education to Improve Quality of Life of Adolescent and Young Adult Cancer Survivors: A Pilot Study. *J Adolesc Young Adult Oncol*. 2021;10(4):454-461. doi:10.1089/jyao.2020.0116
- [205] Depenbusch J, et al. The association between physicians' exercise counseling and physical activity in patients with cancer: Which roles do patients' satisfaction and previous physical activity levels play? *Psychooncology*. 2020;29(11):1856-1863. doi:10.1002/pon.5506
- [206] Dias S, et al. *Network Meta-Analysis for Decision Making*. John Wiley & Sons Ltd.; 2018.
- [207] Dias S, Caldwell DM. Network meta-analysis explained. *Arch Dis Child Fetal Neonatal Ed*. 2019;104(1):F8-f12. doi:10.1136/archdischild-2018-315224
- [208] Dias S, Caldwell DM. Network meta-analysis explained. *Archives of Disease in Childhood - Fetal and Neonatal Edition*. 2019;104(1):F8-F12. doi:10.1136/archdischild-2018-315224
- [209] Dieterich AV, et al. Reducing sedentary behaviour and physical inactivity in the workplace: protocol for a review of systematic reviews. *BMJ Open Sport & Exercise Medicine*. 2020;6(1):e000909. doi:10.1136/bmjsem-2020-000909
- [210] Dieu O, et al. Complementing subjective with objective data in analysing expertise: A machine-learning approach applied to badminton. *J Sports Sci*. 2020;38(17):1943-1952. doi:10.1080/02640414.2020.1764812
- [211] Ding D, et al. Is the COVID-19 lockdown nudging people to be more active: a big data analysis. *British journal of sports medicine*. 2020;doi:10.1136/bjsports-2020-102575
- [212] Ding D, et al. The economic burden of physical inactivity: a systematic review and critical appraisal. *British journal of sports*

- medicine*. 2017;51(19):1392-1409. doi:10.1136/bjsports-2016-097385
- [213] Doi SAR, et al. Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model. *Contemp Clin Trials*. 2015;45:130-138. doi:https://doi.org/10.1016/j.cct.2015.05.009
- [214] Doi SAR, et al. Simulation Comparison of the Quality Effects and Random Effects Methods of Meta-analysis. *Epidemiology*. 2015;26(4):e42-e44. doi:10.1097/ede.0000000000000289
- [215] Dolan E, et al. The influence of acute exercise on bone biomarkers: protocol for a systematic review with meta-analysis. *Systematic Reviews*. 2020;9(1):291. doi:10.1186/s13643-020-01551-y
- [216] Donati Zeppa S, et al. Nine weeks of high-intensity indoor cycling training induced changes in the microbiota composition in non-athlete healthy male college students. *J Int Soc Sports Nutr*. 2021;18(1):74. doi:10.1186/s12970-021-00471-z
- [217] Dong J. Simpson's Paradox. *Wiley StatsRef: Statistics Reference Online*. 2014.
- [218] Doosti-Irani A, et al. What are network meta-analyses (NMAs)? A primer with four tips for clinicians who read NMAs and who perform them (methods matter series). *British journal of sports medicine*. 2021;55(10):520. doi:10.1136/bjsports-2020-102872
- [219] Dos Santos L, et al. Changes in phase angle and body composition induced by resistance training in older women. *Eur J Clin Nutr*. 2016;70(12):1408-1413. doi:10.1038/ejcn.2016.124
- [220] dos Santos MAM, et al. The influence of anthropometric variables, body composition, propulsive force and maturation on 50m freestyle swimming performance in junior swimmers: An allometric approach. *J Sports Sci*. 2021;39(14):1615-1620. doi:10.1080/02640414.2021.1891685
- [221] Dowdy A, et al. Structured visual analysis of single-case experimental design data: Developments and technological advancements. *Journal of Applied Behavior Analysis*. 2022;55(2):451-462. doi:https://doi.org/10.1002/jaba.899
- [222] Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998;52(6):377-84. doi:10.1136/jech.52.6.377
- [223] Dreyer NA, et al. The GRACE Checklist: A Validated Assessment Tool for High Quality Observational Studies of Comparative Effectiveness. *J Manag Care Spec Pharm*. 2016;22(10):1107-13. doi:10.18553/jmcp.2016.22.10.1107
- [224] Drummond GB, et al. Statistics: all together now, one step at a time. *J Physiol*. 2011;589(Pt 8):1859. doi:10.1113/jphysiol.2011.206219
- [225] Drummond GB, Tom BD. How can we tell if frogs jump further? *J Physiol*. 2011;589(Pt 14):3409-13. doi:10.1113/jphysiol.2011.211870
- [226] Drummond GB, Tom BD. Presenting data: can you follow a recipe? *Exp Physiol*. 2011;96(12):1249-52. doi:10.1113/expphysiol.2011.062349
- [227] Drummond GB, Tom BD. Statistics, probability, significance, likelihood: words mean what we define them to mean. *Exp Physiol*. 2011;96(9):817-21. doi:10.1113/expphysiol.2011.060459
- [228] Drummond GB, Vowler SL. Show the data, don't conceal them. *J Physiol*. 2011;589(Pt 8):1861-3. doi:10.1113/jphysiol.2011.205062
- [229] Drummond GB, Vowler SL. Analysis of variance: variably complex. *J Physiol*. 2012;590(6):1303-6. doi:10.1113/jphysiol.2012.229856
- [230] Drummond GB, Vowler SL. Different tests for a difference: how do we do research? *J Physiol*. 2012;590(2):235-8. doi:10.1113/jphysiol.2011.225235
- [231] Drummond GB, Vowler SL. Type I: families, planning and errors. *J Physiol*. 2012;590(20):4971-4. doi:10.1113/jphysiol.2012.244061
- [232] Drummond GB, Vowler SL. Do as you would be done by: write as you would wish to read. *Exp Physiol*. 2013;98(2):355-8. doi:10.1113/expphysiol.2012.071035
- [233] Drummond GB, Vowler SL. Not different is not the same as the same: how can we tell? *Exp Physiol*. 2013;98(2):351-4. doi:10.1113/expphysiol.2012.070052
- [234] Düking P, et al. Proof-of-concept and concurrent validity of a prototype headset to assess peak oxygen uptake without a face mask. *BMC Research Notes*. 2022;15(1):4. doi:10.1186/s13104-021-05850-y
- [235] Dumuid D, et al. The compositional isotemporal substitution model: A method for estimating changes in a health outcome for reallocation of time between sleep, physical activity and sedentary behaviour. *Stat Methods Med Res*. 2019;28(3):846-857. doi:10.1177/0962280217737805
- [236] Dumuid D, et al. Adiposity and the isotemporal substitution of physical activity, sedentary time and sleep among school-aged children: a compositional data analysis approach. *BMC Public Health*. 2018;18(1):311. doi:10.1186/s12889-018-5207-1
- [237] Duncan MJ, et al. TGMD-3 short version: Evidence of validity and associations with sex in Irish children. *J Sports Sci*. 2022;40(2):138-145. doi:10.1080/02640414.2021.1978161
- [238] Dupuy A, et al. Post-exercise heart rate recovery and parasympathetic reactivation are comparable between prepubertal boys and well-trained adult male endurance athletes. *European Journal of Applied Physiology*. 2022;122(2):345-355. doi:10.1007/s00421-021-04823-0
- [239] Dziura JD, et al. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med*. 2013;86(3):343-58.
- [240] Eckstein ML, et al. Physical exercise and non-insulin glucose-lowering therapies in the management of Type 2 diabetes mellitus: a clinical review. *Diabet Med*. 2019;36(3):349-358. doi:10.1111/dme.13865
- [241] Egger F, et al. FIFA Sudden Death Registry (FIFA-SDR): a prospective, observational study of sudden death in worldwide football from 2014 to 2018. *British journal of sports medicine*. 2022;56(2):80-87. doi:10.1136/bjsports-2020-102368
- [242] Ekblom-Bak E, et al. Latent profile analysis patterns of exercise, sitting and fitness in adults – Associations with metabolic risk factors, perceived health, and perceived symptoms. *PLoS One*. 2020;15(4):e0232210. doi:10.1371/journal.pone.0232210
- [243] Ekelund U, et al. Physical activity and mortality: what is the dose response and how big is the effect? *British journal of sports medicine*. 2020;54(19):1125-1126. doi:10.1136/bjsports-2019-101765
- [244] Ekelund U, et al. Joint associations of accelerometer measured physical activity and sedentary time with all-cause mortality: a harmonised meta-analysis in more than 44 000 middle-aged and older individuals. *British journal of sports medicine*. 2020;54(24):1499. doi:10.1136/bjsports-2020-103270
- [245] Ekrami N, et al. Wet-cupping induces anti-inflammatory action in response to vigorous exercise among martial arts athletes: A pilot study. *Complement Ther Med*. 2021;56:102611. doi:10.1016/j.ctim.2020.102611
- [246] El-Medany AYM, et al. What Change in Body Mass Index Is Required to Improve Cardiovascular Outcomes in Childhood and Adolescent Obesity through Lifestyle Interventions: A Meta-Regression. *Childhood Obesity*. 2020;16(7):449-478. doi:10.1089/chi.2019.0286
- [247] Eldridge SM, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ (Clinical research ed)*. 2016;355:i5239. doi:10.1136/bmj.i5239
- [248] Elliott JH, et al. Living systematic review: 1. Introduction-the why, what, when, and how. *J Clin Epidemiol*. 2017;91:23-30. doi:10.1016/j.jclinepi.2017.08.010
- [249] Elliott JH, et al. Living Systematic Reviews: An Emerging Opportunity to Narrow the Evidence-Practice Gap. *PLoS Medicine*. 2014;11
- [250] Emm-Collison LG, et al. Multidimensional motivation for exercise: A latent profile and transition analysis. *Psychology of Sport & Exercise*. 2020;47:101619. doi:https://doi.org/10.1016/j.psychsport.2019.101619
- [251] Meta XL. Version 5.3. EpiGear Intl.; 2016. https://www.epigear.com/index_files/metaxl.html
- [252] Ezzatvar Y, et al. Physical activity and risk of infection, severity and mortality of COVID-19: a systematic review and non-linear dose-response meta-analysis of data from 1 853 610 adults. *British journal of sports medicine*. 2022;bjsports-2022-105733. doi:10.1136/bjsports-2022-105733
- [253] Faraone SV. Interpreting estimates of treatment effects: implications for managed care. *Pharmacy & Therapeutics*. 2008;33(12):700-11.
- [254] Farrance C, et al. Adherence to community based group exercise interventions for older people: A mixed-methods systematic review. *Prev Med*. 2016;87:155-166. doi:10.1016/j.ypmed.2016.02.037

- [255] Faul F, et al. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior research methods*. 2009;41(4):1149-60. doi:10.3758/brm.41.4.1149
- [256] Faul F, et al. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*. 2007;39(2):175-91. doi:10.3758/bf03193146
- [257] Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006;27(8):861-874. doi:https://doi.org/10.1016/j.patrec.2005.10.010
- [258] Fayanju OM, et al. CHEERS Reporting Guidelines for Economic Evaluations. *JAMA Surgery*. 2021;156(7):677-678. doi:10.1001/jamasurg.2021.0540
- [259] Feasel CD, et al. Aerobic reserve capacity in multiple sclerosis-Preliminary evidence. *Acta Neurol Scand*. 2021;doi:10.1111/ane.13441
- [260] Feijen S, et al. Monitoring the swimmer's training load: A narrative review of monitoring strategies applied in research. *Scand J Med Sci Sports*. 2020;doi:10.1111/sms.13798
- [261] Feng C, et al. Log transformation: application and interpretation in biomedical research. *Stat Med*. 2013;32(2):230-239. doi:https://doi.org/10.1002/sim.5486
- [262] Fernández A, et al. *Normas Nacionales. Componentes de salud física: estudiantes costarricenses entre 8 y 17 años*. Editorial Baula; 2001.
- [263] Ferreira AFF, et al. The effects of treadmill exercise in animal models of Parkinson's disease: A systematic review. *Neurosci Biobehav Rev*. 2021;131:1056-1075. doi:10.1016/j.neubiorev.2021.10.019
- [264] Fiedler J, et al. Key facets to build up eHealth and mHealth interventions to enhance physical activity, sedentary behavior and nutrition in healthy subjects - an umbrella review. *BMC Public Health*. 2020;20(1):1605. doi:10.1186/s12889-020-09700-7
- [265] Filho MM, et al. Effects of Different Types of Resistance Training and Detraining on Functional Capacity, Muscle Strength, and Power in Older Women: A Randomized Controlled Study. *Journal of strength and conditioning research*. 2022;36(4):984-990. doi:10.1519/jsc.0000000000004195
- [266] Fitzpatrick J, Panagodage Perera N. The Biathlon Injury and Illness Surveillance (BIIS) project protocol: a prospective cohort study across two World Cup seasons. *BMJ Open Sport & Exercise Medicine*. 2020;6(1):e000862. doi:10.1136/bmjsem-2020-000862
- [267] Forsythe ID, et al. Reproducibility and data presentation. *J Physiol*. 2019;597(22):5313. doi:10.1113/jp277519
- [268] Foster RL. What a pilot study is and what it is not. *Journal for Specialists in Pediatric Nursing*. 2013;18(1):1-2. doi:https://doi.org/10.1111/jspn.12015
- [269] Franke TPC, et al. Minimally important change and smallest detectable change of the OSTRC questionnaire in half- and full-marathon runners. *Scand J Med Sci Sports*. 2021;31(5):1048-1058. doi:https://doi.org/10.1111/sms.13885
- [270] Fridolfsson J, et al. Physical activity spectrum discriminant analysis—A method to compare detailed patterns between groups. *Scand J Med Sci Sports*. 2021;31(12):2333-2342. doi:https://doi.org/10.1111/sms.14052
- [271] Fritz NE, et al. Physical Therapy and Exercise Interventions in Huntington's Disease: A Mixed Methods Systematic Review. *J Huntingtons Dis*. 2017;6(3):217-235. doi:10.3233/jhd-170260
- [272] Fuller D, et al. Why machine learning (ML) has failed physical activity research and how we can improve. *BMJ Open Sport & Exercise Medicine*. 2022;8(1):e001259. doi:10.1136/bmjsem-2021-001259
- [273] Funder DC, Ozer DJ. Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*. 2019;2(2):156-168. doi:10.1177/2515245919847202
- [274] Furuya-Kanamori L, et al. A new improved graphical and quantitative method for detecting bias in meta-analysis. *Int J Evid Based Healthc*. 2018;16(4):195-203. doi:10.1097/xeb.0000000000000141
- [275] Gába A, et al. Replacing school and out-of-school sedentary behaviors with physical activity and its associations with adiposity in children and adolescents: a compositional isotemporal substitution analysis. *Environmental health and preventive medicine*. 2021;26(1):16. doi:10.1186/s12199-021-00932-6
- [276] Gabrys L, et al. Exercise-Based Cardiac Rehabilitation: Secondary Data Analyses of Mortality and Working Capacity in Germany, 2010-2017. *Sports Med Open*. 2021;7(1):88. doi:10.1186/s40798-021-00381-z
- [277] Gage NA, Lewis TJ. Analysis of effect for single-case design research. *Journal of Applied Sport Psychology*. 2013;25(1):46-60.
- [278] Gagnier JJ, et al. The CARE guidelines: consensus-based clinical case reporting guideline development. *Journal of Medical Case Reports*. 2013;7(1):223. doi:10.1186/1752-1947-7-223
- [279] Gallo G, et al. How do world class top 5 Giro d'Italia finishers train? A qualitative multiple case study. *Scand J Med Sci Sports*. 2022;doi:https://doi.org/10.1111/sms.14201
- [280] Gandevia S. Publications, replication and statistics in physiology plus two neglected curves. *J Physiol*. 2021;doi:10.1113/jp281360
- [281] García-Suárez PC, et al. Acute Systemic Response of BDNF, Lactate and Cortisol to Strenuous Exercise Modalities in Healthy Untrained Women. *Dose-Response*. 2020;18(4):1559325820970818. doi:10.1177/1559325820970818
- [282] García-Suárez PC, et al. The effects of interval training on peripheral brain derived neurotrophic factor (BDNF) in young adults: a systematic review and meta-analysis. *Sci Rep*. 2021;11(1):8937. doi:10.1038/s41598-021-88496-x
- [283] Garnica-Caparrós M, Memmert D. Understanding gender differences in professional European football through machine learning interpretability and match actions data. *Sci Rep*. 2021;11(1):10805. doi:10.1038/s41598-021-90264-w
- [284] George AM, et al. Association between socio-economic status and physical activity is mediated by social support in Brazilian students. *J Sports Sci*. 2019;37(5):500-506. doi:10.1080/02640414.2018.1509435
- [285] Gianola S, et al. Effectiveness of treatments for acute and subacute mechanical non-specific low back pain: a systematic review with network meta-analysis. *British journal of sports medicine*. 2022;56(1):41-50. doi:10.1136/bjsports-2020-103596
- [286] Gill JMR. Linking volume and intensity of physical activity to mortality. *Nature Medicine*. 2020;doi:10.1038/s41591-020-1019-9
- [287] Gladin A, et al. Secondary analysis of change in physical function after exercise intervention in older adults with hyperkyphosis and low physical function. *BMC Geriatr*. 2021;21(1):133. doi:10.1186/s12877-021-02062-8
- [288] Godwin M, et al. Intra and intersession reliability of the Run Rocket™ in recreationally trained participants. *Journal of Human Sport and Exercise*. 2021;16(1):53-62. doi:10.14198/jhse.2021.161.05
- [289] Gomes GAO, et al. Twelve year trajectories of physical activity and health costs in mid-age Australian women. *Int J Behav Nutr Phys Act*. 2020;17(1):101. doi:10.1186/s12966-020-01006-6
- [290] González-Xuriguera CG, et al. Introduction to network meta-analysis for evidence synthesis. *Medwave*. 2021;21(6):e8315. doi:https://doi.org/10.5867/medwave.2021.06.8315
- [291] Graham M, et al. The school playground environment as a driver of primary school children's physical activity behaviour: A direct observation case study. *J Sports Sci*. 2021;39(20):2266-2278. doi:10.1080/02640414.2021.1928423
- [292] Gram MCD, et al. Injuries and illnesses among competitive Norwegian rhythmic gymnasts during preseason: a prospective cohort study of prevalence, incidence and risk factors. *British journal of sports medicine*. 2021;55(4):231-236. doi:10.1136/bjsports-2020-102315
- [293] Grandou C, et al. Overtraining in Resistance Exercise: An Exploratory Systematic Review and Methodological Appraisal of the Literature. *Sports Med*. 2020;50(4):815-828. doi:10.1007/s40279-019-01242-2
- [294] Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J*. 2009;26(2):91-108. doi:10.1111/j.1471-1842.2009.00848.x
- [295] Greenfieldboyce N. 'Shrimp On A Treadmill': The Politics Of 'Silly' Studies. 2011. https://www.npr.org/2011/08/23/139852035/shrimp-on-a-treadmill-the-politics-of-silly-studies
- [296] Greenland S, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*. 2016;31(4):337-350. doi:10.1007/s10654-016-0149-3

- [297] Grgic J. Effects of Music on Resistance Exercise Performance: A Narrative Review. *Strength & Conditioning Journal*. 2021;44(4):77-84. doi:10.1519/ssc.0000000000000682
- [298] Grgic J, et al. Health outcomes associated with reallocations of time between sleep, sedentary behaviour, and physical activity: a systematic scoping review of isotemporal substitution studies. *Int J Behav Nutr Phys Act*. 2018;15(1):69. doi:10.1186/s12966-018-0691-3
- [299] Griffin, et al. Training load monitoring in team sports: a practical approach to addressing missing data. *J Sports Sci*. 2021;39(19):2161-2171. doi:10.1080/02640414.2021.1923205
- [300] Grootswagers P, et al. Relative Validity and Reliability of Isometric Lower Extremity Strength Assessment in Older Adults by Using a Handheld Dynamometer. *Sports Health*. 2022;19417381211063847. doi:10.1177/19417381211063847
- [301] Guo F, Huang Q. Signal Recognition Based on APSO-RBF Neural Network to Assist Athlete's Competitive Ability Evaluation. *Comput Intell Neurosci*. 2021;2021:4850020. doi:10.1155/2021/4850020
- [302] Gurd BJ, et al. Incidence of nonresponse and individual patterns of response following sprint interval training. *Applied Physiology, Nutrition & Metabolism*. 2016;41(3):229-34. doi:10.1139/apnm-2015-0449
- [303] Gut V, et al. What do adolescents and young adults strive for in sport and exercise? An explorative study on goal profiles in sport and exercise. *J Sports Sci*. 2022;40(5):571-582. doi:10.1080/02640414.2021.2004703
- [304] Gutiérrez-Sancho OP, et al. The effect of creatine supplementation on biochemical, body composition, and performance outcomes in humans: A meta-analysis. *International Journal of Applied Sports Sciences*. 2006;18(2):12-38.
- [305] Guyatt GH, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical research ed)*. 2008;336(7650):924-6. doi:10.1136/bmj.39489.470347.AD
- [306] Haby MM, et al. What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review. *Health Res Policy Syst*. 2016;14(1):83. doi:10.1186/s12961-016-0155-7
- [307] Hackshaw A. Small studies: strengths and limitations. *European Respiratory Journal*. 2008;32(5):1141-3. doi:10.1183/09031936.00136408
- [308] Haddaway NR, et al. Systematic reviews: A glossary for public health. *Scandinavian Journal of Public Health*. 2022;14034948221074998. doi:10.1177/14034948221074998
- [309] Haenlein M, Kaplan A. A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*. 2019;61(4):5-14. doi:10.1177/0008125619864925
- [310] Hafner M, et al. *The economic benefits of a more physically active population: An international analysis*. RAND Corporation; 2019.
- [311] Hägglund E, et al. Effects of yoga versus hydrotherapy training on health-related quality of life and exercise capacity in patients with heart failure: A randomized controlled study. *Eur J Cardiovasc Nurs*. 2017;16(5):381-389. doi:10.1177/1474515117690297
- [312] Hahn S. Understanding noninferiority trials. *Korean J Pediatr*. 2012;55(11):403-407. doi:10.3345/kjp.2012.55.11.403
- [313] Haiyun Z, Yizhe X. Sports performance prediction model based on integrated learning algorithm and cloud computing Hadoop platform. *Microprocessors and Microsystems*. 2020;79:103322. doi:https://doi.org/10.1016/j.micpro.2020.103322
- [314] Hall M, et al. Comparative effectiveness of exercise programs for psychological well-being in knee osteoarthritis: A systematic review and network meta-analysis. *Semin Arthritis Rheum*. 2021;51(5):1023-1032. doi:10.1016/j.semarthrit.2021.07.007
- [315] Halperin I, Emanuel A. Rating of Perceived Effort: Methodological Concerns and Future Directions. *Sports Med*. 2020;50(4):679-687. doi:10.1007/s40279-019-01229-z
- [316] Hamer M, et al. Effects of substituting sedentary time with physical activity on metabolic risk. *Med Sci Sports Exerc*. 2014;46(10):1946-50. doi:10.1249/mss.0000000000000317
- [317] Han J, et al. Can therapeutic exercises improve proprioception in chronic ankle instability? a systematic review and network meta-analysis. *Arch Phys Med Rehabil*. 2022;doi:10.1016/j.apmr.2022.04.007
- [318] Han L, et al. Corrigendum. *J Physiol*. 2021;doi:https://doi.org/10.1113/JP282313
- [319] Han M, et al. Cardiorepiratory fitness and mortality from all causes, cardiovascular disease and cancer: dose-response meta-analysis of cohort studies. *British journal of sports medicine*. 2022;56(13):733-739. doi:10.1136/bjsports-2021-104876
- [320] Haverinen EH, et al. The effects of long-term physical activity interventions in communities: Scoping review in the Nordic countries. *Scandinavian Journal of Public Health*. 2021;14034948211020599. doi:10.1177/14034948211020599
- [321] Hayden JA, et al. Exercise treatment effect modifiers in persistent low back pain: an individual participant data meta-analysis of 3514 participants from 27 randomised controlled trials. *British journal of sports medicine*. 2020;54(21):1277-1278. doi:10.1136/bjsports-2019-101205
- [322] Hayden MC, et al. Effectiveness of a Three-Week Inpatient Pulmonary Rehabilitation Program for Patients after COVID-19: A Prospective Observational Study. *Int J Environ Res Public Health*. 2021;18(17)doi:10.3390/ijerph18179001
- [323] Hayes AF. Introduction to mediation, moderation, and conditional process analysis: Methodology in the Social Sciences. *Kindle Edition*. 2013;193
- [324] Hayes AF. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications; 2017.
- [325] Hayes EJ, et al. Recovery from resistance exercise in older adults: a protocol for a scoping review. *BMJ Open Sport & Exercise Medicine*. 2022;8(1):e001229. doi:10.1136/bmjsem-2021-001229
- [326] Haynes EMK, et al. Age and sex-related decline of muscle strength across the adult lifespan: a scoping review of aggregated data. *Applied Physiology, Nutrition & Metabolism*. 2020;45(11):1185-1196. doi:10.1139/apnm-2020-0081
- [327] Hazra A. Using the confidence interval confidently. *Journal of Thoracic Disease*. 2017;9(10):4124-4129.
- [328] Healy MJ. Data transformations. *Archives of Disease in Childhood*. 1993;69(2):260-264. doi:10.1136/adc.69.2.260
- [329] Heath MR, et al. Erratum to "Age is More Predictive of Safe Movement Patterns Than Are Physical Activity or Sports Specialization: A Prospective Motion Analysis Study of Young Athletes". *The American Journal of Sports Medicine*. 2021;03635465211047896. doi:10.1177/03635465211047896
- [330] Hecksteden A, et al. Individual response to exercise training - a statistical perspective. *Journal of applied physiology (Bethesda, Md : 1985)*. 2015;118(12):1450-9. doi:10.1152/jappphysiol.00714.2014
- [331] Hecksteden A, et al. Repeated testing for the assessment of individual response to exercise training. *Journal of applied physiology (Bethesda, Md : 1985)*. 2018;124(6):1567-1579. doi:10.1152/jappphysiol.00896.2017
- [332] Held L, et al. Replication power and regression to the mean. *Significance*. 2020;17(6):10-11. doi:https://doi.org/10.1111/1740-9713.01462
- [333] Heneghan NR, et al. Lumbosacral injuries in elite Paralympic athletes with limb deficiency: a retrospective analysis of patient records. *BMJ Open Sport & Exercise Medicine*. 2021;7(1):e001001. doi:10.1136/bmjsem-2020-001001
- [334] Hernández Sampieri R, et al. *Metodología de la investigación*. 6 ed. McGraw Hill; 2014.
- [335] Herold F, et al. Dose-Response Matters! – A Perspective on the Exercise Prescription in Exercise-Cognition Research. *Frontiers in psychology*. 2019;10(2338)doi:10.3389/fpsyg.2019.02338
- [336] Herold F, et al. Causes and Consequences of Interindividual Response Variability: A Call to Apply a More Rigorous Research Design in Acute Exercise-Cognition Studies. *Front Physiol*. 2021;12:682891. doi:10.3389/fphys.2021.682891
- [337] Hertel J. A Picture Tells 1000 Words (but Most Results Graphs Do Not): 21 Alternatives to Simple Bar and Line Graphs. *Clinics in Sports Medicine*. 2018;37(3):441-462. doi:10.1016/j.csm.2018.04.001
- [338] Higgins JP, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ (Clinical research ed)*. 2011;343:d5928. doi:10.1136/bmj.d5928
- [339] Higgins JP, et al. Measuring inconsistency in meta-analyses. *BMJ (Clinical research ed)*. 2003;327(7414):557-60. doi:10.1136/bmj.327.7414.557

- [340] Higuti AY, et al. Effects of listening to music and practicing physical exercise on functional and cognitive aspects in institutionalized older adults with dementia: Pilot study. *EXPLORE*. 2021;17(4):292-296. doi:https://doi.org/10.1016/j.explore.2020.07.006
- [341] Hinman SK, et al. Exercise in Pregnancy: A Clinical Review. *Sports Health*. 2015;7(6):527-31. doi:10.1177/1941738115599358
- [342] Ho RS-T, et al. Active video game on children and adolescents' physical activity and weight management: A network meta-analysis. *Scand J Med Sci Sports*. 2022;32(8):1268-1286. doi:https://doi.org/10.1111/sms.14176
- [343] Hopkins WG. Measures of Reliability in Sports Medicine and Science. *Sports Med*. 2000;30(1):1-15. doi:10.2165/00007256-200030010-00001
- [344] Hopkins WG. A new view of statistics. <http://www.sportsci.org/resource/stats/effectmag.html>
- [345] Hopkins WG. Individual responses made easy. *Journal of applied physiology (Bethesda, Md : 1985)*. 2015;118(12):1444-6. doi:10.1152/jappphysiol.00098.2015
- [346] Hopkins WG. Spreadsheets for analysis of validity and reliability. *Sports Science*. 2015;19:36-42.
- [347] Hopkins WG. A Spreadsheet for Monitoring an Individual's Changes and Trend. *Sports Science*. 2017;21
- [348] Hopkins WG, et al. Progressive Statistics for Studies in Sports Medicine and Exercise Science. *Med Sci Sports Exerc*. 2009;41(1):3-12. doi:10.1249/MSS.0b013e31818cb278
- [349] Hopkins WG, et al. Reliability of Power in Physical Performance Tests. *Sports Med*. 2001;31(3):211-234. doi:10.2165/00007256-200131030-00005
- [350] Horváth Á, et al. Nocebo effects on motor performance: A systematic literature review. *Scand J Psychol*. 2021;62(5):665-674. doi:10.1111/sjop.12753
- [351] Hróbjartsson A, Gøtzsche PC. Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment. *J Intern Med*. 2004;256(2):91-100. doi:10.1111/j.1365-2796.2004.01355.x
- [352] Hrubeniuk TJ, et al. Directions for Exercise Treatment Response Heterogeneity and Individual Response Research. *Int J Sports Med*. 2022;43(1):11-22. doi:10.1055/a-1548-7026
- [353] Hsu C-Y, et al. Regular Exercise and Weight-Control Behavior Are Protective Factors against Osteoporosis for General Population: A Propensity Score-Matched Analysis from Taiwan Biobank Participants. *Nutrients*. 2022;14(3):641.
- [354] Huang X, et al. Comparative efficacy of various exercise interventions on cognitive function in patients with mild cognitive impairment or dementia: A systematic review and network meta-analysis. *J Sport Health Sci*. 2022;11(2):212-223. doi:10.1016/j.jshs.2021.05.003
- [355] Huard Pelletier V, et al. Video games and their associations with physical health: a scoping review. *BMJ Open Sport & Exercise Medicine*. 2020;6(1):e000832. doi:10.1136/bmjsem-2020-000832
- [356] Hughes RA, et al. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *International Journal of Epidemiology*. 2019;48(4):1294-1304. doi:10.1093/ije/dyz032
- [357] Hurst P, et al. Athletes Intending to Use Sports Supplements Are More Likely to Respond to a Placebo. *Med Sci Sports Exerc*. 2017;49(9):1877-1883. doi:10.1249/mss.0000000000001297
- [358] Hurst P, et al. The Placebo and Nocebo effect on sports performance: A systematic review. *Eur J Sport Sci*. 2020;20(3):279-292. doi:10.1080/17461391.2019.1655098
- [359] Iannaccone PM, Jacob HJ. Rats! *Dis Model Mech*. 2009;2(5-6):206-10. doi:10.1242/dmm.002733
- [360] Ikeda T, et al. Maintaining Moderate or Vigorous Exercise Reduces the Risk of Low Back Pain at 4 Years of Follow-Up: Evidence From the English Longitudinal Study of Ageing. *The Journal of Pain*. 2022;23(3):390-397. doi:https://doi.org/10.1016/j.jpain.2021.08.008
- [361] Int'Hout J, et al. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology*. 2014;14(1):25. doi:10.1186/1471-2288-14-25
- [362] Int'Hout J, et al. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*. 2016;6(7):e010247. doi:10.1136/bmjopen-2015-010247
- [363] Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124. doi:10.1371/journal.pmed.0020124
- [364] Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q*. 2016;94(3):485-514. doi:10.1111/1468-0009.12210
- [365] Ioannidis JPA. What Have We (Not) Learnt from Millions of Scientific Papers with P Values? *The American Statistician*. 2019;73(sup1):20-25. doi:10.1080/00031305.2018.1447512
- [366] Ishihara T, et al. The effects of acute aerobic exercise on executive function: A systematic review and meta-analysis of individual participant data. *Neurosci Biobehav Rev*. 2021;128:258-269. doi:10.1016/j.neubiorev.2021.06.026
- [367] Islam H, et al. Repeatability of exercise-induced changes in mRNA expression and technical considerations for qPCR analysis in human skeletal muscle. *Exp Physiol*. 2019;104(3):407-420. doi:10.1113/ep087401
- [368] Islam H, Gurd BJ. Exercise response variability: Random error or true differences in exercise response? *Exp Physiol*. 2020;105(12):2022-2024. doi:10.1113/ep089015
- [369] Ismail NA, et al. Physical Activity and Exergames Among Older Adults: A Scoping Review. *Games Health J*. 2022;11(1):1-17. doi:10.1089/g4h.2021.0104
- [370] Jacques M, et al. Epigenetic changes in healthy human skeletal muscle following exercise- a systematic review. *Epigenetics*. 2019;14(7):633-648. doi:10.1080/15592294.2019.1614416
- [371] Jaén-Carrillo D, et al. Stiffness in Running: A Narrative Integrative Review. *Strength & Conditioning Journal*. 2021;43(2):104-115. doi:10.1519/ssc.0000000000000593
- [372] Jakobsen JC, et al. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology*. 2017;17(1):162. doi:10.1186/s12874-017-0442-1
- [373] James LJ, et al. Does Hypohydration Really Impair Endurance Performance? Methodological Considerations for Interpreting Hydration Research. *Sports Med*. 2019;49(Suppl 2):103-114. doi:10.1007/s40279-019-01188-5
- [374] Jarstad E, Mamen A. Correction: The performance and aerobic endurance effects of high-intensity versus moderate-intensity continuous running. *Applied Physiology, Nutrition & Metabolism*. 2021;46(11):1435. doi:10.1139/apnm-2021-0669
- [375] JASP. Version 0.13.1. The JASP Team; 2020. <https://jasp-stats.org/>
- [376] Jeffreys H. *Theory of probability*. 3rd ed. Oxford University Press; 1961.
- [377] Jiménez-Maldonado A, et al. Effect of exercise training on peripheral brain derived neurotrophic factor (BDNF) levels in Type 2 diabetes patients: a systematic review and meta-analysis. 2022. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021252856
- [378] Jiménez-Morgan S, Hernández-Elizondo J. Efecto del ejercicio sobre la fatiga asociada al cáncer de mama en mujeres: meta-análisis de ensayos clínicos controlados aleatorizados. *Journal of Sport and Health Research*. 2017;9(3):285-290.
- [379] Joensuu L, et al. Precision exercise medicine: predicting unfavourable status and development in the 20-m shuttle run test performance in adolescence with machine learning. *BMJ Open Sport & Exercise Medicine*. 2021;7(2):doi:10.1136/bmjsem-2021-001053
- [380] Johnston BC, et al. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important difference units. *Health and Quality of Life Outcomes*. 2010;8(1):116. doi:10.1186/1477-7525-8-116
- [381] Jouven X, et al. Heart-rate profile during exercise as a predictor of sudden death. *N Engl J Med*. 2005;352(19):1951-8. doi:10.1056/NEJMoa043012
- [382] Juhl CB, Lund H. Do we really need another systematic review? *British journal of sports medicine*. 2018;52(22):1408. doi:10.1136/bjsports-2018-099832
- [383] Kallini J, et al. Operative Treatment of Bipartite Patella in Pediatric and Adolescent Athletes: A Retrospective Comparison With a Nonoperatively Treated Cohort. *Orthopaedic Journal of Sports Medicine*. 2020;8(12):2325967120967125. doi:10.1177/2325967120967125

- [384] Kambhampati SBS, et al. Cyclops Lesions of the Knee: A Narrative Review of the Literature. *Orthopaedic Journal of Sports Medicine*. 2020;8(8)[doi:10.1177/2325967120945671]
- [385] Kang S-w, Xiang X. Physical activity and health services utilization and costs among U.S. adults. *Prev Med*. 2017;96:101-105. doi:https://doi.org/10.1016/j.ypmed.2016.12.043
- [386] Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*. 1958;53(282):457-481. doi:10.1080/01621459.1958.10501452
- [387] Karp N. Embracing variation to improve reproducibility. September 3rd, 2020. <https://blogs.bmj.com/openscience/2020/06/02/embracing-variation-to-improve-reproducibility/>
- [388] Kearns J, et al. A blood biomarker and clinical correlation cohort study protocol to diagnose sports-related concussion and monitor recovery in elite rugby. *BMJ Open Sport & Exercise Medicine*. 2020;6(1):e000948. doi:10.1136/bmjsem-2020-000948
- [389] Keil A, et al. A registered report format for Psychophysiology. *Psychophysiology*. 2020:e13663. doi:10.1111/psyp.13663
- [390] Kelley GA, Kelley KS. Impact of progressive resistance training on lipids and lipoproteins in adults: another look at a meta-analysis using prediction intervals. *Prev Med*. 2009;49(6):473-5. doi:10.1016/j.ypmed.2009.09.018
- [391] Kelley GA, Kelley KS. Exercise and sleep: a systematic review of previous meta-analyses. *Journal of evidence-based medicine*. 2017;10(1):26-36. doi:10.1111/jebm.12236
- [392] Kelley GA, Kelley KS. Systematic reviews and meta-analysis in nutrition research. *British Journal of Nutrition*. 2019;122(11):1279-1294. doi:10.1017/S0007114519002241
- [393] Kelley GA, Kelley KS. Systematic reviews and meta-analysis in nutrition research. *Br J Nutr*. 2019;122(11):1279-1294. doi:10.1017/S0007114519002241
- [394] Kelley GA, et al. Brief communication: use of the minimal important difference for a meta-analysis on exercise and anxiety in adults with arthritis. *Clinical Rheumatology*. 2018;37(7):1997-2000. doi:10.1007/s10067-018-4156-5
- [395] Kelley GA, et al. Brief Report: State-Level Number of Physically Inactive US Adults With Arthritis Who can Improve Their Anxiety and Depression by Exercising. *ACR open rheumatology*. 2020;2(2):92-96. doi:10.1002/acr2.11110
- [396] Kelly L, et al. Economic evaluation of 'Men on the Move', a 'real world' community-based physical activity programme for men. *European Journal of Public Health*. 2020;31(1):156-160. doi:10.1093/eurpub/ckaa152
- [397] Kemmler W, et al. Detraining effects after 18 months of high intensity resistance training on osteosarcopenia in older men-Six-month follow-up of the randomized controlled Franconian Osteopenia and Sarcopenia Trial (FrOST). *Bone*. 2021;142:115772. doi:10.1016/j.bone.2020.115772
- [398] Kerlinger FN, Lee HB. *Foundations of Behavioral Research*. 4 ed. Wadsworth Publishing; 1999.
- [399] Kingsnorth AP, et al. Changes in Device-Measured Physical Activity Patterns in U.K. Adults Related to the First COVID-19 Lockdown. *Journal for the Measurement of Physical Behaviour*. 2021;4(3):247-256. doi:10.1123/jmpb.2021-0005
- [400] Kinugasa T. The Application of Single-Case Research Designs to Study Elite Athletes' Conditioning: An Update. *Journal of Applied Sport Psychology*. 2013;25(1):157-166. doi:10.1080/10413200.2012.709578
- [401] Kinugasa T, et al. Single-subject research designs and data analyses for assessing elite athletes' conditioning. *Sports Med*. 2004;34(15):1035-50. doi:10.2165/00007256-200434150-00003
- [402] Kirk C, et al. The physical demands of mixed martial arts: A narrative review using the ARMSS model to provide a hierarchy of evidence. *J Sports Sci*. 2020;38(24):2819-2841. doi:10.1080/02640414.2020.1802093
- [403] Kirkendall DT, Krstrup P. Studying professional and recreational female footballers: A bibliometric exercise. *Scand J Med Sci Sports*. 2022;32 Suppl 1:12-26. doi:10.1111/sms.14019
- [404] Kiss A, et al. Structure and trends of international sport nutrition research between 2000 and 2018: bibliometric mapping of sport nutrition science. *J Int Soc Sports Nutr*. 2021;18(1):12. doi:10.1186/s12970-021-00409-5
- [405] Kneffel Z, et al. A meta-regression of the effects of resistance training frequency on muscular strength and hypertrophy in adults over 60 years of age. *J Sports Sci*. 2021;39(3):351-358. doi:10.1080/02640414.2020.1822595
- [406] Koenderman L, et al. Science and the social media: Time for a reset. *Eur J Clin Invest*. 2021;51(8):e13643. doi:10.1111/eci.13643
- [407] Koeppel M, et al. Trends in gross body coordination and cardiorespiratory fitness—a hierarchical Bayesian Analysis of 35 000 2nd Graders. *Scand J Med Sci Sports*. 2022;32(6):1026-1040. doi:10.1111/sms.14146
- [408] Konopka MJ, et al. Genetics of long-distance runners and road cyclists—A systematic review with meta-analysis. *Scand J Med Sci Sports*. 2022;32(10):1414-1429. doi:https://doi.org/10.1111/sms.14212
- [409] Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-63. doi:10.1016/j.jcm.2016.02.012
- [410] Ku PW, et al. Response to the letter titled "Double counting individuals in meta-analysis artificially inflates precision". *Scand J Med Sci Sports*. 2020;30(6):1085-1086. doi:10.1111/sms.13692
- [411] Kunze KN, et al. Corrigendum to "Risk Factors for Failure After Osteochondral Allograft Transplantation of the Knee: A Systematic Review and Exploratory Meta-analysis". *The American Journal of Sports Medicine*. 2022;03635465221082057. doi:10.1177/03635465221082057
- [412] Kwasnicka D, Naughton F. N-of-1 methods: A practical guide to exploring trajectories of behaviour change and designing precision behaviour change interventions. *Psychology of Sport & Exercise*. 2020;47:101570. doi:https://doi.org/10.1016/j.psychsport.2019.101570
- [413] Lachenbruch PA. A note on sample size computation for testing interactions. *Stat Med*. 1988;7(4):467-9. doi:10.1002/sim.4780070403
- [414] Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*. 2013;4doi:10.3389/fpsyg.2013.00863
- [415] Lam K, et al. How effective and how expensive are interventions to reduce sedentary behavior? An umbrella review and meta-analysis. *Obes Rev*. 2022:e13422. doi:https://doi.org/10.1111/obr.13422
- [416] Lancaster GA, et al. Design and analysis of pilot studies: recommendations for good practice. *Journal of Evaluation in Clinical Practice*. 2004;10(2):307-312. doi:https://doi.org/10.1111/j.2002.384.doc.x
- [417] Langeard A, et al. High-intensity interval training vs. hydrochlorothiazide on blood pressure, cardiovascular health and cognition: Protocol of a non-inferiority trial. *Contemp Clin Trials*. 2021;102:106286. doi:10.1016/j.cct.2021.106286
- [418] Laranjo L, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. *British journal of sports medicine*. 2020;doi:10.1136/bjsports-2020-102892
- [419] Lavoué C, et al. Analysis of food and fluid intake in elite ultra-endurance runners during a 24-h world championship. *J Int Soc Sports Nutr*. 2020;17(1):36. doi:10.1186/s12970-020-00364-7
- [420] Lavoué C, et al. Correction to: Analysis of food and fluid intake in elite ultra-endurance runners during a 24-h world championship. *J Int Soc Sports Nutr*. 2021;18(1):32. doi:10.1186/s12970-021-00427-3
- [421] Lee DC, et al. Long-term effects of changes in cardiorespiratory fitness and body mass index on all-cause and cardiovascular disease mortality in men: the Aerobics Center Longitudinal Study. *Circulation*. 2011;124(23):2483-90. doi:10.1161/circulationaha.111.038422
- [422] Lee K, et al. Many Papers but Limited Policy Impact? A Bibliometric Review of Physical Activity Research. *Translational Journal of the American College of Sports Medicine*. 2021;6(4):e000167. doi:10.1249/tjx.0000000000000167
- [423] Leow CHW, et al. Cultural differences in hydration practices among physically active individuals: a narrative review. *J Int Soc Sports Nutr*. 2022;19(1):150-163. doi:10.1080/15502783.2022.2057196
- [424] Lepley LK, et al. Eccentric Exercise to Enhance Neuromuscular Control. *Sports Health*. 2017;9(4):333-340. doi:10.1177/1941738117710913

- [425] Levine TR, Hullett CR. Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. *Human Communication Research*. 2006;28(4):612-625. doi:10.1111/j.1468-2958.2002.tb00828.x
- [426] Lewin S, et al. Using Qualitative Evidence in Decision Making for Health and Social Interventions: An Approach to Assess Confidence in Findings from Qualitative Evidence Syntheses (GRADE-CERQual). *PLOS Medicine*. 2015;12(10):e1001895. doi:10.1371/journal.pmed.1001895
- [427] Lewis SJ. Quantifying measurement error. In: Anderson S, ed. *Current and recent research in osteoarchaeology 2: Proceedings of the 4th, 5th and 6th meetings of the Osteoarchaeological Research Group*. 1999:54-55.
- [428] Li J, et al. Breathing exercises in people with COPD: A realist review. *J Adv Nurs*. 2021;77(4):1698-1715. doi:10.1111/jan.14703
- [429] Li X, Li Y. Sports Training Strategies and Interactive Control Methods Based on Neural Network Models. *Comput Intell Neurosci*. 2022;2022:7624578. doi:10.1155/2022/7624578
- [430] Liang X, et al. Hierarchical Bayes Approach to Estimate the Treatment Effect for Randomized Controlled Trials. *Educational and Psychological Measurement*. 2020;80(6):1090-1114. doi:10.1177/0013164420909885
- [431] Liao C-D, et al. Exercise therapy for sarcopenia in rheumatoid arthritis: A meta-analysis and meta-regression of randomized controlled trials. *Clinical rehabilitation*. 2022;36(2):145-157. doi:10.1177/02692155211035539
- [432] Lin LI. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*. 1989;45(1):255-268. doi:10.2307/2532051
- [433] Lindheimer JB, et al. Quantifying the placebo effect in psychological outcomes of exercise training: a meta-analysis of randomized trials. *Sports Med*. 2015;45(5):693-711. doi:10.1007/s40279-015-0303-1
- [434] Lindheimer JB, et al. Advancing the understanding of placebo effects in psychological outcomes of exercise: Lessons learned and future directions. *Eur J Sport Sci*. 2020;20(3):326-337. doi:10.1080/17461391.2019.1632937
- [435] Lindheimer JB, et al. Reconceptualizing the measurement of expectations to better understand placebo and nocebo effects in psychological responses to exercise. *Eur J Sport Sci*. 2020;20(3):338-346. doi:10.1080/17461391.2019.1674926
- [436] Liu J, et al. Image Recognition and Extraction of Students' Human Motion Features Based on Graph Neural Network. *Comput Intell Neurosci*. 2022;2022:6755053. doi:10.1155/2022/6755053
- [437] Liu Q, et al. Reliability and validity of Healthy Fitness Measurement Scale Version 1.0 (HFMS V1.0) in Chinese elderly people. *BMC Public Health*. 2021;21(1):1019. doi:10.1186/s12889-021-11021-2
- [438] Lohbeck A, et al. Children's Motivation Profiles in Sports and Physical Activities: A Latent Profile Analysis and Self-Determination Theory Approach. *J Sport Exerc Psychol*. 2022:1-12. doi:10.1123/jsep.2021-0279
- [439] Longmuir PE, et al. Canadian Agility and Movement Skill Assessment (CAMSA): Validity, objectivity, and reliability evidence for children 8–12 years of age. *J Sport Health Sci*. 2017;6(2):231-240. doi:https://doi.org/10.1016/j.jshs.2015.11.004
- [440] Looney MA. Assessment of interrater and intermethod agreement in the kinesiology literature. *Measurement in Physical Education and Exercise Science*. 2018;22(2):116-128. doi:10.1080/1091367x.2017.1395742
- [441] Lopes JSS, et al. Etiology, prevalence, and severity of reported acute sports injuries in Brazilian Jiu-Jitsu Paradesports: An observational study. *Science & Sports*. 2021;36(2):e43-e50. doi:https://doi.org/10.1016/j.scispo.2020.03.009
- [442] Lopes L, et al. A Narrative Review of Motor Competence in Children and Adolescents: What We Know and What We Need to Find Out. *Int J Environ Res Public Health*. 2020;18(1):doi:10.3390/ijerph18010018
- [443] López-Torres Hidalgo J. Effectiveness of physical exercise in the treatment of depression in older adults as an alternative to antidepressant drugs in primary care. *BMC Psychiatry*. 2019;19(1):21. doi:10.1186/s12888-018-1982-6
- [444] Lopez P, et al. Resistance Training Load Effects on Muscle Hypertrophy and Strength Gain: Systematic Review and Network Meta-analysis. *Med Sci Sports Exerc*. 2021;53(6):1206-1216. doi:10.1249/mss.0000000000002585
- [445] Lopez P, et al. Resistance Exercise Dosage in Men with Prostate Cancer: Systematic Review, Meta-analysis, and Meta-regression. *Med Sci Sports Exerc*. 2020;doi:10.1249/MSS.0000000000002503
- [446] Losada JL. *Metodología observacional*. Penta; 2000.
- [447] Louro-González A, et al. ¿Cómo hacer una Revisión Clínica con fuentes MBE? Accessed 7/1/22, 2021. <https://www.fisterra.com/guias-clinicas/como-hacer-revision-clinica-con-fuentes-mbe/>
- [448] Lowe NK. What Is a Pilot Study? *Journal of Obstetric, Gynecologic & Neonatal Nursing*. 2019;48(2):117-118. doi:10.1016/j.jogn.2019.01.005
- [449] Ludbrook J. Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clinical and Experimental Pharmacology and Physiology*. 2002;29(7):527-536.
- [450] Lusted LB. Signal detectability and medical decision-making. *Science*. 1971;171(3977):1217-9. doi:10.1126/science.171.3977.1217
- [451] Luteberget LS, Gilgien M. Validation methods for global and local positioning-based athlete monitoring systems in team sports: a scoping review. *BMJ Open Sport & Exercise Medicine*. 2020;6(1):doi:10.1136/bmjsem-2020-000794
- [452] Mach M, et al. Transformational leadership and team performance in sports teams: A conditional indirect model. *Applied Psychology*. 2021;doi:https://doi.org/10.1111/apps.12342
- [453] Machado LDS, et al. Pelvic floor evaluation in CrossFit® athletes and urinary incontinence: a cross-sectional observational study. *Women Health*. 2021;61(5):490-499. doi:10.1080/03630242.2021.1927288
- [454] Mack MG, Ragan BG. Development of the mental, emotional, and bodily toughness inventory in collegiate athletes and nonathletes. *J Athl Train*. 2008;43(2):125-132. doi:10.4085/1062-6050-43.2.125
- [455] Magnusson K, et al. High genetic contribution to anterior cruciate ligament rupture: Heritability ~69%. *British journal of sports medicine*. 2020;bjsports-2020-102392. doi:10.1136/bjsports-2020-102392
- [456] Magro-Malosso ER, et al. Exercise during pregnancy and risk of gestational hypertensive disorders: a systematic review and meta-analysis. *Acta Obstet Gynecol Scand*. 2017;96(8):921-931. doi:10.1111/aogs.13151
- [457] Maher CG, et al. Reliability of the PEDro scale for rating quality of randomized controlled trials. *Phys Ther*. 2003;83(8):713-21.
- [458] Maldonado-Villamizar J, Aquino-Guerra A. Experimentación con biomodelos animales en ciencias de la salud. *Avances en Biomedicina*. 2016;5(3):173-177.
- [459] Manley H, et al. Profiles of psychological skill use and their relations with self- and coach-rated mental toughness in Thai athletes. *Sport, Exercise, and Performance Psychology*. 2020;9(4):558-570. doi:10.1037/spy0000200
- [460] Mann T. 'Mean response'disregards the importance of individual variation. *South African Journal of Sports Medicine*. 2011;23(1):30.
- [461] Mansournia MA, et al. A Checklist for statistical Assessment of Medical Papers (the CHAMP statement): explanation and elaboration. *British journal of sports medicine*. 2021;55(18):1009-1017. doi:10.1136/bjsports-2020-103652
- [462] Mansournia MA, et al. CChecklist for statistical Assessment of Medical Papers: the CHAMP statement. *British journal of sports medicine*. 2021;55(18):1002-1003. doi:10.1136/bjsports-2020-103651
- [463] Manterola C, Otzen T. Estudios observacionales: los diseños utilizados con mayor frecuencia en investigación clínica. *International Journal of Morphology*. 2014;32(2):634-645.
- [464] Manterola C, et al. Metodología de los tipos y diseños de estudio más frecuentemente utilizados en investigación clínica. *Revista Médica Clínica Las Condes*. 2019;30(1):36-49. doi:https://doi.org/10.1016/j.rmcl.2018.11.005
- [465] Marda V. Machine learning and transparency: a scoping exercise. *Social Science Research Network* doi:http://dx.doi.org/10.2139/ssrn.3236837
- [466] Marsh CE, et al. Fitness and strength responses to distinct exercise modes in twins: Studies of Twin Responses to Understand Exercise as a Therapy (STRUETH) study. *J Physiol*. 2020;doi:10.1113/JP280048
- [467] Marshall PWM, et al. Physical activity and the mediating effect of fear, depression, anxiety, and catastrophizing on pain related

- disability in people with chronic low back pain. *PLoS One*. 2017;12(7):e0180788. doi:10.1371/journal.pone.0180788
- [468] Martínez-Miranda MD. Análisis Cluster en SPSS. Accessed 10/5/21, 2021. <http://www.ugr.es/~cursps/archivos/Cluster/cluster>
- [469] Mathur MB, VanderWeele TJ. Meta-regression methods to characterize evidence strength using meaningful-effect percentages conditional on study characteristics. *Research Synthesis Methods*. 2021;12(6):731-749. doi:<https://doi.org/10.1002/jrsm.1504>
- [470] Matsel KA, et al. Current Concepts in Arm Care Exercise Programs and Injury Risk Reduction in Adolescent Baseball Players: A Clinical Review. *Sports Health*. 2021;13(3):245-250. doi:10.1177/1941738120976384
- [471] Mattle S, et al. Feasibility of Hypnosis on Performance in Air Rifle Shooting Competition. *International Journal of Clinical and Experimental Hypnosis*. 2020;68(4):521-529. doi:10.1080/00207144.2020.1799655
- [472] Maxwell SE, et al. Sample Size Planning for Statistical Power and Accuracy in Parameter Estimation. *Annual Review of Psychology*. 2008;59(1):537-563. doi:10.1146/annurev.psych.59.103006.093735
- [473] Mazzolari R, et al. Myths and methodologies: The use of equivalence and non-inferiority tests for interventional studies in exercise physiology and sport science. *Exp Physiol*. 2022;107:201-212. doi:<https://doi.org/10.1113/EP090171>
- [474] McBride GB. *Using Statistical Methods for Water Quality Management: Issues, Problems and Solutions*. Wiley; 2005.
- [475] McGee SL, Hargreaves M. Epigenetics and Exercise. *Trends Endocrinol Metab*. 2019;30(9):636-645. doi:10.1016/j.tem.2019.06.002
- [476] McGraw KO, Wong SP. A common language effect size statistic. *Psychol Bull*. 1992;111(2):361.
- [477] McHugh ML. Interrater reliability: the kappa statistic. *Biochemia medica*. 2012;22(3):276-282.
- [478] McKenzie DP, Thomas C. Relative risks and odds ratios: Simple rules on when and how to use them. *Eur J Clin Invest*. 2020:e13249. doi:10.1111/eci.13249
- [479] McKeown S, Mir ZM. Considerations for conducting systematic reviews: evaluating the performance of different methods for de-duplicating references. *Systematic Reviews*. 2021;10(1):38. doi:10.1186/s13643-021-01583-y
- [480] McMahon FG. Placebos in medicine. *Lancet*. 1994;344(8937):1641.
- [481] McNair PJ, et al. Verbal encouragement: effects on maximum effort voluntary muscle action. *British journal of sports medicine*. 1996;30(3):243-5. doi:10.1136/bjsem.30.3.243
- [482] McWilliams M. *Foods: Experimental perspectives*. Prentice Hall; 2001.
- [483] Medicine and Science in Sport and Exercise Editor. Evidence for an Upper Threshold for Resistance Training Volume in Trained Women-Expression of Concern. *Med Sci Sports Exerc*. 2020;52(11):2490. doi:10.1249/mss.0000000000002496
- [484] Medicine and Science in Sport and Exercise Editor. Upper Threshold for Resistance Training Volume in Trained Women-RETRACTION. *Med Sci Sports Exerc*. 2021;53(6):1318. doi:10.1249/mss.0000000000002672
- [485] Mekary RA, Ding EL. Isotemporal Substitution as the Gold Standard Model for Physical Activity Epidemiology: Why It Is the Most Appropriate for Activity Time Research. *Int J Environ Res Public Health*. 2019;16(5):797.
- [486] Mekary RA, et al. Isotemporal substitution analysis for physical activity, television watching, and risk of depression. *American journal of epidemiology*. 2013;178(3):474-83. doi:10.1093/aje/kws590
- [487] Mekary RA, et al. Isotemporal substitution paradigm for physical activity epidemiology and weight change. *American journal of epidemiology*. 2009;170(4):519-27. doi:10.1093/aje/kwp163
- [488] Memon AR, et al. Research Combining Physical Activity and Sleep: A Bibliometric Analysis. *Perceptual and motor skills*. 2020;127(1):154-181. doi:10.1177/0031512519889780
- [489] Mercier MA, et al. Individual factors associated with baseball pitching performance: scoping review. *BMJ Open Sport & Exercise Medicine*. 2020;6(1):e000704. doi:10.1136/bmjsem-2019-000704
- [490] Metcalfe RS, et al. Time-efficient and computer-guided sprint interval exercise training for improving health in the workplace: a randomised mixed-methods feasibility study in office-based employees. *BMC Public Health*. 2020;20(1):313. doi:10.1186/s12889-020-8444-z
- [491] Metcalfe RS, Vollaard NBJ. Heterogeneity and incidence of non-response for changes in cardiorespiratory fitness following time-efficient sprint interval exercise training. *Applied Physiology, Nutrition & Metabolism*. 2021;doi:10.1139/apnm-2020-0855
- [492] Miller DK. *Measurement by the physical educator: Why and how*. 3rd ed. WCB McGraw-Hill; 1996.
- [493] Milroy J, et al. A Latent Profile Analysis to Categorize and Describe Athletes Based on Factors Related to Concussion Disclosure. *J Athl Train*. 2021;56(1):85-91. doi:10.4085/1062-6050-0308.19
- [494] Miočević M, et al. Causal Mediation Analysis in Single Case Experimental Designs: Introduction to the Special Issue. *Evaluation & the Health Professions*. 2022;45(1):3-7. doi:10.1177/01632787211073194
- [495] Moctezuma Viera KR, Linares Márquez P. Utilización de animales en la investigación biomédica y médica: un estudio preliminar. *Revista Iberoamericana de Bioética*. 2020;(12):11.
- [496] Molina Arias M. El metaanálisis en red. *Pediatría Atención Primaria*. 2016;18:279-282.
- [497] Møller AM, Myles PS. What makes a good systematic review and meta-analysis? *British journal of anaesthesia*. 2016;117(4):428-430. doi:10.1093/bja/aew264
- [498] Molloy EJ, Bearer CF. When research goes wrong: the importance of clinical trials methodology. *Pediatric Research*. 2020;doi:10.1038/s41390-020-0984-5
- [499] Moncada-Jiménez J. Métodos estadísticos utilizados en las ciencias del movimiento humano. *Revista Educación*. 2004;28(2):279-287.
- [500] Moncada-Jiménez J. 2005: Año Internacional del Deporte y la Educación Física. *Revista Educación*. 2005;29(2):233-247.
- [501] Moncada-Jiménez J. *Estadística: para las ciencias del movimiento humano*. San José, Costa Rica. Editorial de la Universidad de Costa Rica; 2005.
- [502] Moncada-Jiménez J. Metodología de la investigación: el enfoque meta-analítico para la síntesis cuantitativa de la literatura en las ciencias del movimiento humano. *Actualidades Investigativas en Educación*. 2006;6(1):1-25.
- [503] Moncada-Jiménez J. Revisiones de literatura, revisiones sistemáticas de literatura y meta análisis. In: Moncada Jiménez J, ed. *Los principales efectos del ejercicio y el entrenamiento físico en el desarrollo, el rendimiento y la salud humana*. Editorial de la Universidad de Costa Rica; 2013:15-26.
- [504] Moncada-Jiménez J, Lachowetz A. Escalas alométricas: un justo criterio en el proceso de evaluación en las ciencias del movimiento humano. *Revista Educación*. 2001;25(1):95-102.
- [505] Moncada-Jiménez J, et al. La recolección, utilización y almacenamiento de datos biométricos sensibles en deportistas: insumos para la carrera de Educación Física. *Revista Educación*. 2021;45(1):640-652.
- [506] Moncada-Jiménez J, et al. Collection, utilization and storage of athletes' sensitive biometric data: Investment for the Physical Education career. *Wearable Technology*. 2022;3(1):96-104.
- [507] Moncada-Jiménez J, et al. ¿Cómo evaluar estadísticamente los puntajes de métodos distintos que miden el mismo constructo? *Revista MHSalud*. 2021:1-11.
- [508] Moncada-Jiménez J, et al. Fuentes de varianza e índices de varianza explicada en las ciencias del movimiento humano. *Revista de Ciencias del Ejercicio y la Salud*. 2002;2(2):70-74.
- [509] Moniz SC, et al. Mechanistic and methodological perspectives on the impact of intense interval training on post-exercise metabolism. *Scand J Med Sci Sports*. 2020;30(4):638-651. doi:10.1111/sms.13610
- [510] Montenegro-Montenegro E, et al. Do not Be Afraid of Missing Data: Modern Approaches to Handle Missing Information. *Actualidades en Psicología*. 2015;29(119):29-42. doi:<https://doi.org/10.15517/ap.v29i119.18812>
- [511] Montero D, Lundby C. Refuting the myth of non-response to exercise training: 'non-responders' do respond to higher dose of training. *J Physiol*. 2017;595(11):3377-3387. doi:10.1113/jp273480
- [512] Moreno-Pérez V, et al. Intra and inter-tester reliability of a novel device to assess gluteal muscle strength in professional football players. *Research in Sports Medicine*. 2022;30(2):156-168. doi:10.1080/15438627.2020.1868466

- [513] Morera-Siercovich PL, Moncada-Jiménez J. A Meta-Analysis of the Effect of Small-Sided Games on the Repeated Sprint Ability in Soccer. *International Journal of Physical Education, Fitness and Sports*. 2020;9(4):70-84. doi:10.34256/ijpefs2048
- [514] Morishima S, et al. Intensive, prolonged exercise seemingly causes gut dysbiosis in female endurance runners. *J Clin Biochem Nutr*. 2021;68(3):253-258. doi:10.3164/jcbn.20-131
- [515] Morrow JR, et al. *Measurement and Evaluation in Human Performance*. 4th ed. Human Kinetics; 2011.
- [516] Moulds K, et al. Sink or Swim? A survival analysis of sport dropout in Australian youth swimmers. *Scand J Med Sci Sports*. 2020;30(11):2222-2233. doi:10.1111/sms.13771
- [517] Muniz-Pardos B, et al. Ethical dilemmas and validity issues related to the use of new cooling technologies and early recognition of exertional heat illness in sport. *BMJ Open Sport & Exercise Medicine*. 2021;7(2):e001041. doi:10.1136/bmjsem-2021-001041
- [518] Munn Z, et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*. 2018;18(1):143. doi:10.1186/s12874-018-0611-x
- [519] Munn Z, et al. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Medical Research Methodology*. 2018;18(1):5. doi:10.1186/s12874-017-0468-4
- [520] Munoz CX, et al. Impact of Nutrient Intake on Hydration Biomarkers Following Exercise and Rehydration Using a Clustering-Based Approach. *Nutrients*. 2020;12(5)doi:10.3390/nu12051276
- [521] Munthe-Kaas H, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 3: how to assess methodological limitations. *Implement Sci*. 2018;13(Suppl 1):9. doi:10.1186/s13012-017-0690-9
- [522] Myrcik D, et al. Influence of Physical Activity on Pain, Depression and Quality of Life of Patients in Palliative Care: A Proof-of-Concept Study. *J Clin Med*. 2021;10(5)doi:10.3390/jcm10051012
- [523] Naci H, et al. How does exercise treatment compare with antihypertensive medications? A network meta-analysis of 391 randomised controlled trials assessing exercise and medication effects on systolic blood pressure. *British journal of sports medicine*. 2019;53(14):859-869. doi:10.1136/bjsports-2018-099921
- [524] Naranjo Orellana J, Muela Galán S. Retrograde extrapolation of VO₂max from recovery values recorded breath by breath. *Retos*. 2021;41:695-700. doi:https://doi.org/10.47197/retos.v41i0.84525
- [525] Nassiss GP, et al. Elite football of 2030 will not be the same as that of 2020: Preparing players, coaches, and support staff for the evolution. *Scand J Med Sci Sports*. 2020;30(6):962-964. doi:10.1111/sms.13681
- [526] National Heart Lung and Blood Institute. Study Quality Assessment Tools. US Department of Health and Human Services. Accessed Nov 1st, 2021. <https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools>
- [527] Navarro DJ, Foxcroft DR. *Learning statistics with jamovi: a tutorial for psychology students and other beginners (Version 0.75)*. 2019.
- [528] Nevill AM, et al. The dose-response association between VO₂peak and self-reported physical activity in children. *J Sports Sci*. 2020;38(16):1829-1835. doi:10.1080/02640414.2020.1756682
- [529] Nevill AM, et al. Scaling physiological measurements for individuals of different body size. *European Journal of Applied Physiology and Occupational Physiology*. 1992;65(2):110-117. doi:10.1007/BF00705066
- [530] Nevill AM, et al. Are adult physiques geometrically similar? The dangers of allometric scaling using body mass power laws. *American Journal of Physical Anthropology*. 2004;124(2):177-182. doi:https://doi.org/10.1002/ajpa.10351
- [531] Ng YL, et al. Effectiveness of Outdoor Exercise Parks on Health Outcomes in Older Adults-A Mixed-Methods Systematic Review and Meta-Analysis. *J Aging Phys Act*. 2021;29(4):695-707. doi:10.1123/japa.2020-0031
- [532] Nguyen P, et al. Economics of sedentary behaviour: A systematic review of cost of illness, cost-effectiveness, and return on investment studies. *Prev Med*. 2022;156:106964. doi:https://doi.org/10.1016/j.ypmed.2022.106964
- [533] Nielsen K, Miraglia M. What works for whom in which circumstances? On the need to move beyond the 'what works?' question in organizational intervention research. *Human Relations*. 2017;70(1):40-62. doi:10.1177/0018726716670226
- [534] Nielsen RO, et al. Methods matter and the 'too much, too soon' theory (part 2): what is the goal of your sports injury research? Are you describing, predicting or drawing a causal inference? *British journal of sports medicine*. 2020;54(22):1307-1309. doi:10.1136/bjsports-2020-102144
- [535] Notley SR, et al. Exercise Thermoregulation in Prepubertal Children: A Brief Methodological Review. *Med Sci Sports Exerc*. 2020;52(11):2412-2422. doi:10.1249/mss.0000000000002391
- [536] Nummela A, et al. Variability in hemoglobin mass response to altitude training camps. *Scand J Med Sci Sports*. 2021;31(1):44-51. doi:https://doi.org/10.1111/sms.13804
- [537] Nunes A, et al. Effects of Integrative Neuromuscular Training and Detraining on Countermovement Jump Performance in Youth Volleyball Players. *Journal of strength and conditioning research*. 2021;35(8):2242-2247. doi:10.1519/jsc.0000000000003092
- [538] O'Donoghue G, et al. What exercise prescription is optimal to improve body composition and cardiorespiratory fitness in adults living with obesity? A network meta-analysis. *Obes Rev*. 2021;22(2):e13137. doi:10.1111/obr.13137
- [539] O' Cathain A, et al. Three techniques for integrating data in mixed methods studies. *BMJ (Clinical research ed)*. 2010;341:c4587. doi:10.1136/bmj.c4587
- [540] Okada M. Big data and real-world data-based medicine in the management of hypertension. *Hypertension Research*. 2021;44(2):147-153. doi:10.1038/s41440-020-00580-3
- [541] Orozco-Díaz A, et al. Allelic and genotypic frequencies of actinin 3 gene (ACTN3) in the Costa Rican population and its association with physical fitness in college students. *Journal of Biology of Exercise*. 2019;15(1):1-12. doi:https://doi.org/10.4127/jbe.2019.0142
- [542] Owen PJ, et al. Which specific modes of exercise training are most effective for treating low back pain? Network meta-analysis. *British journal of sports medicine*. 2020;54(21):1279. doi:10.1136/bjsports-2019-100886
- [543] Pace R, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *Int J Nurs Stud*. 2012;49(1):47-53. doi:10.1016/j.ijnurstu.2011.07.002
- [544] Padilla J, et al. Identifying responders versus non-responders: Incorporation of controls is required for sound statistical inference. *Exp Physiol*. 2021;106:375-376. doi:https://doi.org/10.1113/EP089142
- [545] Paffenbarger RS, et al. Physical Activity, All-Cause Mortality, and Longevity of College Alumni. *New England Journal of Medicine*. 1986;314(10):605-613. doi:10.1056/nejm198603063141003
- [546] Pal A, et al. No Evidence for Effect of Exercise on Transcriptome of NK Cells in Breast Cancer Patients Undergoing Adjuvant Therapy: Results From a Pilot Study. *Frontiers in Physiology*. 2019;10(959)doi:10.3389/fphys.2019.00959
- [547] Papageorgiou G, et al. Statistical primer: how to deal with missing data in scientific research?†. *Interactive Cardiovascular and Thoracic Surgery*. 2018;27(2):153-158. doi:10.1093/icvts/ivy102
- [548] Papakonstantinou T, et al. CINEMA: Software for semiautomated assessment of the confidence in the results of network meta-analysis. *Campbell Systematic Reviews*. 2020;16(1):e1080. doi:https://doi.org/10.1002/cl2.1080
- [549] Papic C, et al. Improving data acquisition speed and accuracy in sport using neural networks. *J Sports Sci*. 2021;39(5):513-522. doi:10.1080/02640414.2020.1832735
- [550] Park HY, et al. Estimated Artificial Neural Network Modeling of Maximal Oxygen Uptake Based on Multistage 10-m Shuttle Run Test in Healthy Adults. *Int J Environ Res Public Health*. 2021;18(16)doi:10.3390/ijerph18168510
- [551] Park J, et al. Real-Time Exercise Feedback through a Convolutional Neural Network: A Machine Learning-Based Motion-Detecting Mobile Exercise Coaching Application. *Yonsei Med J*. 2022;63(Suppl):S34-s42. doi:10.3349/ymj.2022.63.S34
- [552] Patrician PA. Multiple imputation for missing data. *Res Nurs Health*. 2002;25(1):76-84. doi:10.1002/nur.10015
- [553] Pawlowsky-Glahn V, et al. *Modeling and analysis of compositional data*. John Wiley & Sons; 2015.
- [554] Pawson R, et al. Realist review - a new method of systematic review designed for complex policy interventions. *Journal of Health*

- Services Research & Policy.* 2005;10(1_suppl):21-34. doi:10.1258/1355819054308530
- [555] Peralta-Brenes M, et al. The chronic effect of exercise on phase angle in non-athletes: a systematic review with meta-analysis of experimental studies. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021197966
- [556] Perry CG, et al. Repeated transient mRNA bursts precede increases in transcriptional and mitochondrial proteins during training in human skeletal muscle. *J Physiol.* 2010;588(Pt 23):4795-810. doi:10.1113/jphysiol.2010.199448
- [557] Pérusse L, et al. Advances in exercise, fitness, and performance genomics in 2012. *Med Sci Sports Exerc.* 2013;45(5):824-31. doi:10.1249/MSS.0b013e31828b28a3
- [558] Pérusse L, et al. The human gene map for performance and health-related fitness phenotypes: the 2002 update. *Med Sci Sports Exerc.* 2003;35(8):1248-64. doi:10.1249/01.Mss.0000078938.84161.22
- [559] Pescatello LS, et al. Do the combined blood pressure effects of exercise and antihypertensive medications add up to the sum of their parts? A systematic meta-review. *BMJ Open Sport & Exercise Medicine.* 2021;7(1):e000895. doi:10.1136/bmjsem-2020-000895
- [560] Peters MDJ, et al. Scoping Reviews (2020 version). In: Aromataris E, Munn Z, eds. *JBI Manual for Evidence Synthesis, JBI, 2020.* Joanna Briggs Institute; 2020:chap 11.
- [561] Peters MDJ, et al. Scoping reviews: reinforcing and advancing the methodology and application. *Systematic Reviews.* 2021;10(1):263. doi:10.1186/s13643-021-01821-3
- [562] Pham B, et al. Text mining to support abstract screening for knowledge syntheses: a semi-automated workflow. *Systematic Reviews.* 2021;10(1):156. doi:10.1186/s13643-021-01700-x
- [563] Phillips BE, et al. A Practical and Time-Efficient High-Intensity Interval Training Program Modifies Cardio-Metabolic Risk Factors in Adults with Risk Factors for Type II Diabetes. *Front Endocrinol (Lausanne).* 2017;8:229. doi:10.3389/fendo.2017.00229
- [564] Phillips MR, et al. A clinician's guide to network meta-analysis. *Eye.* 2022;doi:10.1038/s41433-022-01943-5
- [565] Pickering C, Kiely J. Do Non-Responders to Exercise Exist and If So, What Should We Do About Them? *Sports Med.* 2019;49(1):1-7. doi:10.1007/s40279-018-01041-1
- [566] Piedra A, et al. Monitoring Training Loads in Basketball: A Narrative Review and Practical Guide for Coaches and Practitioners. *Strength & Conditioning Journal.* 2021;43(5):12-35. doi:10.1519/ssc.0000000000000620
- [567] Pieper D, Rombey T. Where to prospectively register a systematic review. *Systematic Reviews.* 2022;11(1):8. doi:10.1186/s13643-021-01877-1
- [568] Pierce KC, et al. Weightlifting for Children and Adolescents: A Narrative Review. *Sports Health.* 2021;19417381211056094. doi:10.1177/19417381211056094
- [569] Pillon NJ, et al. Transcriptomic profiling of skeletal muscle adaptations to exercise and inactivity. *Nat Commun.* 2020;11(1):470. doi:10.1038/s41467-019-13869-w
- [570] Pinheiro LSP, et al. Prevalence and incidence of injuries in para athletes: a systematic review with meta-analysis and GRADE recommendations. *British journal of sports medicine.* 2021;55(23):1357-1365. doi:10.1136/bjsports-2020-102823
- [571] Podsiadlo D, Richardson S. The Timed "Up & Go": A Test of Basic Functional Mobility for Frail Elderly Persons. *Journal of the American Geriatrics Society.* 1991;39(2):142-148. doi:https://doi.org/10.1111/j.1532-5415.1991.tb01616.x
- [572] Polli A, et al. When Environment Meets Genetics: A Clinical Review of the Epigenetics of Pain, Psychological Factors, and Physical Activity. *Arch Phys Med Rehabil.* 2019;100(6):1153-1161. doi:10.1016/j.apmr.2018.09.118
- [573] Ponnusamy V, et al. Latent profiles of elite Malaysian athletes' use of psychological skills and techniques and relations with mental toughness. *PeerJ.* 2018;6:e4778. doi:10.7717/peerj.4778
- [574] Popay J, et al. Guidance on the conduct of narrative synthesis in systematic reviews. *A product from the ESRC methods programme Version.* 2006;1:b92.
- [575] Popoff E, et al. Aligning text mining and machine learning algorithms with best practices for study selection in systematic literature reviews. *Systematic Reviews.* 2020;9(1):293. doi:10.1186/s13643-020-01520-5
- [576] Porrati-Paladino G, Cuesta-Barriuso R. Effectiveness of Plyometric and Eccentric Exercise for Jumping and Stability in Female Soccer Players-A Single-Blind, Randomized Controlled Pilot Study. *Int J Environ Res Public Health.* 2021;18(1):doi:10.3390/ijerph18010294
- [577] Portell Vidal M, et al. Guidelines for reporting evaluations based on observational methodology. *Psicothema.* 2015;27(3):283-289.
- [578] Porter AL, et al. Research profiling: Improving the literature review. *Scientometrics.* 2002;53(3):351. doi:10.1023/a:1014873029258
- [579] Poyatos-León R, et al. Effects of exercise during pregnancy on mode of delivery: a meta-analysis. *Acta Obstet Gynecol Scand.* 2015;94(10):1039-47. doi:10.1111/aogs.12675
- [580] Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods.* 2008;40(3):879-891. doi:10.3758/BRM.40.3.879
- [581] Prieto G, Delgado AR. Fiabilidad y validez. *Papeles del psicólogo.* 2010;31(1):67-74.
- [582] Prud'homme D, et al. Sensitivity of maximal aerobic power to training is genotype-dependent. *Med Sci Sports Exerc.* 1984;16(5):489-93. doi:10.1249/00005768-198410000-00012
- [583] Pyne D. The Art and Practice of Presenting Results. *Int J Sports Physiol Perform.* 2007;2:345-346.
- [584] Pyne D. Preparing a High-Quality and Impactful Sport Science Manuscript. *Int J Sports Physiol Perform.* 2020;15(5):598-599. doi:10.1123/ijsp.2020-0129
- [585] Qin P, Feng W. Design of the Exercise Load Data Monitoring System for Exercise Training Based on the Neural Network. *J Healthc Eng.* 2021;2021:7340140. doi:10.1155/2021/7340140
- [586] *R: A language and environment for statistical computing.* Version 4.0.3. R Foundation for Statistical Computing; 2020. <https://www.r-project.org/>
- [587] Raglin J, et al. Understanding placebo and nocebo effects in the context of sport: A psychological perspective. *Eur J Sport Sci.* 2020;20(3):293-301. doi:10.1080/17461391.2020.1727021
- [588] Rahlf AL, Zech A. Comparison of 10 vs. 20 min neuromuscular training for the prevention of lower extremity injuries in male youth football: A cluster randomised controlled trial. *J Sports Sci.* 2020;38(19):2177-2185. doi:10.1080/02640414.2020.1776459
- [589] Raiteri BJ. The unknowable probability of replication. *J Physiol.* 2021;doi:https://doi.org/10.1113/JP281472
- [590] Ramadan FA, et al. Association of sedentary and physical activity behaviours with body composition: a genome-wide association and Mendelian randomisation study. *BMJ Open Sport & Exercise Medicine.* 2022;8(3):e001291. doi:10.1136/bmjsem-2021-001291
- [591] Ramirez-Campillo R, et al. Methodological characteristics and future directions for plyometric jump training research: A scoping review update. *Scand J Med Sci Sports.* 2020;30(6):983-997. doi:10.1111/sms.13633
- [592] Ramkumar PN, et al. Sports Medicine and Artificial Intelligence: A Primer. *Am J Sports Med.* 2021;3635465211008648. doi:10.1177/03635465211008648
- [593] Ramos-Vera CA, Ogundokun RO. The use of the Bayes factor for statistical inference. *Journal of Human Sport and Exercise.* 2021;16(2):473-475.
- [594] Rankin A, et al. Infographic. COVID-19 RT-PCR testing for elite athletes. *British journal of sports medicine.* 2021;55(14):818-820. doi:10.1136/bjsports-2020-103751
- [595] Rankinen T, et al. The human gene map for performance and health-related fitness phenotypes: the 2005 update. *Med Sci Sports Exerc.* 2006;38(11):1863-88. doi:10.1249/01.mss.0000233789.01164.4f
- [596] Rankinen T, et al. The human gene map for performance and health-related fitness phenotypes. *Med Sci Sports Exerc.* 2001;33(6):855-67. doi:10.1097/00005768-200106000-00001
- [597] Rankinen T, et al. The human gene map for performance and health-related fitness phenotypes: the 2001 update. *Med Sci Sports Exerc.* 2002;34(8):1219-1233. doi:10.1097/00005768-200208000-00001
- [598] Rankinen T, et al. The human gene map for performance and health-related fitness phenotypes: the 2003 update. *Med Sci Sports Exerc.* 2004;36(9):1451-69. doi:10.1249/01.mss.0000139902.42385.5f
- [599] Rantilä A, et al. High Responders to Hypertrophic Strength Training Also Tend to Lose More Muscle Mass and Strength During

- Detraining Than Low Responders. *Journal of strength and conditioning research*. 2021;35(6):1500-1511. doi:10.1519/jsc.0000000000004044
- [600] Rasbash J. Statistical Software Review: Mixed models in SPSS. *British Journal of Mathematical and Statistical Psychology*. 2004;57(1):189-190.
- [601] Rastogi T, et al. Advanced analytical methods to assess physical activity behaviour using accelerometer raw time series data: a protocol for a scoping review. *Systematic Reviews*. 2020;9(1):259. doi:10.1186/s13643-020-01515-2
- [602] Rebar AL, et al. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. *Health Psychology Review*. 2015;9(3):366-78. doi:10.1080/17437199.2015.1022901
- [603] Reeh HE, Reilly KJ. *A quasi meta analysis of the health locus of control construct*. 1995.
- [604] Rees H, et al. Epidemiology of injuries in senior men's field hockey: A two-season prospective observational injury surveillance study. *J Sports Sci*. 2020;38(24):2842-2849. doi:10.1080/02640414.2020.1802094
- [605] Reichardt CS. The Counterfactual Definition of a Program Effect. *American Journal of Evaluation*. 2022:1098214020975485. doi:10.1177/1098214020975485
- [606] Renner KE, et al. The Reliability and Validity of the Loadsol[®] under Various Walking and Running Conditions. *Sensors*. 2019;19(2):265.
- [607] Rettedal EA, et al. Short-term high-intensity interval training exercise does not affect gut bacterial community diversity or composition of lean and overweight men. *Exp Physiol*. 2020;105(8):1268-1279. doi:10.1113/ep088744
- [608] Rey Lopez JP, et al. Do vigorous-intensity and moderate-intensity physical activities reduce mortality to the same extent? A systematic review and meta-analysis. *BMJ Open Sport & Exercise Medicine*. 2020;6(1):e000775. doi:10.1136/bmjsem-2020-000775
- [609] Rice DB, et al. Dealing with predatory journal articles captured in systematic reviews. *Systematic Reviews*. 2021;10(1):175. doi:10.1186/s13643-021-01733-2
- [610] Ripper A, et al. Aspartame. In: O'Brien L, Gelardi RC, eds. *Alternative sweeteners*. Marcel Dekker, Inc.; 1986:43-70:chap 3.
- [611] Robertson PS. Man & machine: Adaptive tools for the contemporary performance analyst. *J Sports Sci*. 2020;38(18):2118-2126. doi:10.1080/02640414.2020.1774143
- [612] Roe KM. The Gift of Peer Review: A Note to My Younger Self. *Health Promotion Practice*. 2021;22(2):149-151. doi:10.1177/1524839920987848
- [613] *WebPlotDigitizer, version 4.5*. 2021. <https://automeris.io/WebPlotDigitizer>
- [614] Rommers N, et al. A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players. *Med Sci Sports Exerc*. 2020;52(8):1745-1751. doi:10.1249/MSS.0000000000002305
- [615] Roossien CC, et al. Can breathing gases be analyzed without a mouth mask? Proof-of-concept and concurrent validity of a newly developed design with a mask-less headset. *Applied Ergonomics*. 2021;90:103266. doi:<https://doi.org/10.1016/j.apergo.2020.103266>
- [616] Rosenberger ME, et al. The 24-Hour Activity Cycle: A New Paradigm for Physical Activity. *Med Sci Sports Exerc*. 2019;51(3):454-464. doi:10.1249/mss.0000000000001811
- [617] Rosenthal R. Designing, analyzing, interpreting, and summarizing placebo studies. In: White L, et al., eds. *Placebo: Theory, research and mechanisms*. The Guilford Press; 1985:110-126:chap 7.
- [618] Ross R, et al. Precision exercise medicine: understanding exercise response variability. *British journal of sports medicine*. 2019;53(18):1141. doi:10.1136/bjsports-2018-100328
- [619] Roth SM, et al. Advances in exercise, fitness, and performance genomics in 2011. *Med Sci Sports Exerc*. 2012;44(5):809-17. doi:10.1249/MSS.0b013e31824f28b6
- [620] Rowlands A, et al. Wrist-worn accelerometers: recommending ~1.0 mg as the minimum clinically important difference (MCID) in daily average acceleration for inactive adults. *British journal of sports medicine*. 2021;55(14):814-815. doi:10.1136/bjsports-2020-102293
- [621] Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581-592. doi:10.1093/biomet/63.3.581
- [622] Ruiz-Iglesias P, et al. Protective Effect of a Cocoa-Enriched Diet on Oxidative Stress Induced by Intensive Acute Exercise in Rats. *Antioxidants (Basel)*. 2022;11(4)doi:10.3390/antiox11040753
- [623] Russell AAM, et al. What has preclinical systematic review ever done for us? *BMJ Open Science*. 2022;6(1):e100219. doi:10.1136/bmjos-2021-100219
- [624] Ryom K, et al. Tackling physical inactivity in Scandinavia: a narrative review of reviews supplemented by expert interviews. *Scandinavian Journal of Public Health*. 2021:14034948211042956. doi:10.1177/14034948211042956
- [625] Saco-Ledo G, et al. Physical exercise and epicardial adipose tissue: A systematic review and meta-analysis of randomized controlled trials. *Obes Rev*. 2021;22(1):e13103. doi:<https://doi.org/10.1111/obr.13103>
- [626] Saeidi A, et al. The effects of physical activity on adipokines in individuals with overweight/obesity across the lifespan: A narrative review. *Obes Rev*. 2021;22(1):e13090. doi:10.1111/obr.13090
- [627] Sáez de Asteasu ML, et al. Assessing the impact of physical exercise on cognitive function in older medical patients during acute hospitalization: Secondary analysis of a randomized trial. *PLoS Med*. 2019;16(7):e1002852. doi:10.1371/journal.pmed.1002852
- [628] Sainani KL, et al. Call to increase statistical collaboration in sports science, sport and exercise medicine and sports physiotherapy. *British journal of sports medicine*. 2021;55(2):118-122. doi:10.1136/bjsports-2020-102607
- [629] Sakr S, et al. Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford Exercise Testing (FIT) Project. *PLoS One*. 2018;13(4):e0195344. doi:10.1371/journal.pone.0195344
- [630] Samnani SS, et al. Review Typology: The Basic Types of Reviews for Synthesizing Evidence for the Purpose of Knowledge Translation. *J Coll Physicians Surg Pak*. 2017;27(10):635-641.
- [631] Santabàrbara J. Cálculo del tamaño de muestra necesario para estimar el coeficiente de correlación de Pearson mediante sintaxis en SPSS. *REIRE Revista d'Innovació i Recerca en Educació*. 2021;14(1):1-7.
- [632] Sarzynski MA, et al. Genomic and transcriptomic predictors of response levels to endurance exercise training. *J Physiol*. 2017;595(9):2931-2939. doi:10.1113/jp272559
- [633] Sarzynski MA, et al. Advances in Exercise, Fitness, and Performance Genomics in 2015. *Med Sci Sports Exerc*. 2016;48(10):1906-16. doi:10.1249/mss.0000000000000982
- [634] Saueressig T, et al. Primary surgery versus primary rehabilitation for treating anterior cruciate ligament injuries: a living systematic review and meta-analysis. *British journal of sports medicine*. 2022;bjsports-2021-105359. doi:10.1136/bjsports-2021-105359
- [635] Schaefer J, et al. Competition Anxiety, Motivation, and Mental Toughness in Golf. *Journal of Applied Sport Psychology*. 2016;28(3):309-320. doi:10.1080/10413200.2016.1162219
- [636] Schild C, et al. Linking personality traits to objective foul records in (semi-)professional youth basketball. *Journal of Research in Personality*. 2020;87:103987. doi:<https://doi.org/10.1016/j.jrp.2020.103987>
- [637] Schinke RJ, et al. Getting Published: Suggestions and Strategies from Editors of Sport and Exercise Psychology Journals. *Journal of Applied Sport Psychology*. 2020;doi:10.1080/10413200.2020.1741725
- [638] Schmidt W, Prommer N. Impact of alterations in total hemoglobin mass on VO₂max. *Exerc Sport Sci Rev*. 2010;38(2):68-75. doi:10.1097/JES.0b013e3181d4957a
- [639] Schultz HD. Visualizing data in research articles. *J Physiol*. 2018;596(16):3431-3432. doi:10.1113/jp276501
- [640] Schumacher YO, et al. Resuming professional football (soccer) during the COVID-19 pandemic in a country with high infection rates: a prospective cohort study. *British journal of sports medicine*. 2021;55(19):1092-1098. doi:10.1136/bjsports-2020-103724
- [641] Seo YG, et al. What Is the Optimal Exercise Prescription for Patients With Dilated Cardiomyopathy in Cardiac Rehabilitation? A systematic review. *J Cardiopulm Rehabil Prev*. 2019;39(4):235-240. doi:10.1097/hcr.0000000000000382
- [642] Shakespear-Druery J, et al. Muscle-Strengthening Exercise Questionnaire (MSEQ): an assessment of concurrent validity and test-retest reliability. *BMJ Open Sport & Exercise Medicine*. 2022;8(1):e001225. doi:10.1136/bmjsem-2021-001225

- [643] Shamseer L, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ (Clinical research ed)*. 2015;350:g7647. doi:10.1136/bmj.g7647
- [644] Sharp P, et al. Infographic. One small step for man, one giant leap for men's health: a meta-analysis of behaviour change interventions to increase men's physical activity. *British journal of sports medicine*. 2021;55(14):816-817. doi:10.1136/bjsports-2020-102976
- [645] Shea BJ, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ (Clinical research ed)*. 2017;358:j4008. doi:10.1136/bmj.j4008
- [646] Shi L, Lin L. The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses. *Medicine*. 2019;98(23):e15987. doi:10.1097/md.00000000000015987
- [647] Shirai T, et al. Effect of endurance exercise duration on muscle hypertrophy induced by functional overload. *FEBS Open Bio*. 2021;11(1):85-94. doi:10.1002/2211-5463.13028
- [648] Shojania KG, et al. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med*. 2007;147(4):224-33. doi:10.7326/0003-4819-147-4-200708210-00179
- [649] Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Methods Med Res*. 1998;7(3):301-317. doi:10.1177/096228029800700306
- [650] Sillanpää E, et al. Polygenic Score for Physical Activity Is Associated with Multiple Common Diseases. *Med Sci Sports Exerc*. 2022;54(2):280-287. doi:10.1249/mss.0000000000002788
- [651] Sirichana W, et al. Reference Values for Chronotropic Index from 1280 Incremental Cycle Ergometry Tests. *Med Sci Sports Exerc*. 2020;52(12):2515-2521. doi:10.1249/mss.0000000000002417
- [652] Sirichana W, et al. Response. *Med Sci Sports Exerc*. 2021;53(6):1316. doi:10.1249/mss.0000000000002651
- [653] Siwek J, Barton S. Evidence at the point of care. *Am Fam Physician*. 2002;65(1):27-8.
- [654] Smart NA, et al. Validation of a new tool for the assessment of study quality and reporting in exercise training studies: TESTEX. *Int J Evid Based Healthc*. 2015;13(1):9-18. doi:10.1097/xeb.0000000000000020
- [655] Smith CJ, Havenith G. Upper body sweat mapping provides evidence of relative sweat redistribution towards the periphery following hot-dry heat acclimation. *Temperature (Austin)*. 2019;6(1):50-65. doi:10.1080/23328940.2019.1570777
- [656] Smith CT, et al. Individual participant data meta-analyses compared with meta-analyses based on aggregate data. *The Cochrane database of systematic reviews*. 2016;9:MR000007.
- [657] Smith E, et al. Physical behaviors and fundamental movement skills in British and Iranian children: An isotemporal substitution analysis. *Scand J Med Sci Sports*. 2021;31(2):398-404. doi:https://doi.org/10.1111/sms.13837
- [658] Sobreiro P, et al. Swimmer Dropout Rate: A Survival Analysis. *Apunts Educación Física y Deportes*. 2022;147:74-83. doi:https://doi.org/10.5672/apunts.2014-0983.es.(2022/1).147.08
- [659] Sollanek KJ, et al. The accurate prediction of sweat rate from energy expenditure and air temperature: a proof-of-concept study. *Applied Physiology, Nutrition & Metabolism*. 2020;45(11):1299-1305. doi:10.1139/apnm-2020-0236
- [660] Solomon TPJ. Sources of Inter-individual Variability in the Therapeutic Response of Blood Glucose Control to Exercise in Type 2 Diabetes: Going Beyond Exercise Dose. *Frontiers in Physiology*. 2018;9(896)doi:10.3389/fphys.2018.00896
- [661] Soto-Fernández A, et al. LINCE PLUS: Research software for behavior video analysis. *Apunts Educación Física y Deportes*. 2019;3(137):149-153.
- [662] Soto-Fernández A, et al. LINCE PLUS software for systematic observational studies in sports and health. *Behavior research methods*. 2022;54(3):1263-1271. doi:10.3758/s13428-021-01642-1
- [663] Springham M, et al. Prior workload has moderate effects on high-intensity match performance in elite-level professional football players when controlling for situational and contextual variables. *J Sports Sci*. 2020;38(20):2279-2290. doi:10.1080/02640414.2020.1778355
- [664] Spurk D, et al. Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*. 2020;120:103445. doi:https://doi.org/10.1016/j.jvb.2020.103445
- [665] Spurlock D, Jr. Scholarship During a Pandemic: Secondary Data Analysis. *J Nurs Educ*. 2020;59(5):245-247. doi:10.3928/01484834-20200422-02
- [666] Stamatakis E, et al. Sitting Time, Physical Activity, and Risk of Mortality in Adults. *Journal of the American College of Cardiology*. 2019;73(16):2062-2072. doi:10.1016/j.jacc.2019.02.031
- [667] Stamatis A, et al. Offseason Fitness Tests a Collegiate Basketball Strength Coach Should Choose to Predict In-Season Performance Based on Sex. *International Journal of Computer Science in Sport*. 2021;20(2):164-174. doi:doi:10.2478/ijcss-2021-0010
- [668] Stangroom J. Social Science Statistics. 2020. https://www.socscistatistics.com/
- [669] Steel EA, et al. Beyond Calculations: A Course in Statistical Thinking. *The American Statistician*. 2019;73(sup1):392-401. doi:10.1080/00031305.2018.1505657
- [670] Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol*. 2001;54(10):1046-55. doi:10.1016/s0895-4356(01)00377-8
- [671] Sterne JAC, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ (Clinical research ed)*. 2019;366:14898. doi:10.1136/bmj.14898
- [672] Stewart-Williams S, Podd J. The placebo effect: dissolving the expectancy versus conditioning debate. *Psychol Bull*. 2004;130(2):324-40. doi:10.1037/0033-2909.130.2.324
- [673] Stout JL, et al. Rasch analysis of items from two self-report measures of motor function: determination of item difficulty and relationships with children's ability levels. *Developmental Medicine & Child Neurology*. 2012;54(5):443-50. doi:10.1111/j.1469-8749.2012.04231.x
- [674] Strain T, et al. Wearable-device-measured physical activity and future health risk. *Nature Medicine*. 2020;doi:10.1038/s41591-020-1012-3
- [675] Stroup DF, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *Jama*. 2000;283(15):2008-12. doi:10.1001/jama.283.15.2008
- [676] Su X, et al. Application of network meta-analysis in the field of physical activity and health promotion. *Journal of Sport and Health Science*. 2020;9(6):511-520. doi:https://doi.org/10.1016/j.jshs.2020.07.011
- [677] Sullivan GM, Feinn R. Using Effect Size-or Why the P Value Is Not Enough. *Journal of graduate medical education*. 2012;4(3):279-82. doi:10.4300/jgme-d-12-00156.1
- [678] Swinton PA, et al. A Statistical Framework to Interpret Individual Response to Intervention: Paving the Way for Personalized Nutrition and Exercise Prescription. *Frontiers in nutrition*. 2018;5:41. doi:10.3389/fnut.2018.00041
- [679] Sylow L. Three challenges of being a scientist in an age of misinformation. *J Physiol*. 2021;599(7):1937-1938. doi:https://doi.org/10.1113/jp281434
- [680] Syväoja HJ, et al. How physical activity, fitness, and motor skills contribute to math performance: Working memory as a mediating factor. *Scand J Med Sci Sports*. 2021;31(12):2310-2321. doi:https://doi.org/10.1111/sms.14049
- [681] Tabachnick BG, et al. *Using multivariate statistics*. vol 5. Pearson Boston, MA; 2007.
- [682] Taddei UT, et al. Effects of a foot strengthening program on foot muscle morphology and running mechanics: A proof-of-concept, single-blind randomized controlled trial. *Physical Therapy in Sport*. 2020;42:107-115. doi:https://doi.org/10.1016/j.ptsp.2020.01.007
- [683] Taniguchi H, et al. Effects of short-term endurance exercise on gut microbiota in elderly men. *Physiological reports*. 2018;6(23):e13935. doi:10.14814/phy2.13935
- [684] Taniguchi Y, et al. Evidence that dog ownership protects against the onset of disability in an older community-dwelling Japanese population. *PLoS One*. 2022;17(2):e0263791. doi:10.1371/journal.pone.0263791
- [685] Tarp J, et al. Double counting individuals in meta-analysis artificially inflates precision. Comment on "Device-measured light-intensity

- physical activity and mortality: A meta-analysis". *Scand J Med Sci Sports*. 2020;30(6):1083-1084. doi:10.1111/sms.13693
- [686] Tate RL, et al. Reprint of "the single-case reporting guideline in behavioural interventions (SCRIBE) 2016: Explanation and elaboration". *Pratiques Psychologiques*. 2019;25(2):119-151. doi:10.1016/j.prps.2019.03.001
- [687] Taylor CE, et al. Blood pressure status and post-exercise hypotension: an example of a spurious correlation in hypertension research? *Journal of Human Hypertension*. 2010;24(9):585-592. doi:10.1038/jhh.2009.112
- [688] Taylor JK, et al. A systematic review and meta-analysis of the acute and chronic effects of exercise training on paraoxonase-1 (PON1) activity. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018092693
- [689] Taylor JK, et al. The Effects of Acute and Chronic Exercise on Paraoxonase-1 (PON1): A Systematic Review with Meta-Analysis. *Research Quarterly for Exercise and Sports*. 2022;93(1):130-143. doi:<https://doi.org/10.1080/02701367.2020.1812493>
- [690] Teixeira DS, et al. Motivational patterns in persistent swimmers: A serial mediation analysis. *Eur J Sport Sci*. 2020;20(5):660-669. doi:10.1080/17461391.2019.1675768
- [691] Tenenbaum G, et al. Perceived discomfort in running: Scale development and theoretical considerations. *J Sports Sci*. 1999;17(3):183-196.
- [692] Thomas HJ, et al. Studies of Twin Responses to Understand Exercise Therapy (STRUETH): Body Composition. *Med Sci Sports Exerc*. 2021;53(1):58-67. doi:10.1249/MSS.0000000000002461
- [693] Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*. 2008;8:45. doi:10.1186/1471-2288-8-45
- [694] Thomas JR, et al. *Research Methods in Physical Activity*. 7 ed. Human Kinetics; 2015.
- [695] Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*. 2002;21(11):1559-1573. doi:<https://doi.org/10.1002/sim.1187>
- [696] Thuan M, et al. Is there any difference between "amateur" and "recreational" runners? A latent class analysis. *Motriz: Revista de Educação Física*. 2020;26(4):e10200140. doi:<http://dx.doi.org/10.1590/S1980-65742020000400140>
- [697] Tian S. The effects of physical exercise on epicardial adipose tissue: Some methodological issues could be considered. *Obes Rev*. 2021:e13233. doi:<https://doi.org/10.1111/obr.13233>
- [698] Tipton E, et al. A history of meta-regression: Technical, conceptual, and practical developments between 1974 and 2018. *Research Synthesis Methods*. 2019;10(2):161-179. doi:<https://doi.org/10.1002/rsm.1338>
- [699] Tobias A, et al. Impacts of ambient temperature and seasonal changes on sports injuries in Madrid, Spain: a time-series regression analysis. *BMJ Open Sport & Exercise Medicine*. 2021;7(4):e001205. doi:10.1136/bmjsem-2021-001205
- [700] Tong C. Statistical Inference Enables Bad Science; Statistical Thinking Enables Good Science. *The American Statistician*. 2019;73(sup1):246-261. doi:10.1080/00031305.2018.1518264
- [701] Tousignant M, et al. Tai Chi-based exercise program provided via telerehabilitation compared to home visits in a post-stroke population who have returned home without intensive rehabilitation: study protocol for a randomized, non-inferiority clinical trial. *Trials*. 2014;15:42. doi:10.1186/1745-6215-15-42
- [702] Tricco AC, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467-473. doi:10.7326/m18-0850
- [703] Trussell J. Pills or Placebos? *Science*. 1999;284(5416):913-913. doi:10.1126/science.284.5416.913c
- [704] Tsiros MD, et al. Helping Children with Obesity "Move Well" To Move More: An Applied Clinical Review. *Curr Sports Med Rep*. 2021;20(7):374-383. doi:10.1249/jsr.0000000000000861
- [705] Tully MA, et al. Sedentary behavior, physical activity, and mental health in older adults: An isomorphous substitution model. *Scand J Med Sci Sports*. 2020;doi:10.1111/sms.13762
- [706] Turner AN, et al. Assessing group-based changes in high-performance sport. Part 2: effect sizes and embracing uncertainty through confidence intervals. *Strength & Conditioning Journal*. 2021;43(4):68-77. doi:10.1519/SSC.0000000000000613
- [707] Turner AN, et al. Assessing group-based changes in high-performance sport. Part 1: null hypothesis significance testing and the utility of p values. *Strength & Conditioning Journal*. 2021;43(3):112-116. doi:10.1519/SSC.0000000000000625
- [708] Turner MJ, et al. "I must do this!": A latent profile analysis approach to understanding the role of irrational beliefs and motivation regulation in mental and physical health. *J Sports Sci*. 2022;40(8):934-949. doi:10.1080/02640414.2022.2042124
- [709] Tyers M, Rappilberg Laboratory. BoxPlotR: a web-tool for generation of box plots. Université de Montréal. 2020. <http://shiny.chemgrid.org/boxplotr/>
- [710] van den Akker OR, et al. Preregistration of secondary data analysis: A template and tutorial. *Meta-psychology*. 2021;5doi:<https://doi.org/10.15626/MP.2020.2625>
- [711] van der Fels IMJ, et al. Effects of aerobic exercise and cognitively engaging exercise on cardiorespiratory fitness and motor skills in primary school children: A cluster randomized controlled trial. *J Sports Sci*. 2020;38(17):1975-1983. doi:10.1080/02640414.2020.1765464
- [712] van Eck NJ, Waltman L. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of Informetrics*. 2014;8(4):802-823. doi:<https://doi.org/10.1016/j.joi.2014.07.006>
- [713] VOSviewer. Centre for Science and Technology Studies (CWTS); 2020. <https://www.vosviewer.com>
- [714] van Houwelingen HC, et al. Likelihood Modelling: Advanced methods in Meta-Analysis: Multivariate Approach and Meta-Regression. *Tutorials in Biostatistics*. 2004:289-324.
- [715] van Kleef LA, et al. Objectively Measured Physical Activity Is Inversely Associated With Nonalcoholic Fatty Liver Disease: The Rotterdam Study. *Official journal of the American College of Gastroenterology | ACG*. 2022;117(2):311-318. doi:10.14309/ajg.0000000000001584
- [716] van Middelkoop M, et al. Sharing data-taming the beast: barriers to meta-analyses of individual patient data (IPD) and solutions. *British journal of sports medicine*. 2020;54(14):822-824. doi:10.1136/bjsports-2019-101892
- [717] Vandenberghe JP, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg*. 2014;12(12):1500-24. doi:10.1016/j.ijssu.2014.07.014
- [718] Vanderburgh PM, et al. Allometric scaling of grip strength by body mass in college-age men and women. *Research Quarterly for Exercise and Sports*. 1995;66(1):80-84.
- [719] Vasconcelos NN, et al. High Intensity Physical Exercise before the Brain Ischemia Promotes Increase in Brain Injury. *J Stroke Cerebrovasc Dis*. 2021;30(12):106148. doi:10.1016/j.jstrokecerebrovasdis.2021.106148
- [720] Verhagen E, et al. Tennis-specific extension of the International Olympic Committee consensus statement: methods for recording and reporting of epidemiological data on injury and illness in sport 2020. *British journal of sports medicine*. 2021;55(1):9-13. doi:10.1136/bjsports-2020-102360
- [721] Video Journal of Sports Medicine. Sage Publishing. Accessed Jan 25th, 2021. <https://www.sportsmed.org/journals/vjssm>
- [722] Viljoen C, et al. Trail running injury risk factors: a living systematic review. *British journal of sports medicine*. 2022;bjsports-2021-104858. doi:10.1136/bjsports-2021-104858
- [723] Villaça CBP, et al. Beneficial effects of physical exercise for β -cell maintenance in a type 1 diabetes mellitus animal model. *Exp Physiol*. 2021;106(7):1482-1497. doi:10.1113/ep088872
- [724] Villaseñor ÁB, et al. Data analysis techniques in observational designs applied to the environment-behaviour relation1. *Medio Ambiente y Comportamiento Humano*. 2003;4(2):111-126.
- [725] Villaseñor AB, et al. Diseños observacionales, cuestión clave en el proceso de la metodología observacional. *Metodología de las Ciencias del Comportamiento*. 2001;3(2):135-160.
- [726] Voelker DK, et al. Toward understanding of coaches' role in athletes' eating pathology: A systematic review and ecological application to advance research. *Psychology of Sport & Exercise*. 2022;58doi:10.1016/j.psychsport.2021.102059
- [727] Vohra S, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *BMJ (Clinical research ed)*. 2015;350:h1738. doi:10.1136/bmj.h1738

- [728] Voisin S, et al. Statistical Considerations for Exercise Protocols Aimed at Measuring Trainability. *Exerc Sport Sci Rev.* 2019;47(1):37-45. doi:10.1249/JES.0000000000000176
- [729] von Elm E, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol.* 2008;61(4):344-9. doi:10.1016/j.jclinepi.2007.11.008
- [730] von Rosen P, et al. Latent profile analysis of physical activity and sedentary behavior with mortality risk: A 15-year follow-up. *Scand J Med Sci Sports.* 2020;doi:10.1111/sms.13761
- [731] Voorn EL, et al. Individualized Aerobic Exercise in Neuromuscular Diseases: A Pilot Study on the Feasibility and Preliminary Effectiveness to Improve Physical Fitness. *Phys Ther.* 2021;101(3)doi:10.1093/ptj/pzaa213
- [732] Wallace BC, et al. OpenMEE: Intuitive, open-source software for meta-analysis in ecology and evolutionary biology. *Methods in Ecology and Evolution.* 2017;8:941-947. doi:10.1111/2041-210x.12708
- [733] Walsh JJ, et al. Interindividual variability and individual responses to exercise training in adolescents with obesity. *Applied Physiology, Nutrition & Metabolism.* 2020;45(1):45-54. doi:10.1139/apnm-2019-0088
- [734] Wan L, Liu J. Cyclic Anaerobic Exercise Performance and Neuromuscular Activity Based on Artificial Intelligence Genetic Algorithm. *Mobile Information Systems.* 2022;2022:3960120. doi:10.1155/2022/3960120
- [735] Wang M, Hanges PJ. Latent Class Procedures: Applications to Organizational Research. *Organizational Research Methods.* 2011;14(1):24-31. doi:10.1177/1094428110383988
- [736] Wang R, et al. Exercise for low back pain: A bibliometric analysis of global research from 1980 to 2018. *Journal of rehabilitation medicine.* 2020;52(4):jrm00052. doi:10.2340/16501977-2674
- [737] Wang YJ. Saccharides: Modifications and applications. In: Tomasik P, ed. *Chemical and functional properties of food saccharides.* CRC Press; 2004:35-46.
- [738] Warkentin LM, et al. Weight loss required by the severely obese to achieve clinically important differences in health-related quality of life: two-year prospective cohort study. *BMC Medicine.* 2014;12(1):175. doi:10.1186/s12916-014-0175-5
- [739] Washabaugh EP, et al. Functional Resistance Training Differentially Alters Gait Kinetics After Anterior Cruciate Ligament Reconstruction: A Pilot Study. *Sports Health.* 2022:19417381221104042. doi:10.1177/19417381221104042
- [740] Wasserstein RL, Lazar NA. The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician.* 2016;70(2):129-133. doi:10.1080/00031305.2016.1154108
- [741] Wasserstein RL, et al. Moving to a World Beyond "p<0.05". *The American Statistician.* 2019;73(sup1):1-19. doi:10.1080/00031305.2019.1583913
- [742] Watts A, et al. Intra-Individual Variability of Physical Activity in Older Adults With and Without Mild Alzheimer's Disease. *PLoS One.* 2016;11(4):e0153898. doi:10.1371/journal.pone.0153898
- [743] Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of strength and conditioning research.* 2005;19(1):231-40. doi:10.1519/15184.1
- [744] Weissgerber TL, et al. Reveal, Don't Conceal. *Circulation.* 2019;140(18):1506-1518. doi:10.1161/CIRCULATIONAHA.118.037777
- [745] Weissgerber TL, et al. Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol.* 2015;13(4):e1002128. doi:10.1371/journal.pbio.1002128
- [746] Weissgerber TL, et al. Data visualization, bar naked: A free tool for creating interactive graphics. *J Biol Chem.* 2017;292(50):20592-20598. doi:10.1074/jbc.RA117.000147
- [747] Werkhausen A, et al. Adaptations to explosive resistance training with partial range of motion are not inferior to full range of motion. *Scand J Med Sci Sports.* 2021;31(5):1026-1035. doi:10.1111/sms.13921
- [748] *Free Statistics Software.* Version 1.2.1. Office for Research Development and Education; 2020. <https://www.wessa.net/>
- [749] Weston M, et al. Effects of Low-Volume High-Intensity Interval Training (HIT) on Fitness in Adults: A Meta-Analysis of Controlled and Non-Controlled Trials. *Sports Med.* 2014;44(7):1005-1017. doi:10.1007/s40279-014-0180-z
- [750] Wilburn DT, et al. Comparison of skeletal muscle ultrastructural changes between normal and blood flow-restricted resistance exercise: A case report. *Exp Physiol.* 2021;106(11):2177-2184. doi:10.1113/ep089858
- [751] Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. *American journal of epidemiology.* 1986;124(1):17-27. doi:10.1093/oxfordjournals.aje.a114366
- [752] Williams CA, James DV. *Science for exercise and sport.* Psychology Press; 2001.
- [753] Williams CJ, et al. A Multi-Center Comparison of VO2peak Trainability Between Interval Training and Moderate Intensity Continuous Training. *Front Physiol.* 2019;10:19. doi:10.3389/fphys.2019.00019
- [754] Windt J, et al. Making everyone's job easier. How do data scientists fit as a critical member of integrated support teams? *British journal of sports medicine.* 2021;55(2):73-75. doi:10.1136/bjsports-2020-102938
- [755] Winters M, et al. Stay alive! What are living systematic reviews and what are their advantages and challenges? *British journal of sports medicine.* 2020:bjsports-2020-103490. doi:10.1136/bjsports-2020-103490
- [756] Winters M, et al. Comparative effectiveness of treatments for patellofemoral pain: a living systematic review with network meta-analysis. *British journal of sports medicine.* 2020:bjsports-2020-102819. doi:10.1136/bjsports-2020-102819
- [757] Wolf V, et al. Does transcutaneous auricular vagus nerve stimulation affect vagally mediated heart rate variability? A living and interactive Bayesian meta-analysis. *Psychophysiology.* 2021;58(11):e13933. doi:https://doi.org/10.1111/psyp.13933
- [758] Wolfarth B, et al. The human gene map for performance and health-related fitness phenotypes: the 2004 update. *Med Sci Sports Exerc.* 2005;37(6):881-903. doi:10.1249/01.mss.0000168663.55604.1d
- [759] Wolfarth B, et al. Advances in exercise, fitness, and performance genomics in 2013. *Med Sci Sports Exerc.* 2014;46(5):851-9. doi:10.1249/mss.0000000000000300
- [760] Wong G, et al. RAMESES publication standards: realist syntheses. *BMC Med.* 2013;11:21. doi:10.1186/1741-7015-11-21
- [761] Wright S, et al. Research Design Considerations. *Journal of graduate medical education.* 2016;8(1):97-8. doi:10.4300/JGME-D-15-00566.1
- [762] Wu PP-Y, et al. Bayesian prediction of winning times for elite swimming events. *J Sports Sci.* 2022;40(1):24-31. doi:10.1080/02640414.2021.1976485
- [763] Wu QL, Brannon GE. Factors related to moderate exercise during COVID-19 for overweight and obese individuals: A secondary analysis of HINTS data. *PEC Innov.* 2022;1:100058. doi:10.1016/j.pecinn.2022.100058
- [764] Wu Z, Qi S. Computer-assisted children physical fitness detection and exercise intervention evaluation based on artificial intelligence model. *IEEE;* 2020:13-16.
- [765] Xing A, Lin L. Effects of treatment classifications in network meta-analysis. *Research Methods in Medicine & Health Sciences.* 2020;1(1):12-24. doi:10.1177/2632084320932756
- [766] Yang S, Yuan Q. Are scientometrics, informetrics, and bibliometrics different. *Data Science and Informetrics.* 2020;1(1):103597. doi:https://doi.org/10.4236/dsi.2020.11003
- [767] Yarborough M. Moving towards less biased research. *BMJ Open Science.* 2021;5(1):e100116. doi:10.1136/bmjopen-2020-100116
- [768] Yaremkov RM, et al. *Reference Handbook of Research and Statistical Methods in Psychology: For Students and Professionals.* Harper & Row; 1982.
- [769] Zavorsky GS. Incorrect Terminology Confuses Article's Purpose and Usefulness. *Med Sci Sports Exerc.* 2021;53(6):1315. doi:10.1249/mss.00000000000002650
- [770] Završnik J, et al. The impact of physical activity to the child's quality of life: a bibliometric study. *F1000Research.* 2019;8:672. doi:10.12688/f1000research.18838.1
- [771] Zeiger JS, Zeiger RS. Mental toughness latent profiles in endurance athletes. *PLoS One.* 2018;13(2):e0193071. doi:10.1371/journal.pone.0193071
- [772] Zeng L, et al. When applying GRADE, how do we decide the target of certainty of evidence rating? *Evid Based Ment Health.* 2021;24:121-123. doi:10.1136/ebmental-2020-300170

- [773] Zhang Y, et al. Prediction of exercise sudden death in rabbit exhaustive swimming using deep neural network. *Biomed Eng Online*. 2021;20(1):87. doi:10.1186/s12938-021-00925-0
- [774] Zhao Y, et al. Empirical comparisons of heterogeneity magnitudes of the risk difference, relative risk, and odds ratio. *Systematic Reviews*. 2022;11(1):26. doi:10.1186/s13643-022-01895-7
- [775] Zhu W. Should total scores from a rating scale be used directly? *Research Quarterly for Exercise and Sport*. 1996;67(3):363-72. doi:https://doi.org/10.1080/02701367.1996.10607966
- [776] Zhu W. Sadly, the earth is still round ($p < 0.05$). *J Sport Health Sci*. 2012;1(1):9-11. doi:https://doi.org/10.1016/j.jshs.2012.02.002
- [777] Zhu W. Reliability: What type, please! *J Sport Health Sci*. 2013;2(1):62-64. doi:https://doi.org/10.1016/j.jshs.2012.11.001
- [778] Zhu W, et al. Many-Faceted Rasch Modeling Expert Judgment in Test Development. *Measurement in Physical Education and Exercise Science*. 1998;2(1):21-39. doi:10.1207/s15327841mpee0201_2
- [779] Zhu W, Kurz KA. Rasch Partial Credit Analysis of Gross Motor Competence. *Perceptual and motor skills*. 1994;79(2):947-961. doi:10.2466/pms.1994.79.2.947
- [780] Zignoli A, et al. Oxynet: A collective intelligence that detects ventilatory thresholds in cardiopulmonary exercise tests. *Eur J Sport Sci*. 2022;22(3):425-435. doi:10.1080/17461391.2020.1866081
- [781] Zignoli A, et al. Expert-level classification of ventilatory thresholds from cardiopulmonary exercising test data with recurrent neural networks. *Eur J Sport Sci*. 2019;19(9):1221-1229. doi:10.1080/17461391.2019.1587523
- [782] Zimmerman J, et al. Iterative guided machine learning-assisted systematic literature reviews: a diabetes case study. *Systematic Reviews*. 2021;10(1):97. doi:10.1186/s13643-021-01640-6

Acerca de los autores

José Moncada Jiménez, Ph.D.



Doctor en Ciencias Biomédicas de la Universidad de Costa Rica. Investigación post doctoral en Baylor University y estudios de posgrado en The Ohio State University. Es Máster en Ciencias del Movimiento con especialización en Fisiología del Ejercicio de Springfield College (EE.UU.). Es Bachiller en Educación Física de la Universidad de Costa Rica. Fundador y primer director del Centro de Investigación en Ciencias del Movimiento Humano de la Universidad de Costa Rica. Asesor de posgrado en Costa Rica, España, México, Brasil y Portugal. Profesor de grado y posgrado en la Escuela de Educación Física y Deportes y en la Facultad de Medicina de la Universidad de Costa Rica. Miembro del American College of Sports Medicine (ACSM) y de la Red Americana de Investigación en Ciencias del Movimiento Humano (RAICIMH).

Pedro Carazo Vargas, Ph.D.



Doctor en Ciencias del Movimiento Humano de la Universidad Nacional, Heredia y la Universidad de Costa Rica. Especialista en Alto Rendimiento Deportivo de la Universidad de Castilla La Mancha, Toledo, España. Máster en Ciencias del Movimiento Humano y Bachiller en la Enseñanza de la Educación Física de la Universidad de Costa Rica. Investigador del Centro de Investigación en Ciencias del Movimiento Humano de la Universidad de Costa Rica. Miembro de la Red Americana de Investigación en Ciencias del Movimiento Humano (RAICIMH).

Jessenia Hernández Elizondo, Ph.D.



Doctora en Nutrición y Tecnología de los Alimentos y Máster en Nutrición Humana de la Universidad de Granada, España. Máster en Ciencias del Movimiento Humano y Bachiller en Educación Física de la Universidad de Costa Rica. Sus temas de interés son la investigación en ciencias del movimiento humano y prescripción del ejercicio físico en la enfermedad, ejercicio físico en el tratamiento de patologías, entrenamiento de la fuerza, potencia y velocidad, y valoración nutricional y de actividad física. Miembro de la Red Americana de Investigación en Ciencias del Movimiento Humano (RAICIMH).

Isaura Castillo Hernández, Ph.D.



Doctora en Kinesiología con Énfasis en Fisiología del Ejercicio de la Universidad de Georgia, EE.UU. Obtuvo una Certificación de Posgrado en Gerontología de la Universidad de Georgia, EE.UU. Máster en Ciencias del Movimiento Humano y Bachiller en la Enseñanza de la Educación Física de la Universidad de Costa Rica. Graduada de la primera generación del Programa Técnico en Gestión del Deporte de la Universidad de Costa Rica, la Federación Internacional de Fútbol Asociado (FIFA) y el Centro Internacional de Estudios del Deporte (CIES) de Suiza. Investigadora del Centro de Investigación en Ciencias del Movimiento Humano y profesora de cursos de grado y posgrado en la Escuela de Educación Física y Deportes, ambos de la Universidad de Costa Rica. Miembro de la Red Americana de Investigación en Ciencias del Movimiento Humano (RAICIMH).

Gerardo Araya Vargas, Ed.D.



Doctor en Educación de la Universidad La Salle, Costa Rica. Máster en Ciencias del Movimiento Humano de la Universidad de Costa Rica. Bachiller en la Enseñanza de la Educación Física con Concentración en Deportes de la Universidad Nacional, Costa Rica. Es entrenador nivel 1 y 2 de fútbol de la Dirección General de Educación Física y Deportes de Costa Rica. Es investigador en psicología del ejercicio y el deporte, actividad física y salud y rendimiento deportivo. Investigador del Centro de Investigación en Ciencias del Movimiento Humano de la Universidad de Costa Rica. Miembro de la Red Americana de Investigación en Ciencias del Movimiento Humano (RAICIMH).

EDUFI
Escuela de
Educación Física y
Deportes

CIMOHU Centro de
Investigación en Ciencias
del Movimiento Humano





SCULPTURE
BY [unreadable]
1977

ISBN: 978-9968-03-198-1



9 789968 031981