

Juan Osvaldo Talavera^{1a}, Ivonne Roy-García^{2b}, Sofía Teresa Díaz-Torres^{1c}, Lino Palacios-Cruz^{3d}, Alejandro Noguez-Ramos^{1e}, Marcela Pérez-Rodríguez^{2f}, Miguel Ángel Martínez^{1g}, Jessica E. Silva-Guzmán^{1h}, Rodolfo Rivas-Ruiz²ⁱ

Resumen

El manejo de datos “*tras bambalinas*” se refiere a los procesos de recopilación, limpieza, imputación y demarcación; los cuales, aun siendo indispensables, usualmente suelen ser descuidados, por lo que generan información errónea. Durante la recopilación son errores: omisión de covariables, desvío del objetivo, y calidad insuficiente. La omisión de covariables distorsiona el resultado atribuido a la maniobra principal. El desvío del objetivo primario es común cuando el desenlace es raro, tardado o subjetivo y promueve la sustitución por variables subrogadas no equivalentes. Además, la calidad insuficiente, sucede por instrumentos inadecuados, omisión del procedimiento de medición, o medición fuera de contexto -como atribución a destiempo o equivalente-. Por otro lado, la limpieza implica identificar valores erróneos, extremos y faltantes, que podrán ser o no imputados, dependiendo del porcentaje se imputará comúnmente por la medida de resumen. Nunca se imputan los valores de la maniobra ni del desenlace, ni se eliminan pacientes por falta de valores. Finalmente, la demarcación de cada variable busca un significado clínico en referencia al desenlace, para ello se sigue una secuencia jerárquica de criterios: 1) estudio clínico previo, 2) acuerdo de expertos, 3) juicio clínico del investigador/investigadores y 4) estadística. Actuar sin controles de calidad en el manejo de datos provoca frecuentemente mentiras involuntarias y confunde en lugar de esclarecer.

Abstract

Data management “*behind the scenes*” refers to collection, cleaning, imputation, and demarcation; and despite of being indispensable processes, they are usually neglected and thus, generate erroneous information. During the collection are errors: omission of covariates, deviation from the objective, and insufficient quality. The omission of covariates distorts the result attributed to the main manoeuvre. Deviation from the primary objective commonly occurs when the outcome is rare, delayed, or subjective and promotes substitution by non-equivalent surrogate variables. Moreover, insufficient quality occurs due to inadequate instruments, omission of the measurement procedure, or measurements out of context, such as attribution at the wrong time or equivalent. Furthermore, cleaning implies identifying erroneous, extreme, and missing values, which may or may not be imputed, depending on the percentage. The values of the manoeuvre or the outcome are never imputed, nor are patients eliminated due to a lack of values. Finally, the demarcation of each variable seeks to give it a clinical meaning about the outcome, for which a hierarchical sequence of criteria is followed: 1) previous clinical study, 2) expert agreement, 3) clinical judgment of the investigator/investigators, and 4) statistics. Acting without quality controls in data management frequently causes involuntary lies and confuses instead of clarifying.

¹Centro Médico ABC, Subdirección de Enseñanza e Investigación. Ciudad de México, México

²Instituto Mexicano del Seguro Social, Coordinación de Investigación en Salud, Centro de Adiestramiento en Investigación Clínica. Ciudad de México, México

³Instituto Nacional de Psiquiatría Dr. Ramón de la Fuente, Subdirección de Investigaciones Clínicas, Departamento de Epidemiología Clínica. Ciudad de México, México

ORCID: 0000-0003-2272-7369^a, 0000-0002-1859-3866^b, 0000-0003-2300-7662^c, 0000-0001-9533-2996^d, 0000-0002-4784-1793^e, 0000-0003-3417-3201^f, 0000-0001-6987-8354^g, 0000-0002-7814-6785^h, 0000-0002-5967-7222ⁱ

Palabras clave

Recolección de Datos
Manejo de Datos
Epidemiología Clínica
Estadística

Keywords

Data Collection
Data Management
Clinical Epidemiology
Statistics

Fecha de recibido: 13/01/2023

Fecha de aceptado: 15/02/2023

Comunicación con:

Juan Osvaldo Talavera

 jotalaverap@abchospital.com

.....
Cómo citar este artículo: Talavera JO, Roy-García I, Díaz-Torres ST et al. Expresión numérica del curso clínico de la enfermedad. Manejo de datos. Rev Med Inst Mex Seguro Soc. 2023;61 Supl 3:S503-9. doi: 10.5281/zenodo.8319834

Si la manifestación de la vida fuera toda en blanco o negro, con mencionarla se le estaría describiendo. Sin embargo, la realidad recorre un amplio rango por lo que requiere ser resumida. La historia natural y/o el curso clínico de la enfermedad no escapan a esta diversidad de expresión, por lo tanto, requieren de un resumen numérico que sitúe tanto a los factores de riesgo, como a las manifestaciones de la enfermedad y su desenlace -pronóstico- dentro de estos rangos.^{1,2,3,4,5,6}

En esta serie de artículos relacionados con el resumen numérico, mostraremos la razón conceptual y la descripción operativa de la información -manejo de datos-, la manera de ejercer la expresión numérica -diseño del análisis de los datos y expresión tabular y gráfica-, pasando por las frecuencias mínimas de datos para garantizar una estimación correcta -tamaño de muestra-.^{1,7,8,9,10,11,12}

I) Manejo de datos “*tras bambalinas*” -recopilación, limpieza, imputación y demarcación-

Este proceso, resulta oscuro y rara vez se describe de manera explícita en los artículos.

1. Recopilación

Durante la **recopilación**, existe el riesgo de cometer una serie de errores, entre los más comunes encontramos: 1.1. La omisión de covariables, 1.2. El desvío del objetivo y 1.3. La calidad insuficiente.

1.1. La **omisión de covariables** que participan en el proceso de causalidad, independientes del efecto de la maniobra principal, es consecuencia de que se les suele atribuir baja relevancia o, incluso, por ignorancia de estas. Esta omisión provoca una distorsión del resultado atribuido a la maniobra principal. Por ejemplo, el beneficio de un tratamiento antineoplásico probado en una población bien nutrida es diferente a lo que ocurre cuando se aplica en pacientes desnutridos, ya que en estos, incluso, puede ser deletéreo por toxicidad.^{13,14} Es así como los pacientes ideales del estudio no son los *reales*, en quienes debemos considerar una serie de covariables antes de poder asumir el beneficio observado. Otra manera en la que se presenta esta distorsión es atribuir el efecto de una maniobra a todos los participantes del estudio, cuando, en realidad, el impacto observado es el promedio obtenido en la población estudiada y para las características predominantes, no para las subrepresentadas. Lo cual significa que hay sujetos que, incluso, se beneficiarían más que el promedio y otros que, además de no beneficiarse, podrían sufrir un efecto deletéreo, dependiendo de las covariables que presenten.

Para evitar la omisión de variables durante la conformación del proyecto de investigación hay que realizar una búsqueda estructurada¹⁵ que identifique los componentes dentro del modelo arquitectónico que participan en el fenómeno de causalidad. Posteriormente, el investigador debe dejar en claro el manejo que realiza con las covariables, ya sea a través una selección muy restrictiva, estratificación y/o ajuste en el análisis.¹⁶

1.2. El **desvío del objetivo primario** es algo común cuando el desenlace resulta poco frecuente, requiere mucho tiempo de seguimiento o se considera subjetivo -por ausencia de un sistema métrico para registrarse-. La consecuencia es la sustitución del objetivo primario por una variable subrogada, pero esto no significa que dicha variable sea equivalente. Por ejemplo, en el caso del paciente diabético durante el estudio de daño a órgano blanco -micro y macrovascular-, dado que se requiere de seguimientos prolongados, se utilizó como subrogado el control glucémico, no obstante, se ha demostrado que el descontrol glucémico no es equivalente a daño a órgano blanco, el descontrol glucémico tiene un número necesario a dañar (NND) de 200-250 sujetos, por lo que resulta un mal subrogado.¹⁷ Sin embargo, cuando la glucosa ha sido utilizada como subrogado de las manifestaciones agudas debidas al descontrol metabólico del paciente diabético, como es la triada característica de polifagia, polidipsia y poliuria, resulta sumamente eficiente.^{18,19}

La recomendación es mantener el desenlace de interés y buscar estrategias que permitan ajustar o completar el tamaño de muestra,^{20,21,22} reconocer grupos de riesgo que requieran menor tiempo de seguimiento y/o menor tamaño de muestra²³ y objetivar la variable cuando se considera subjetiva,²⁴ a través de la descripción metódica -ejemplo: semiología del síntoma o signo-, y/o la agregación de datos asociados. Por ejemplo, en el caso de la cefalea, describir la localización, extensión, duración, cambios y/o agregar síntomas acompañantes como náuseas, vómito, palpitaciones, etc.²⁴

1.3. La **calidad insuficiente** puede ocurrir por *instrumentos inadecuados*, por *falta u omisión del procedimiento de medición*, y por *medición fuera de contexto*. Los dos primeros son muy evidentes,^{25,26} pero la medición fuera de contexto suele pasar desapercibida y darse de diferentes maneras: *atribución a destiempo*, que consiste en atribuir una característica a un tiempo diferente al que se midió, -por ejemplo, cuando el resultado de algún estudio realizado el lunes se atribuye a la condición de ingreso del paciente cuando este ocurrió el viernes-; *atribución equivalente*, considerar una condición actual similar entre dos o más pacientes como

equivalente, sin considerar el tiempo de exposición -por ejemplo, dos pacientes con la misma obesidad, donde uno está en esa condición hace un par de años y el otro desde hace un par de décadas-.^{27,28}

La estrategia para contar con información de calidad suficiente es la descripción conceptual y operativa de las variables de estudio, en estas se describe el porqué, el cómo y cuándo de cada medición.^{29,30}

2. Limpieza de la base de datos

Una vez que los valores correspondientes a cada una de las variables se encuentran en el formato de reporte de casos y/o en la base de datos se contrastan con su referente, con el propósito de identificar valores extremos factibles pero poco probables y valores faltantes. Estos valores se revisan contra el documento fuente para garantizar que son reales y no un error de transcripción o captura. Cuando los valores extremos son reales hay que considerar la deformación que provocan a la medida de resumen y/o de dispersión de la variable a la que corresponden, razón por la que habrá que considerar alguna estrategia para su uso, como eliminar aquellos valores cuantitativos que se encuentran fuera de 3 desviaciones estándar, mismos que posteriormente podrán ser imputados con las reglas correspondientes, o transformar la variable a ordinal o dicotómica con lo que al integrarse a su categoría correspondiente dejaría de ser un valor extremo.³¹

3. Imputación

Manejo de valores faltantes. Inicialmente, hay que identificar los valores (datos) faltantes por variable. Y si bien no existe una regla precisa de hasta qué porcentaje de valores faltantes se les debe atribuir un valor estimado, en general se hace cuando los datos faltantes de una variable no rebasan el 10%; cuando los valores faltantes superan el 20% no se imputan, y cuando se encuentran en algún punto entre > 10 y 20% la decisión será de acuerdo con la experiencia de los investigadores o por lo referido en la literatura para esa variable.^{32,33} *Nunca se sustituirán aquellos valores faltantes perteneciente al desenlace, ni los de la maniobra principal.*

3.1. Si consideramos la **imputación** (atribución) de un valor estimado, cuando los valores faltantes en una variable son menores o iguales al 10%, la manera más común de hacerlo es por la medida de resumen de dicha variable calculada con los valores registrados, así será con la media cuando se trate de una variable cuantitativa con distribución normal, la *mediana* en caso de una variable cuantitativa sin distribución normal o de una

variable ordinal y la *moda* para una variable dicotómica. Existen otras maneras de obtener el valor estimado, como la predictiva, donde la variable con datos faltantes se trata de explicar a partir de otras variables relacionadas por medio de un modelo de regresión (nunca incluir como variables predictoras la maniobra principal, ni al desenlace). Sin embargo, en estos casos hay que prevenir algunos problemas posteriores, como colinealidad entre la variable explicada y dichas variables.^{34,35,36}

3.2. Cuando se decide la **no imputación** de los valores faltantes se crea para dicha variable una categoría de *valor faltante* (en inglés, *missing data*); una variable inicialmente cuantitativa se transforma en ordinal o dicotómica, a la variable ordinal o dicotómica se le añade una categoría correspondiente a valor faltante o *missing data*. La razón de crear una categoría de valor faltante o *missing data* es para que los pacientes que presenten esta categoría en dichas variables no sean excluidos del análisis multivariado -3 o más variables-. Si en 2 variables, faltan 5 valores en pacientes diferentes, en el análisis bivariado faltarán esos 5 pacientes, pero provocará que en el multivariado falten 10. Si esto ocurriese en 3 variables en el bivariado continuarían faltando solo 5 en cada variable, pero en el multivariado 15 y así sucesivamente. La inclusión de una categoría de *missing data* permite evaluar el efecto de la variable en el análisis multivariado sin excluir pacientes, ponderando dichos valores faltantes, de esta forma se mantendrá el tamaño total de la población.³⁷

3.3. **Cuando el valor faltante está en la maniobra o en el desenlace este no se imputa.** Para el caso de la maniobra se genera la categoría de valor faltante o *missing data*, mientras que para el desenlace el valor se deja vacío y durante la descripción del curso clínico se hará con la totalidad de la información existente hasta su última medición.

Por ejemplo, si en un estudio solo se hubiesen perdido los pacientes al momento del desenlace, al realizar el análisis se puede mostrar la comparación de las condiciones basales con la maniobra en la totalidad de la población (lo cual habitualmente aparece en el primer cuadro de los artículos), se puede mostrar la calidad de la maniobra principal y su relación con las maniobras periféricas (habitualmente se muestra en el cuadro II de los artículos). Sin embargo, al mostrar la relación de la maniobra principal, condiciones basales y maniobras periféricas contra el desenlace solo se incluirán los pacientes en quienes se tiene dicha medición (habitualmente se muestra en un artículo en los cuadros III y IV).

Para el caso de los eventos adversos se muestran los resultados con el total de la población.

3.4. Un error común causado por la falta de valores, sobre todo de la maniobra o el desenlace, es **dejar fuera del estudio a pacientes**, provocando un desvío en el espectro de la enfermedad, ya que estos pacientes pudieron haber cursado con alguna característica particular que impidió que se midieran dichos valores.^{1,12,38,39,40} Incluirlos, referir la situación y reportar el tipo de acciones que se realizaron con la información faltante es lo que resulta pertinente, y le ofrece al lector una idea de la probable influencia de dichos datos faltantes sobre los resultados.

3.5. Una situación diferente es la **exclusión del paciente a consecuencia del diseño del estudio**. Recordemos que en los estudios de casos y controles, y en la encuesta transversal, la pesquisa se realiza sin relación al ensamble de la población, sino más bien con relación al desenlace. En consecuencia, los participantes son los supervivientes de una o múltiples cohortes preexistentes, razón por la cual estos diseños siempre se acompañan del sesgo de transferencia, con la consecuente sobreestimación de la buena evolución del paciente.⁴¹ De igual manera, este fenómeno se observa en los estudios que recientemente se han popularizado como “*Estudios de la Vida Real*”, que son estudios que evalúan el efecto al manejo con diferentes maniobras, entre ellas las farmacológicas, donde solo continúan con dichas maniobras quienes han tenido una respuesta aceptable y quienes no han presentado grandes efectos adversos.

4. Demarcación

El último apartado del manejo de datos “*tras bambalinas*” se refiere a los puntos de demarcación de los valores al interior de cada variable. Las variables que serán extraídas del interrogatorio, de la exploración, los datos de laboratorio o gabinete, o del expediente clínico, son aquellas que permiten contar la historia o el curso clínico de la enfermedad en relación con el desenlace primario de interés. La demarcación de cada valor, al interior de cada variable, sigue la lógica de contar con un significado en referencia a dicho desenlace y, consecuentemente, redundante en utilidad clínica. Para ello, es pertinente considerar una secuencia jerárquica de criterios de referencia en base a: 1) estudio clínico previo, 2) acuerdo de expertos, 3) juicio clínico del investigador (o grupo de investigadores) y 4) estadística.

4.1. El uso de **puntos de corte descritos en estudios clínicos previos** que hayan demostrado utilidad resulta el criterio ideal. Incluso, cuando se plantean nuevos puntos de corte es indispensable inicialmente mostrar reproducibilidad de los resultados en relación con los puntos de corte de estudios previos. Si no existen

puntos de corte previos en la misma condición basal (enfermedad, estadio, etc.) y desenlace de interés (supervivencia, curación, etc.), se pueden explorar los puntos de corte de condiciones basales similares con el mismo desenlace o de distinto desenlace en la misma condición basal, y en el peor de los casos de condiciones basales o desenlaces similares.^{42,43}

4.2. Cuando no existe una referencia clínica previa como resultado de investigación se pueden utilizar los **puntos de corte referidos por un grupo de expertos**, como son los consensos de especialistas sobre el tema de interés.^{44,45}

4.3. El **clínico o grupo de clínicos** que ha generado la hipótesis, además de contar con experiencia sobre el tema, debieron revisar de manera exhaustiva durante la construcción de la justificación la existencia de respuesta, o de posibles respuestas a la hipótesis, y haber previsto los puntos de corte clínicamente relevantes.⁴⁶

4.4. Cuando no es posible ninguna de las estrategias previas se puede recurrir a la **estadística para definir los puntos de corte**. Dos modalidades comunes son: 1) utilizar **percentiles** en los que se garantiza distribución homogénea del número de sujetos en cada apartado (ejemplo, quintiles que divide a la totalidad de sujetos en 5 partes iguales: 150 sujetos / 5 grupos = 30 sujetos en cada grupo, del quintil 1 donde están los 30 sujetos con los menores valores de la variable, quintil 5, donde están los 30 sujetos con los mayores valores de la variable),^{47,48} o 2) utilizar puntos de demarcación de la variable, que pueden ser o no equidistantes: los equidistantes garantizan la misma repartición de valores entre una categoría y otra (ejemplo, 5 puntos de demarcación para una variable con valores de 1 a 15, primer punto de demarcación es con los valores de 1-3, quinto punto de demarcación es con valores de 13-15.), mientras que los no equidistantes podrán seguir cualquier distribución de valores. En ambos casos puede ser que en algunas de las categorías se acumulen más sujetos que en otras, e incluso puede haber categorías sin sujetos.^{49,50}

Consideraciones finales

Emplear de forma adecuada los datos permite describir con mayor precisión la historia natural y el curso clínico de la enfermedad. La industria farmacéutica tiene gran cuidado sobre el manejo de los datos, pues el conocimiento que tiene sobre el tema, y la alta inversión económica que se realiza para el monitoreo de estos, garantiza la calidad de información de sus estudios. Lamentablemente, en la mayoría de las instituciones de salud no existen controles

de calidad para el proceso de manejo de datos, por lo que resulta difícil garantizar la veracidad de los resultados. No es necesario actuar de mala fe por parte de los generadores de resultados, basta con actuar sin controles de calidad para mentir involuntariamente y confundir, en lugar de aclarar. Por lo tanto, se exhorta a establecer al interior de nuestras instituciones de salud controles de calidad en el manejo de la información.

Declaración de conflicto de interés: los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno que tuviera relación con este artículo.

Cuadro I Manejo de datos

Recopilación	Omisión de covariables	Omitir variables que participan en el proceso de causalidad puede distorsionar el resultado
	Desvío del objetivo	Cuando el desenlace resulta poco frecuente, requiere mucho tiempo de seguimiento o se considera subjetivo se suele sustituir el objetivo primario por una variable subrogada, pero esto no significa que dicha variable sea equivalente. Cuando eso sucede, se deben interpretar con cautela los resultados
	Calidad insuficiente	Puede ocurrir por instrumentos inadecuados, por falta u omisión del procedimiento de medición, y por medición fuera de contexto, atribución a destiempo o equivalente
Limpieza de base de datos	Revisar valores extremos en documento fuente y verificar datos faltantes Si son reales, aquellos que se encuentran fuera de 3 desviaciones estándar considerar eliminar el dato	
Imputación (atribución de un valor estimado) cuando:	Imputar con media	Variable cuantitativa de distribución normal
	Imputar con mediana	Variable cuantitativa sin distribución normal o variable ordinal
	Imputar con moda	Variable dicotómica
	Se crea una categoría de "Valor faltante o <i>missing data</i> "	No se imputa el dato, solo se crea una categoría más, para ponderar su efecto y no excluir pacientes en el análisis multivariado
Datos faltantes	<ul style="list-style-type: none"> • ≤ 10%: Imputar • > 10-20%: Considerar imputar • > 20%: No imputar 	
Nunca imputar datos de maniobra principal o desenlace		
Demarcación de los datos dentro de cada variable	<p>Es pertinente considerar una secuencia jerárquica de criterios de referencia en base a: 1) estudio clínico previo, 2) acuerdo de expertos, 3) juicio clínico del investigador (o grupo de investigadores) y 4) estadística</p> <p>El uso de puntos de corte descritos en estudios clínicos previos que hayan demostrado utilidad resulta el criterio ideal. De no existir, se pueden utilizar los puntos de corte referidos por un grupo de expertos o por el grupo de clínicos que realizan la investigación. Cuando no es posible ninguna de las estrategias previas se puede recurrir a la estadística para definir los puntos de corte (percentiles o puntos de demarcación de la variable)</p>	

Referencias

1. Talavera JO, Roy-García IA, Pérez-Rodríguez M, Palacios-Cruz L, Rivas-Ruiz R. De vuelta a la clínica. Métodos II. Arquitectura de la investigación clínica. Interacción sujeto, maniobras y enfermedad a través del tiempo. *Gac Med Mex*. 2020;156(5):438-446. doi:10.24875/GMM.20000159
2. Talavera J. Juicio Clínico: el método científico aplicado a la clínica. *Rev Med Inst Mex Seguro Soc*. 2019;57(5):267-268.
3. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health*. 2005;95(SUPPL. 1):S144-S150.
4. Evans AS, Henle W, Huebner R, Johnson R, Lilienfeld A. Causation and Disease: The Henle-Koch Postulates Revisited. *Yale J Biol Med*. 1976;49:175-195.
5. Yekushalmu J, Palmer CE. On the methodology of investigations of etiologic factors in chronic diseases. *Journal of Chronic Disease*. Published online 1959:27-40.
6. Elster J. *Ulises y Las Sirenas*. 1st ed. Fondo de Cultura Económica. México; 1979.
7. Bradford-Hill A. The Environment and Disease: Association or Causation? *Proc R Soc Med*. Published online. 1965:295-300.
8. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. (Seigafuse S, Bierig L, eds.). Wolters Kluwer, Lippincott Williams & Wilkins; 2008.
9. Spinoza B. *Ethics Proved in Geometrical Order*. 1st ed. (Kisner MJ, ed.). Cambridge University Press; 2018.

10. Feinstein AR. Clinical biostatistics XLVII. Scientific standards vs. statistical associations and biologic logic in the analysis of causation. *Clin Pharmacol Ther.* 1979;25(4):481-492. doi:10.1002/cpt1979254481
11. Feinstein AR. Statistical reductionism and clinicians' delinquencies in humanistic research. *Clin Pharmacol Ther.* 1999; 60(3):211-217.
12. Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Investigación clínica III. Estudios de Causalidad. *Rev Med Inst Mex Seguro Soc.* 2011;49(3):289-294.
13. López-Facundo N, Talavera JO, Tejocote-Romero I. Mortalidad temprana en niños con leucemia linfoblástica aguda en un país en vías de desarrollo; factores asociados con el pronóstico. *Gaceta Médica de Oncología.* 2008;7(3):93-101.
14. López-Facundo NA, Tejocote-Romero I, Rodríguez-Castillejos C, Jaimes-García Y. Impacto de la obesidad en el pronóstico de supervivencia y recaída en niños con leucemia aguda linfoblástica del estado de México. *Gaceta Mexicana de Oncología.* 2015;14(5):242-249. doi: 10.1016/j.gamo.2015.11.004
15. Rivas-Ruiz R, Talavera JO. Investigación clínica VII. Búsqueda sistemática: cómo localizar artículos para resolver una pregunta clínica. *Rev Med Inst Mex Seguro Soc.* 2012;50(1): 53-58.
16. Horwitz R, Singer BH, Makuch RW, Viscoli CM. Can Treatment That Is Helpful on Average Be Harmful to Some Patients? A Study of thk Conflicting information Needs of Clinical Inquiry and Drug Regulation. *J Clin Epidemiol.* 1996;49 (4):395-400.
17. UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *The Lancet.* 352AD;352:855-865.
18. Donner T, Muñoz M. Update on Insulin Therapy for Type 2 Diabetes. *J Clin Endocrinol Metabol.* 2012;97(5):1405-1413.
19. Delahanty LM, Nathan DM. Research Navigating the Course of Clinical Practice in Diabetes. *J Am Diet Assoc.* 2004;104 (12):1846-1853. doi: 10.1016/j.jada.2004.09.028
20. Guyatt G, Sackett D, Taylor W, Chong J, Roberts R, Pugsley S. Determining optimal therapy- Randomized Trials in individual patients. *New England Journal of Medicine.* 1986;314 (14):889-892.
21. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The n-of-1 Randomized Controlled Trial: Clinical Usefulness. Our Three-Year Experience. *Ann Intern Med.* 1990;112:293-299.
22. Levin WC, Fink DJ, Porter FS, Hall TC, Loeb V. Cooperative Clinical Investigation. A Modality of Medical Science. *JAMA.* 1974;227(11):1295-1296.
23. Feinstein AR. Clinical biostatistics XLVIII. Efficacy of different research structures in preventing bias in the analysis of causation. *Clin Pharmacol Ther.* 1979;26(1):129-141. doi: 10.1002/cpt1979261129
24. Feinstein AR. Clinical biostatistics XLI. Hard science, soft data, and the challenges of choosing clinical variables in research. *Clin Pharmacol Ther.* 1977;22(4):485-498. doi: 10.1002/cpt1977224485
25. Streiner DL, Norman GR. "Precision" and "accuracy": Two terms that are neither. *J Clin Epidemiol.* 2006;59:327-330. doi: 10.1016/j.jclinepi.2005.09.005
26. Feinstein AR. *Clinometrics.* 1st ed. Yale University Press; 1987.
27. Lawlor DA, Hart CL, Hole DJ, Smith GD. Reverse Causality and Confounding and the Associations of Overweight and Obesity with Mortality. *Obesity.* 2006;14(12):2294-2304.
28. Banack HR, Bea JW, Kaufman JS, et al. The effects of reverse causality and selective attrition on the relationship between Body Mass Index and mortality in postmenopausal women. *Am J Epidemiol.* 2019;188(10):1838-1848. doi: 10.1093/aje/kwz160
29. Vasudeva D, Asi R. An examination of the role of conceptualization and operationalization in empirical social research. *ZENITH International Journal of Multidisciplinary Research.* 2013;3:108-114.
30. Feinstein AR. Clinical biostatistics. XLV. The purposes and functions of criteria. *Clin Pharmacol Ther.* 1978;24(4):479-492. doi: 10.1002/cpt1978244479
31. Dixon WJ. Analysis of Extreme Values. *The Annals of Mathematical Statistics.* 1950;21(4):488-506.
32. Portney LG. *Foundations of Clinical Research: Applications to Evidence-Based Medicine.* 4th ed. F.A. Davis; 2020.
33. Streiner DL. *Statistics Commentary Series: Commentary # 10 - Dealing with Drop-Outs.* *J Clin Psychopharmacol.* 2015;35 (5):496-498. doi: 10.1097/JCP.0000000000000299
34. Little RJA, Rubin DB. *Statistical Analysis with Missing Data.* 3rd ed. Wiley; 2020.
35. Nguyen CD, Carlin JB, Lee KJ. Practical strategies for handling breakdown of multiple imputation procedures. *Emerg Themes Epidemiol.* 2021;18(5). doi: 10.1186/s12982-021-00095-3
36. Padgett CR, Skilbeck CE, Summers MJ. Missing data: The importance and impact of missing data from clinical research. *Brain Impairment.* 2014;15(1):1-9. doi: 10.1017/BrImp.2014.2
37. Steele AJ, Denaxas SC, Shah AD, Hemingway H, Luscombe NM. Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. *PLoS One.* 2018; 13(8). doi: 10.1371/journal.pone.0202344
38. Feinstein AR. Clinical biostatistics III. The architecture of clinical research. *Clin Pharmacol Ther.* 1970;11(3):432-441. doi: 10.1002/cpt1970113432
39. Feinstein AR. Clinical biostatistics IV. The architecture of clinical research (continued). *Clin Pharmacol Ther.* 1970;11(4):595-610. doi: 10.1002/cpt1970114595
40. Feinstein AR. Clinical biostatistics V. The architecture of clinical research (concluded). *Clin Pharmacol Ther.* 1970;11(5):755-771. doi: 10.1002/cpt1970115755
41. Talavera JO, Roy-García I, Palacios-Cruz L, Rivas-Ruiz R, Hoyo I, Pérez-Rodríguez M. De vuelta a la clínica. Métodos I. Diseños de investigación. Mayor calidad de información, mayor certeza a la respuesta. *Gac Med Mex.* 2019;155(4):399-405. doi: 10.24875/GMM.19005226
42. Belk RA, Pilling M, Rogers KD, Lovell K, Young A. The theoretical and practical determination of clinical cut-offs for the British Sign Language versions of PHQ-9 and GAD-7. *BMC Psychiatry.* 2016;16(372):1-12. doi: 10.1186/s12888-016-1078-0
43. He L, Khanal P, Morse CI, Williams A, Thomis M. Differentially methylated gene patterns between age-matched sarcopenic and non-sarcopenic women. *J Cachexia Sarcopenia Muscle.* 2019;10(6):1295-1306. doi: 10.1002/jcsm.12478
44. Gordijn SJ, Beune IM, Thilaganathan B, et al. Consensus definition of fetal growth restriction: a Delphi procedure. *Ultrasound Obstet Gynecol.* 2016;48:333-339. doi: 10.1002/uog.15884
45. Tinazzi M, Geroi C, Bhidayasiri R, et al. Task Force Consensus on Nosology and Cut-Off Values for Axial Postural Abnormalities in Parkinsonism. *Mov Disord Clin Pract.* 2022; 9(5):594-603. doi: 10.1002/mdc3.13460

46. Duckett Jones T. The Diagnosis of Reumatic Fever. *JAMA*. 1944;126(8):481-484.
47. Xu H, Simonet F, Luo ZC. Optimal birth weight percentile cut-offs in defining small- or large-for-gestational-age. *Acta Paediatr*. 2010;99(4):550-555. doi: 10.1111/j.1651-2227.2009.01674.x
48. Brown JD, Alipour-Haris G, Pahor M, Manini TM. Association between a Deficit Accumulation Frailty Index and Mobility Outcomes in Older Adults: Secondary Analysis of the Lifestyle Interventions and Independence for Elders (LIFE) Study. *J Clin Med*. 2020;9(11):1-13. doi: 10.3390/jcm9113757
49. González R, Hidalgo G, Salazar J, Preciado M. Instrumento Para Medir La Calidad de Vida En El Trabajo CVT-GOHISA-LO. Manual Para Su Aplicación e Interpretación. 1st ed. Editorial de la Luna; 2010.
50. Gholami T, Pahlavian AH, Akbarzadeh M, Motamedzade M, Moghaddam RH. The role of burnout syndrome as a mediator for the effect of psychosocial risk factors on the intensity of musculoskeletal disorders: a structural equation modeling approach. *International Journal of Occupational Safety and Ergonomics*. 2016;22(2):283-290. doi: 10.1080/10803548.2016.1147876