

MInCir-Pr2: Escala para Determinar Calidad Metodológica de Artículos Primarios de Pronóstico. Explicación de sus Dominios e Ítems, y una Pauta para su Aplicación

MInCir-Pr2: A Scale for Determining the Methodological Quality of Primary Prognosis Articles Explanation of its Domains and Items, and a Guideline for its Application

Carlos Manterola^{1,2,3}; Juan Pablo Holguín-Carvajal^{1,4}; Josue Rivadeneira^{1,3} & Tamara Otzen^{1,3}

MANTEROLA, C.; HOLGUÍN-CARVAJAL, J. P.; RIVADENEIRA, J. & OTZEN, T. MInCir-Pr2: Escala para determinar calidad metodológica de artículos primarios de pronóstico. Explicación de sus dominios e ítems, y una pauta para su aplicación. *Int. J. Morphol.*, 42(5):1394-1402, 2024.

RESUMEN: La evaluación de la calidad metodológica (CM) en la investigación biomédica es un área que ha sufrido un desarrollo dinámico en los últimos 20 años a nivel global. CM se puede definir como un constructo multidimensional y complejo compuesto por ítems y dominios. Distintos grupos de investigación han propuesto herramientas para su valoración, entre los cuales se encuentra nuestro equipo con el desarrollo de las escalas MInCir. Con el paso de los años, hemos diseñado y validado escalas para valorar CM de estudios de tratamiento o procedimientos terapéuticos (MInCir-T), diagnóstico (MInCir-Dg) y pronóstico (MInCir-Pr en sus dos versiones, MInCir-Pr1 y MInCir-Pr2). Sin embargo, como ocurre con cualquier instrumento de medición, es fundamental especificar de forma detallada y precisa la forma en que se aplica, de forma tal de poder estandarizar las mediciones y reducir los sesgos de información. Se presenta una descripción pormenorizada de los 4 dominios y 8 ítems que componen la escala MInCir-Pr2, precisando para cada uno de ellos las características a valorar y puntuar en los artículos primarios de pronóstico, con una sumatoria que puede dar una puntuación total entre 7 y 60 puntos, con un punto de corte para definir el constructo CM de 33 puntos (dicotómico), que permite clasificar como CM adecuada o alta, cuando el artículo evaluado recibe una puntuación de 33 o más puntos; y CM inadecuada o baja, si la puntuación es de 32 puntos o menos. El objetivo de este artículo fue exponer y aclarar la aplicación de la escala MInCir-Pr2 de CM para estudios de pronóstico, mediante una guía que pretende estandarizar las mediciones, para optimizar la confiabilidad intra e inter observador.

PALABRAS CLAVE: Escalas; Reproducibilidad de resultados; Validez de las pruebas; Calidad metodológica; Pronóstico.

INTRODUCCIÓN

La investigación clínica requiere un enfoque sistemático con una planificación meticulosa y ejecución para obtener resultados válidos y confiables. Por lo tanto, es crucial que los investigadores posean una comprensión integral de los distintos tipos de diseños y metodologías de investigación, y de su impacto en la calidad metodológica (CM) final del producto (Kiani *et al.*, 2022).

Desde la década de 1990, se han desarrollado diversas herramientas para evaluar y mejorar la calidad de la investigación biomédica (Malic'ki *et al.*, 2019). Sin embargo, aún falta un concepto claro sobre lo que es necesario y cómo se deben informar los resultados de la investigación en salud

(Armijo-Olivo *et al.*, 2021; Muchadeyi *et al.*, 2022). Por lo tanto, se han ideado diversos enfoques para abordar esta faceta del proceso, como el desarrollo de instrumentos para medir CM de los estudios (da Silva *et al.*, 2013; Ma *et al.*, 2020; Sargeant *et al.*, 2022; Prill *et al.*, 2023). Sin embargo y a pesar de todo, algunas de estas herramientas, como la escala de Newcastle-Ottawa (NOS), han sido muy criticadas (Stang *et al.*, 2010; Stang *et al.*, 2018).

La CM puede conceptualizarse como un constructo complejo y multidimensional, que puede incluir diversos elementos en forma de ítems y dominios como el tipo de diseño, el tamaño de la muestra utilizada, diversos aspectos

¹ Programa de Doctorado en Ciencias Médicas, Universidad de La Frontera, Chile.

² Centro de Estudio Morfológicos y Quirúrgicos (CEMyQ), Universidad de La Frontera, Chile.

³ Núcleo de Sociomedicina. SOCIOMED

⁴ Universidad del Azuay, Ecuador.

metodológicos, la calidad del análisis realizado, la calidad de la presentación, entre otros (Manterola *et al.*, 2006). Este concepto puede representarse mediante una figura geométrica con tantos lados y ángulos como ítems o dimensiones se consideren en la definición del constructo (Fig. 1).

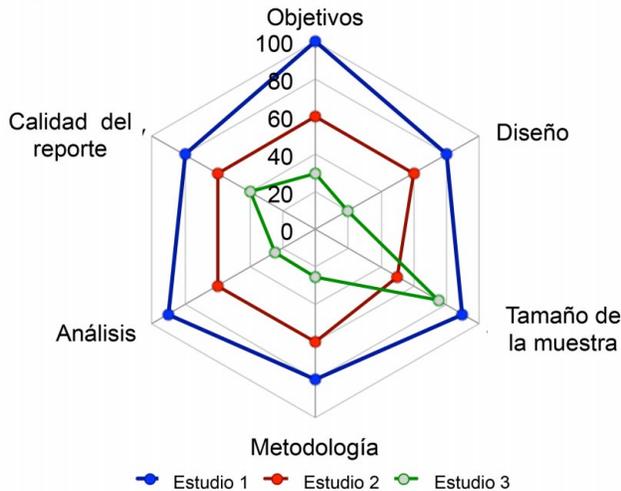


Fig. 1. CM de un artículo científico (concepto o constructo multidimensional). Se aprecia un gráfico polar en el que se miden 6 dominios para explicar el constructo de CM. Se puede observar la configuración de tres artículos primarios que ocupan diferentes áreas del hexágono. El número uno indica aquel con el mejor desempeño del constructo CM.

Hace 20 años, desarrollamos una escala para evaluar CM en estudios sobre tratamientos o procedimientos terapéuticos (Manterola *et al.*, 2006, 2009; Moraga *et al.*, 2013; Aravena *et al.*, 2013; Cartes-Velasquez *et al.*, 2014). Esto nos permitió no solo llevar a cabo estudios de medición de CM, sino que también estudios bibliométricos y revisiones sistemáticas (RS), con una evaluación ponderada de la evidencia basada en la CM de los estudios primarios que generaron dicha evidencia (Manterola *et al.*, 2009), superando las limitaciones inherentes al método tradicional de RS con metaanálisis, que consideraba exclusivamente a los ensayos clínicos (EC) como estudios primarios.

Posteriormente, contribuimos al desarrollo de tres escalas adicionales para determinar CM. Una para evaluar la CM en estudios diagnósticos o de pruebas diagnósticas, a la que denominamos MInCir-Dg (Burgos *et al.*, 2011; Manterola *et al.*, 2016); y otras dos para evaluar CM en estudios de pronóstico, MinCir-Pr1 y MinCir-Pr2 (Losada *et al.*, 2009; Zavando, 2011; Manterola *et al.*, 2015, 2018). Estos instrumentos también se han aplicado en distintas RS (Castillo *et al.*, 2021; Duque *et al.*, 2022a,b), lo que ha permitido establecer la viabilidad de su aplicación (Carvajal *et al.*, 2011), como también su utilidad como escala de medición (Luján-Tangarife & Cardona-Arias, 2015).

Sin embargo, la validez y confiabilidad de las mediciones dependen no solo de las propiedades psicométricas del instrumento a utilizar, sino también de su correcta aplicación por diferentes investigadores, especialmente al medir una variable no física, como es el constructo CM (Manterola *et al.*, 2009; Keszei *et al.*, 2010).

El objetivo de este artículo fue elucidar y clarificar la aplicación de la escala MinCir-Pr2 para determinar CM en estudios de pronóstico, a través de una guía destinada a estandarizar las mediciones, con el fin de mejorar tanto la confiabilidad intraobservador como la interobservador.

OBJETIVOS Y USO DE MInCir-Pr2

MInCir-Pr2 es una escala para determinar CM en estudios primarios en los que se haya estudiado un escenario clínico de pronóstico; es decir estudios que permitan determinar el riesgo de presentar uno o más eventos de interés en un grupo de personas con una condición determinada. Por ejemplo, estudios de cohortes (prospectivas y retrospectivas), estudio de casos y controles, estudios de corte transversal, entre otros.

Su intención es proporcionar orientación sobre cómo fueron desarrollados algunos aspectos metodológicos fundamentales de un estudio de pronóstico (diseño, variable resultado/exposición, seguimiento, análisis, etc.), por ende, no constituyen recomendaciones para diseñar o llevar a cabo este tipo de estudios.

DESARROLLO DE MInCir-Pr2

Los primeros pasos del desarrollo de MInCir-Pr comenzaron en 2006, con una tesis para la obtención del grado de Magíster de uno de nuestros colaboradores. En esa oportunidad, se realizó una búsqueda de los diferentes ítems para la construcción del primer borrador, utilizando estrategias cualitativas. Posteriormente, se realizó un análisis factorial para reducción de ítems y dominios con lo que se construyó el segundo borrador con el cual se determinó consistencia interna de los ítems, aplicando alfa de Cronbach (Losada *et al.*, 2009). Esto dio origen a MInCir-Pr1.

Posteriormente, en el desarrollo de otra tesis de Magíster, el diseño de la escala se desarrolló en tres pasos: selección de ítems para generar el primer borrador (revisión de la literatura de CM de estudios pronósticos en BIREME-BVS, PubMed, OVID, Scopus, WoS y SciELO). Luego, un panel de 5 expertos (epidemiólogos clínicos y bioestadístico), sugirió los ítems y dominios para construir el constructo basándose en la revisión de la literatura y su experiencia personal en la materia. Luego, se llevó a cabo un estudio de

validez de contenido que permitió generar el segundo borrador (el que fue evaluado por investigadores con títulos de maestría o doctorado en Ciencias Médicas, y con al menos dos publicaciones en revistas de la base de datos WoS relacionadas con CM, provenientes de EE. UU., España y Chile). Finalmente, se realizó un estudio piloto con estudiantes de posgrado de los programas de doctorado y maestría en ciencias médicas de nuestra universidad, con el objetivo de optimizar el uso y la comprensión de la escala (de esta manera, se obtuvo el tercer borrador), que dio posteriormente origen a MInCir-Pr2 (Zavando, 2011; Manterola *et al.*, 2015, 2018).

COMPONENTES DE MInCir-Pr2. La escala MInCir-Pr2 está compuesta por 4 dominios y 8 ítems (Tabla I). El Dominio 1 (de un solo ítem) considera el tipo de diseño; el Dominio 2 (también de un solo ítem) asigna puntuaciones al tamaño de la población estudiada, incluyendo un factor de justificación (si el estudio incluye el cálculo del tamaño de muestra necesario para su realización); el Dominio 3 consta de 5 ítems que evalúan la metodología del estudio; y el Dominio 4 consta de 4 ítems que evalúan el análisis estadístico y las conclusiones del estudio (Manterola *et al.*, 2018).

Tabla I. MInCir-Pr2. Escala para determinar CM de estudios de pronóstico.

Dominios e ítems de la escala		Score
Dominio 1: Diseño del estudio		
Estudio de cohorte concurrente o prospectiva	EC CAA con doble enmascaramiento	15
Estudio de cohorte histórica o retrospectiva	EC SAA enmascaramiento simple o ausente	10
Estudio de casos y controles		8
Estudio de corte transversal		6
Reporte de casos y series de casos		3
Dominio 2: Población en estudio x factor de justificación *		
n > 501		7 o 15
n = 201 – 500		6 o 12
n = 151- 200		5 o 10
n = 101 – 150		4 o 8
n = 51 – 100		3 o 6
n = 31 – 50		2 o 4
n ≤ 30		1 o 2
Dominio 3: Metodología		
Objetivos		
Objetivos claros, precisos y concretos		3
Objetivos vagos		2
Sin objetivos		1
Diseño utilizado		
Se identifica claramente el tipo de diseño utilizado		3
Diseño no claro, o no mencionado		1
Variables		
Variable resultado definida adecuadamente		1
Variable de exposición definida adecuadamente		1
Variables de confusión definidas adecuadamente		1
Tamaño de la muestra		
Incluye estimación o cálculo del tamaño de la muestra		3
No incluye estimación o cálculo del tamaño de la muestra		1
Seguimiento (follow-up)		
Menciona el porcentaje de pérdida de seguimiento de la(s) cohorte (s)		1
El seguimiento de la(s) cohorte (s) es mayor de 80%		1
Se explican las causas de pérdida de seguimiento de la(s) cohorte (s)		1
Dominio 4: Análisis y conclusiones		
Incluye cálculo de medidas de riesgo		5
Los datos reportados permiten realizar el cálculo de medidas de riesgo		2
Incluye modelos predictivos o de asociación		5
Existe consistencia entre objetivo, metodología y resultados		3
Total		7 - 60

ECCAA: Ensayo clínico controlado con asignación aleatoria / ECSAA: Ensayo clínico sin asignación aleatoria. Estudios que justifican el tamaño de la muestra empleado tienen puntaje doble.

IMPLICANCIAS Y LIMITACIONES DE MInCir-Pr2. La escala MInCir-Pr2 permite determinar CM en estudios primarios de pronóstico, facilitando el trabajo de lectores, autores, revisores y editores de revistas biomédicas. De la misma manera, puede utilizarse para llevar a cabo estudios bibliométricos y para ponderar la calidad de la evidencia a través de metaanálisis de la información obtenida de RS en escenarios de pronóstico con estudios primarios de diferentes tipos de diseños.

Por otra parte, MInCir-Pr2 no debe interpretarse como un intento de prescribir la presentación de una investigación ni como una lista de chequeo o de verificación.

ESTUDIO DE VALIDACIÓN DE MInCir-Pr2. La escala MInCir-Pr2 posee validez de fachada y contenido determinada a través de un panel de expertos, así como validez de criterio, y confiabilidad interobservador (Manterola *et al.*, 2009).

Validez de criterio: La validez de criterio se determinó contrastando las puntuaciones de CM con la propuesta de niveles de evidencia del Centro de Medicina Basada en la Evidencia de Oxford (The Centre for Evidence-Based Medicine, 2009). Se determinó mediante un ANOVA de un factor y una prueba de Duncan. Los niveles de evidencia de los artículos fueron 1b (21 artículos; 17,6 %), 2b (15 artículos; 12,6 %) y 4 (83 artículos; 69,8 %), que tuvieron puntuaciones promedio de CM de 45,3±3,0; 40,8±2,8; y 27,3±2,5, respectivamente ($p < 0,0001$).

Validez de constructo: La validez de constructo se determinó mediante un análisis de grupos extremos al dicotomizar el factor de impacto (FI) de las 13 revistas de las cuales de forma aleatoria se obtuvieron 119 artículos primarios referentes a pronóstico. Las revistas fueron agrupadas según su FI en: alto

[$\geq 4,0$], medio [3,99 a 0,50] y bajo [$\leq 0,49$]. La muestra se obtuvo aplicando los criterios de Streiner (al menos 10 artículos por ítem 1 de escala [Streiner & Geoffrey, 2003]). Se determinó aplicando t-test. Los artículos con alto y bajo FI obtuvieron puntuaciones medias de CM de 38,8±8,0 y 27,6±7,6 respectivamente ($p < 0,0001$).

Confiabilidad interobservador: Utilizando las dos evaluaciones independientes (niveles de evidencia de Oxford y FI), se determinó el grado de acuerdo entre los evaluadores. Se determinó mediante coeficiente de correlación intraclase, el que fue de 0,91 ($p = 0,1082$).

Luego, se confeccionaron curvas ROC, se calcularon sensibilidad, especificidad, valores predictivos, razones de verosimilitud y área bajo la curva según los puntos de corte. El área bajo la curva ROC fue del 82,6 % (Fig. 2). El análisis de estos parámetros permitió determinar que el mejor punto de corte para definir el constructo CM fue 33 (Tabla II).

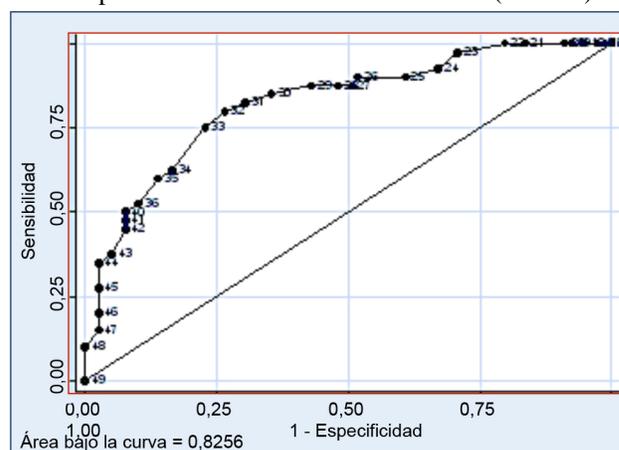


Tabla II. MInCir-Pr2. Escala para determinar CM de estudios de pronóstico. Parámetros psicométricos de algunos puntos de corte.

Parámetros	Puntos de corte			
	31	32	33	34
Sensibilidad (%)	85,0	82,5	80,0	75,0
Especificidad (%)	64,6	69,6	73,4	77,2
Valor predictivo positivo (%)	79,4	79,5	82,6	83,0
Valor predictivo negativo (%)	73,1	73,9	72,0	66,7
LHR (+) o RVP	2,39	2,72	3,01	3,29
LHR (-) o RVN	0,23	0,25	0,27	0,32
Clasificación correcta (%)	71,4	73,9	75,6	76,5
Área bajo la curva (%)	78,1	80,9	82,6	81,9
Asociación artículos/escala (OR)	10,6	10,9	12,2	9,8

LHR (+): Likelihood ratio positivo, o razón de verosimilitud positiva (RVP). LHR (-): Likelihood ratio negativo, o razón de verosimilitud negativa (RVN). OR: Odds ratio

El punto de corte 33 muestra un área bajo la curva del 82,6 %, una sensibilidad del 80,0 %, un valor predictivo positivo del 82,6 %, y razones de verosimilitud positiva y negativa de 3,01 y 0,27, respectivamente (Tabla II, Fig. 2).

INSTRUCCIONES PARA SU UTILIZACIÓN

La aplicación de MInCir-Pr2 por parte del usuario requiere una evaluación crítica de cada ítem y dominio para asignar la puntuación respectiva, cuya suma puede dar como resultado una puntuación total que oscila entre 7 y 60 puntos. El punto de corte para definir el constructo es de 33 puntos (dicotómico); permitiendo diferenciar entre CM adecuada o alta cuando el artículo evaluado recibe una puntuación ≥ 33 puntos; y CM inadecuada o baja si la puntuación es ≤ 32 puntos (Zavando, 2011; Manterola *et al.*, 2015, 2018).

A continuación, se proporciona una guía detallada para comprender y aplicar cada dominio e ítem de esta escala, permitiendo al usuario obtener una comprensión precisa para la aplicación correcta de MInCir-Pr2.

Dominio 1: Diseño de investigación.

El tipo de diseño del estudio debe aparecer en la sección "material y método" del artículo, y debe ser el mismo que se declara en el resumen y eventualmente en el título. Esto es lo que se debe evaluar, la descripción del tipo de diseño de investigación según las puntuaciones indicadas. De este modo, un estudio prospectivo o de cohorte concurrente y un EC con asignación aleatoria recibirán 15 puntos en este dominio; y en el otro extremo, una serie de casos solo recibirá 3 puntos. Lamentablemente, a menudo resulta difícil identificar claramente el diseño del estudio, especialmente cuando los autores no lo mencionan, o simplemente "inventan" el diseño, por ejemplo:

- a) Artículos en los que los autores no hacen una mención o descripción del diseño utilizado en la sección "material y método" ni en el resumen; por lo tanto, es necesario determinarlo a través de una lectura detenida de todo el artículo.
- b) En otros artículos, los autores hacen una breve descripción de algunas características del diseño, como: "estudio prospectivo randomizado" en lugar de denominarlo ensayo clínico con asignación aleatoria; o "estudio retrospectivo" en lugar de serie de casos retrospectiva. Esta situación no es ideal, pero al menos es aceptable para puntuar este ítem.
- c) En otras ocasiones, el diseño del estudio no es consistente con el reportado por los autores. En estas situaciones, lo que se observa con mayor frecuencia es una suerte de magnificación del tipo de diseño, por ejemplo, declarar que se trata de un estudio de cohorte, cuando en realidad es una serie de casos, incluso sin seguimiento (en ocasiones, este se trata de una simple serie de casos retrospectiva).

- d) Por último, algunos artículos mencionan un diseño que no existe, por ejemplo, "caso-control de cohorte" o "estudio prospectivo de cohorte no concurrente". En estos casos, la mayoría de las veces se trata de una serie de casos.

Es muy inusual que los autores desarrollen un diseño avanzado o complejo y no lo informen de forma adecuada en la metodología, y que los resultados no sean acordes a ello. Por lo tanto, cuando no se menciona el diseño utilizado o se menciona un diseño que no existe, generalmente se trata de una serie de casos (por lo general, retrospectiva). Por otro lado, si los autores informan un estudio de cohorte, los resultados del artículo deben contener un valor de incidencia, riesgo absoluto o relativo con sus respectivos intervalos de confianza del 95 %.

Dominio 2: Población en estudio.

En este dominio, la puntuación debe asignarse según el número de pacientes que participaron efectivamente en el estudio, es decir, sólo aquellos incluidos en el análisis de los resultados informados. Además, se debe considerar si este número de pacientes fue estimado con un procedimiento estadístico adecuado en la planificación del estudio, es decir, si se calculó o estimó el tamaño de la muestra necesario para conducir la investigación. Si se realizó, entonces se aplica el "factor de justificación", que puede duplicar la puntuación del dominio. Por ejemplo, un estudio con 560 pacientes tendría al menos 7 puntos, pero si el tamaño mínimo de la muestra para llevar a cabo el estudio fue estimado (cálculo de la muestra), entonces se asignarán 15 puntos. En el caso de un estudio con 58 pacientes, se asignarán 6 o 3 puntos dependiendo si se calculó o no el tamaño de la muestra. Reconocer el "factor de justificación" es muy importante, ya que en la mayor parte de los casos, los estudios se realizan incluyendo solo a los pacientes disponibles, sin considerar la estimación del tamaño mínimo de muestra para asegurar el poder estadístico de los resultados.

Dominio 3: Metodología.

Este dominio evalúa 5 aspectos de la metodología: objetivos, diseño, variables, tamaño de la muestra y seguimiento realizado.

Ítem 1. Objetivos: Por lo general, se encuentran al final de la introducción y en el resumen. En algunos casos, y dependiendo de las normas de la revista, podrían ubicarse al principio de la sección de metodología. Debe haber coherencia entre los objetivos declarados en el texto principal y el resumen. En artículos más cortos, como las series de casos, el objetivo podría mencionarse solo en el resumen. Siguiendo esta línea de argumentación, los objetivos pueden evaluarse en tres niveles:

1. **Objetivos claros, precisos y concretos:** En este caso, es posible reconocer claramente lo que se medirá, de qué manera, por quién, dónde y cómo se seguirá. Si la metodología se basa en este tipo de objetivo. Se asignan tres puntos.
2. **Objetivos vagos:** Cuando alguno de los elementos mencionados anteriormente está ausente o redactado de forma incorrecta. Es posible reconocer en esta categoría enunciados como: "El objetivo de este artículo es informar nuestra experiencia en ...", "... en comparación con la opción convencional ...", etc. En estos casos, no es posible reconocer claramente el propósito del estudio. Se asignan dos puntos.
3. **Sin objetivos:** Estos no son visibles en ninguno de los apartados mencionados anteriormente. Se asigna un punto.

Ítem 2. Diseño: Se evaluará en la sección de metodología, a veces llamada "Material y Métodos", "Métodos", o "Pacientes y Métodos". Al igual que el objetivo, el diseño debe mencionarse también en el resumen del artículo o incluso en el título. En el caso de un informe de caso o una serie de casos, el diseño podría mencionarse solo en el título o en el resumen. Las opciones posibles para este ítem son las siguientes:

1. **Identificación clara del diseño:** Es imperativo que el diseño reportado se ajuste al diseño real del estudio. También es aceptable la descripción de las características de un diseño, por ejemplo: "estudio prospectivo aleatorizado", en lugar de EC con asignación aleatoria; o "estudio longitudinal, descriptivo prospectivo", en lugar de serie prospectiva de casos. Se asignan tres puntos.
2. **Diseño desconocido:** No es posible identificar el diseño mencionado, o el diseño no se ajusta a lo que realmente se ejecuta. Por ejemplo, los autores informan que el diseño es un estudio de cohorte, cuando en realidad es una serie de casos. Se asigna un punto.

Ítem 3. Variables: En este ítem se evalúa la adecuada definición de las variables de resultado, exposición y confusión. Se asignará un punto por una definición completa de cada una de ellas.

1. **Variable de resultado o dependiente:** Es el evento de interés, identificado específicamente en su definición conceptual y operacional. Generalmente, corresponde a una enfermedad o condición, como: morbilidad postoperatoria, dolor postoperatorio, supervivencia libre de enfermedad a 5 años, etc.

2. **Variable de exposición o independiente:** Definidas como una o más variables incidentes en la ocurrencia del evento de interés. Al igual que en el caso anterior, la definición debe incluir la operacionalización de ésta. Por ejemplo, la existencia de comorbilidades (diabetes mellitus, dislipidemia, enfermedad cardíaca, etc.); que se analizó de manera dicotómica (presente o ausente) o categorizada por tipo de tratamiento requerido (fármacos/cirugía/ninguno).
3. **Variable de confusión:** La definición de aquellas variables que pueden alterar la asociación entre el evento de interés y la exposición; arriesgando una interpretación errónea del efecto real de estas. También deberá incluir una definición conceptual y operacionalización. Por ejemplo: La edad como variable continua o dicotomizada en menos o más de 45 años.

Ítem 4. Tamaño de la muestra: Deberá evaluar la justificación del número de sujetos incluidos en el estudio, existen las siguientes opciones:

1. Existe un cálculo o estimación clara del tamaño de la muestra; o se trabajó con el total de la población blanco. Se asignan tres puntos.
2. No hay cálculo ni estimación del tamaño de la muestra. No es posible encontrar ninguna referencia para entender la inclusión del número de pacientes reportado. Se asigna un punto.

Ítem 5. Seguimiento: Al igual que en el ítem 3 (variables); en este, se evalúan tres aspectos de forma independiente, asignando un punto a cada uno si están adecuadamente reportados.

1. **Se reporta el porcentaje de sujetos con seguimiento:** se menciona la proporción de pacientes que completaron el período de seguimiento previamente definido, o el porcentaje de pérdidas durante el seguimiento. Debe indicarse el seguimiento para la población general y para cada uno de los brazos, grupos o cohortes comparados en el estudio.
2. **Seguimiento superior al 80 %:** Este límite debe aplicarse no solo a la población total, sino a cada uno de los brazos, grupos o cohortes incluidos en el estudio. Por ejemplo, "la cohorte de pacientes diabéticos tuvo una pérdida del 11 %, y del 24 % para pacientes no diabéticos, al final del seguimiento"; en este caso se asigna cero, ya que una de las cohortes no alcanzó el mínimo del 80 %.
3. **Indicar las causas de la pérdida de seguimiento:** se describen claramente las razones de las pérdidas para cada

uno de los grupos o cohortes, así como las estrategias implementadas para su control. Por ejemplo: "... durante el seguimiento, 4 pacientes cambiaron de ciudad, 2 fallecieron por accidentes no relacionados con la patología en estudio y 3 rechazaron continuar participando; el protocolo contempló al menos una comunicación mensual con cada paciente para mantener el vínculo con el equipo de investigación y disminuir pérdidas", en este caso se asignará un punto.

4. Dominio 4: Análisis y conclusiones.

Este dominio evalúa 5 aspectos del análisis, los resultados y su interpretación. Estos se relacionan con el cálculo de medidas de riesgo y desarrollo de modelos predictivos o de modelos de asociación (o que se haya reportado información suficiente para realizar ambos cálculos); y conclusiones coherentes con el objetivo y la metodología empleada.

Ítem 1. Cálculo de medidas de riesgo: Se reportan las medidas de riesgo relevantes para el objetivo declarado y el tipo de diseño. Se asignan 5 puntos. Por ejemplo: en un estudio de casos y controles o en un estudio de corte transversal, deberían mencionarse las odds ratio (OR); para un estudio de cohortes y en un EC, deberían reportarse los riesgos absolutos (RA) y el riesgo relativo (RR); así como otras medidas complementarias como la reducción absoluta del riesgo (RAR) y el número necesario para tratar (NNT) o a dañar (NND), con sus respectivos intervalos de confianza del 95 %. Todas estas medidas son crudas o no ajustadas. Si no se informan, el ítem se evalúa con cero punto.

Ítem 2. Información para el cálculo de medidas de riesgo: Se dispone de información suficiente para que el lector pueda calcular y confirmar el valor de las medidas de riesgo que se informan u otras relevantes. Se asignan dos puntos. En general, esta información debería presentarse en tablas comparando cada uno de los grupos de estudio o cohortes según la variable de resultado. Si no se informa, el ítem se evalúa con cero punto.

Ítem 3. Modelos predictivos o de asociación: Deben informarse los modelos utilizados para evaluar el impacto general de cada exposición y variables de confusión sobre el resultado. Se asignan 5 puntos. Los modelos utilizados pueden ser de regresión logística o lineal, y curvas de supervivencia. Estos modelos deben permitir obtener medidas de riesgo ajustadas y sus respectivos intervalos de confianza. Si no se informan, el ítem se evalúa con cero punto.

Ítem 4. Consistencia objetivo-metodología-resultados: Los resultados informados responden al objetivo y la

metodología. Se asignan tres puntos. Por ejemplo: En el caso de estudios que se centran en enfermedades raras o a largo plazo, el diseño apropiado sería un estudio de casos y controles, aplicando cálculos de odds ratios (OR) como medida de riesgo. Deben concluir sobre el impacto de las variables de exposición en el evento de interés; en este caso hay consistencia, y se asignan tres puntos. En otros casos, hay una metodología incorrecta para el objetivo declarado, un seguimiento insuficiente de una cohorte o la inclusión de pacientes que no se ajustan al objetivo. En estos casos, se asignan cero punto.

También es posible observar estudios en los que la conclusión no es consistente con el diseño. Esto se ve con frecuencia en series de casos que plantean conclusiones absolutas e incluso recomendaciones. En estos casos, se asignan cero punto.

DISCUSIÓN

En el caso específico de constructos medidos con escalas, es esencial verificar la validez y confiabilidad mediante el uso de diversas estrategias (Rojahn *et al.*, 2011; Aravena *et al.*, 2014). Sin embargo, hay otros factores relevantes, como el modo de aplicación y el uso del instrumento por parte del evaluador. A lo largo del ejercicio de diseño, construcción, validación y aplicación de escalas para medir la CM por parte del grupo MInCir, surge la clara necesidad de explicar detalladamente la forma de puntuar cada ítem del instrumento (Moraga *et al.*, 2014a). A pesar de que los niveles de puntuación para cada ítem parecen ser auto explicativos, en la práctica, la evaluación de los artículos se vuelve difícil porque la mayoría de los artículos tienen una baja calidad de informe y CM, a menudo con redacción o estructura confusa y un uso indebido de la terminología metodológica.

Dada la situación anterior, es necesario contar con instrucciones como estas. Estas instrucciones especifican de manera más detallada la aplicación de la escala, incluyendo ejemplos que faciliten la identificación de situaciones específicas pero frecuentes. De acuerdo con lo anterior, estas instrucciones facilitan la calibración de múltiples evaluadores (Aravena *et al.*, 2014), para llevar a cabo estudios bibliométricos más extensos (Manterola *et al.*, 2006) o RS (Moraga *et al.*, 2014b).

En el caso de la escala MInCir-T, el uso de instrucciones sobre la aplicación del instrumento (Moraga *et al.*, 2014b) permitió obtener medidas confiables cercanas al 100 % en los campos quirúrgico (Moraga *et al.*, 2013) y dental (Cartes-Velasquez *et al.*, 2014), demostrando la utilidad de esta guía.

Hay otra herramienta para evaluar la CM en estudios de pronóstico, la Quality in Prognosis Studies (QUIPS), que evalúa la calidad de los estudios de pronóstico en revisiones sistemáticas. Sin embargo, el enfoque que utiliza esta herramienta es diferente, ya que se centra en identificar sesgos de investigación (Hayden *et al.*, 2013).

MANTEROLA, C.; HOLGUÍN-CARVAJAL, J.P.; RIVADENEIRA, J. & OTZEN, T. MInCir-Pr2: A scale for determining the methodological quality of primary prognosis articles. Explanation of its domains and items, and a guideline for its application. *Int. J. Morphol.*, 42(5):1394-1402, 2024.

SUMMARY: The assessment of methodological quality (MQ) in biomedical research is an area that has undergone dynamic development in the last 20 years on a global scale. MQ can be defined as a multidimensional and complex construct composed of items and domains. Different research groups have proposed tools for its assessment, among which our team is included with the development of the MInCir scales. Over the years, we have developed and validated scales to assess MQ in treatment or therapeutic procedure studies (MInCir-T), in diagnostic studies (MInCir-Dg), and prognosis studies (MInCir-Pr in its two versions, MInCir-Pr1 and MInCir-Pr2). However, as is the case with any measurement instrument, it is crucial to specify in a detailed and precise manner how it is applied, aiming to standardize measurements and minimize information biases. A detailed description is provided for the 4 domains and 8 items comprising the MInCir-Pr2 scale, specifying for each of them the characteristics to be assessed and scored in primary prognosis articles. The scale total score ranges from 7 to 60 points, with a cutoff point of 33 points (dichotomous) to define the MQ construct. This cutoff point allows classification into either adequate or high MQ (when the evaluated article receives a score of 33 points or higher), and inadequate or low MQ if the score is 32 points or less. The aim of this article was to elucidate and clarify the application of the MInCir-Pr2 scale for MQ in prognosis studies, through a guide intended to standardize MQ measurements using the MInCir-Pr2 scale, with the goal of enhance both intra- and inter-observer reliability.

KEY WORDS: Scales; Weights and Measures; Reproducibility of Results; Validity of results; Methodological quality; Prognosis.

REFERENCIAS BIBLIOGRÁFICAS

- Aravena, P. C.; Moraga, J.; Cartes-Velásquez, R. & Manterola, C. Validity and reliability in dental research. *Int. J. Odontostomat.*, 8(1):69-75, 2014.
- Aravena, P.; Cartes-Velásquez, R. & Manterola, C. Productividad y calidad metodológica de artículos clínicos en cirugía oral y maxilofacial en Chile: Período 2001-2012. *Rev. Chil. Cir.*, 65(5):382-8, 2013.
- Armijo-Olivo, S.; Patrini, M.; Oliveira-Souza, A. I. S.; Dennett, L.; Arienti, C.; Dahchi, M. & Negrini, S. Tools to assess the risk of bias and reporting quality of randomized controlled trials in rehabilitation. *Arch. Phys. Med. Rehabil.*, 102(8):1606-13, 2021.
- Burgos, D. M. E.; Manterola, D. C. & Sanhueza, C. A. Diseño de una escala para evaluar calidad metodológica de estudios de pruebas diagnósticas. Estudio piloto. *Rev. Chil. Cir.*, 63(5):493-7, 2011.
- Cartes-Velásquez, R.; Manterola, C.; Aravena, P. & Moraga, J. Reliability and validity of MINCIR scale for methodological quality in dental therapy research. *Braz. Oral Res. (São Paulo)*, 28(4):1-5, 2014.
- Carvajal, A.; Centeno, C.; Watson, R.; Martínez, M. & Sanz Rubiales, Á. ¿Cómo validar un instrumento de medida de la salud? *An. Sist. Sanit. Navarra*, 34(1):63-72, 2011.
- Castillo, S.; Manterola, C.; Grande, L. & Rojas, C. Infected hepatic echinococcosis. Clinical, therapeutic, and prognostic aspects. A systematic review. *Ann. Hepatol.*, 22:100237, 2021.
- da Silva, F. C.; Valdivia Arancibia, B. A.; da Rosa Iop, R.; Barbosa Gutierrez Filho, P. J. & da Silva, R. Escalas y listas de evaluación de la calidad de estudios científicos. *Rev. Cuba. Inf. Cienc. Salud*, 24(3):295-312, 2013.
- Duque, G.; Manterola, C.; Otzen, T.; Arias, C.; Galindo, B.; Mora, M.; Guerrero, E. & García, N. Clinical utility of liquid biopsy in breast cancer: A systematic review. *Clin. Genet.*, 101(3):285-95, 2022a.
- Duque, G.; Manterola, C.; Otzen, T.; Arias, C.; Palacios, D.; Mora, M.; Galindo, B.; Holguín, J. P. & Albarracín, L. Cancer Biomarkers in Liquid Biopsy for Early Detection of Breast Cancer: A Systematic Review. *Clin. Med. Insights Oncol.*, 16:11795549221134831, 2022b.
- Keszei, A. P.; Novak, M. & Streiner, D. L. Introduction to health measurement scales. *J. Psychosom. Res.*, 68(4):319-23, 2010.
- Kiani, A. K.; Naureen, Z.; Pheby, D.; Henahan, G.; Brown, R.; Sieving, P.; Sykora, P.; Marks, R.; Falsini, B.; Capodicasa, N.; et al. Methodology for clinical research. *J. Prev. Med. Hyg.*, 63(2 Suppl. 3):E267-E278, 2022.
- Losada, H.; Manterola, C.; Pineda, V.; Vial, M. & Sanhueza A. Diseño de una escala para la evaluación de calidad metodológica de estudios de pronóstico. *Rev. Chil. Cir.*, 61(1):59-72, 2009.
- Luján-Tangarife, J. A. & Cardona-Arias, J. A. Construcción y validación de escalas de medición en salud: revisión de propiedades psicométricas. *Arch. Med.*, 11(3):1-10, 2015.
- Ma, L. L.; Wang, Y. Y.; Yang, Z. H.; Huang, D.; Weng, H. & Zeng, X. T. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Mil. Med. Res.*, 7:7, 2020.
- Malic'ki, M.; Aalbersberg, I. J.; Bouter, L. & Ter Riet, G. Journals' instructions to authors: A cross-sectional study across scientific disciplines. *PLoS One*, 14(9):e0222157, 2019.
- Manterola, C.; Cartes-Velásquez, R. & Otzen, T. Instructions for using the MInCir scale to assess methodological quality in diagnostic accuracy studies. *Int. J. Morphol.*, 34(1):78-84, 2016.
- Manterola, C.; Cartes-Velásquez, R. & Otzen, T. Instructions for the use of MInCir Scale to assess methodological quality in prognosis studies. *Int. J. Morphol.*, 33(4):1553-8, 2015.
- Manterola, C.; Pineda, V.; Vial, M.; Losada, H. & MINCIR Group. What is the methodologic quality of human therapy studies in ISI surgical publications? *Ann. Surg.*, 244(5):827-32, 2006.
- Manterola, C.; Vial, M.; Pineda, V.; Sanhueza, A. Systematic review of literature with different types of designs. *Int. J. Morphol.*, 27(4):1179-86, 2009.
- Manterola, C.; Zavando, D.; Cartes-Velásquez, R. & Otzen, T. & Sanhueza, A. Initial validation of a scale to measure methodological quality in prognosis studies. The MInCir proposal. *Int. J. Morphol.*, 36(2):762-7, 2018.
- Moraga, J.; Manterola, C.; Cartes-Velásquez, R. Single-incision laparoscopic appendectomy versus conventional laparoscopy in adults. A systematic review. *Acta Cir. Bras.*, 29(12):826-31, 2014b.
- Moraga, J.; Manterola, C.; Cartes-Velásquez, R.; Burgos, M. E.; Aravena, P. & Urrutia, S. Instructions for the use of MINCIR scale to assess methodological quality in therapy studies. *Int. J. Morphol.*, 32(1):294-8, 2014a.
- Moraga, J.; Manterola, C.; Sanhueza, A.; Cartes-Velásquez, R. & Urrutia, S. Confiabilidad de la escala MInCir para valorar calidad metodológica de estudios de terapia. *Rev. Chil. Cir.*, 65(3):222-7, 2013.
- Muchadeyi, M. T.; Hernandez-Villafuerte, K. & Schlander, M. Quality appraisal for systematic literature reviews of health state utility values: a descriptive analysis. *BMC Med. Res. Methodol.*, 22(1):303, 2022.

- Prill, R.; Królikowska, A.; de Girolamo, L.; Becker, R. & Karlsson, J. Checklists, risk of bias tools, and reporting guidelines for research in orthopedics, sports medicine, and rehabilitation. *Knee Surg. Sports Traumatol. Arthrosc.*, 31(8):3029-33, 2023.
- Rojahn, J.; Rowe, E. W.; Kasdan, S.; Moore, L. & van Ingen, D. J. Psychometric properties of the Aberrant Behavior Checklist, the Anxiety, Depression and Mood Scale, the Assessment of Dual Diagnosis and the Social Performance Survey Schedule in adults with intellectual disabilities. *Res. Dev. Disab.*, 32(6):2309-20, 2011.
- Sargeant, J. M.; Brennan, M. L. & O'Connor, A. M. Levels of evidence, quality assessment, and risk of bias: evaluating the internal validity of primary research. *Front. Vet. Sci.*, 9:960957, 2022.
- Stang, A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur. J. Epidemiol.*, 25(9):603-5, 2010.
- Stang, A.; Jonas, S. & Poole, C. Case study in major quotation errors: a critical commentary on the Newcastle-Ottawa scale. *Eur. J. Epidemiol.*, 33(11):1025-31, 2018.
- Streiner, D. & Geoffrey, R. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 3rd ed. Oxford, Oxford University Press, 2003.
- The Centre for Evidence-Based Medicine. *Levels of Evidence* (March 2009). Oxford, Centre for Evidence-Based Medicine, 2009. Available from: <http://www.cebm.net/index.aspx?o=1025>
- Zavando, D. *Diseño y Validación de una Escala para Medir Calidad Metodológica en Estudios de Pronóstico*. Tesis. Temuco, Universidad de La Frontera, 2011.

Dirección para correspondencia:

Dr. Carlos Manterola
CEMyQ
Universidad de La Frontera
CHILE

E-mail: carlos.manterola@ufrontera.cl

Dirección para correspondencia:

Dr. Josue Rivadeneira
Programa de Doctorado en Ciencias Médicas
Universidad de La Frontera

E-mail: j.rivadeneira01@ufromail.cl