

Chapter 15 / Capítulo 15

Critical reading of scientific articles on
health: A practical guide (Spanish Edition)
ISBN: 978-9915-9928-3-9
DOI: 10.62486/978-9915-9928-3-9.Ch15
Pages: 243-262



© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

Quasi-Experimental Studies

Estudios cuasiexperimentales

Jhossmar Cristians Auza-Santivañez  

Pamela Gutierrez Villegas  

Aaron Eduardo Carvajal-Tapia  

1 Hospital del Gran Chaco “Fray Quebracho” de III nivel. Instituto Académico Científico Quispe-Cornejo. Tarija, Bolivia

2 Hospital Obrero Número 30. La Paz, Bolivia.

3 Hospital de la Mujer “Dr. Percy Boland Rodríguez”, Unidad de Cuidados Intensivos Neonatales. Santa Cruz, Bolivia.

ABSTRACT

Introduction. Quasi-experimental studies constitute an essential methodological alternative for evaluating clinical interventions when randomization is not feasible for ethical, logistical, or administrative reasons. Their application allows for the generation of evidence in real-world healthcare settings, although with greater susceptibility to selection bias and confounding. **Development.** This chapter describes the main quasi-experimental designs, their strengths, limitations, and strategies for controlling threats to internal validity, including multivariable adjustments, propensity score matching, and sensitivity analysis. It also addresses the critical interpretation of results, the evaluation of methodological quality, and the use of international tools for assessing the risk of bias and clinical applicability. **Conclusions.** Quasi-experimental studies represent a valuable source of evidence for decision-making when randomized controlled trials are not feasible. Their proper interpretation requires a critical analysis of the design, potential biases, and adjustment methods employed, promoting a reflective and contextualized application of their findings in clinical practice and public health.

Keywords: Quasi-experimental Studies

RESUMEN

Introducción. Los estudios cuasiexperimentales constituyen una alternativa metodológica esencial para evaluar intervenciones clínicas cuando la aleatorización no es factible por razones éticas, logísticas o administrativas. Su aplicación permite generar evidencia en escenarios asistenciales reales, aunque con mayor susceptibilidad a sesgos de selección y confusión. **Desarrollo.** El capítulo describe los principales diseños cuasiexperimentales, sus fortalezas, limitaciones y estrategias para controlar amenazas a la validez interna, incluyendo ajustes multivariados, puntuación de propensión y análisis de sensibilidad. Asimismo, aborda la interpretación crítica de los resultados, la evaluación de la calidad metodológica y el uso de herramientas internacionales para valorar el riesgo de sesgo y la aplicabilidad clínica. **Conclusiones:** Los estudios cuasiexperimentales representan una fuente valiosa de evidencia para la toma de decisiones cuando los ensayos clínicos aleatorizados no son viables. Su adecuada interpretación exige un análisis crítico del diseño, los sesgos potenciales y los métodos de ajuste empleados, favoreciendo una aplicación reflexiva y contextualizada de sus hallazgos en la práctica clínica y la salud pública.

Palabras clave. Estudios cuasiexperimentales

1. CASO CLÍNICO

En un Hospital de tercer nivel ubicado a 3600 metros sobre el nivel del mar, se implementó un Programa Intensivo de Rehabilitación Neuromotora (PIRN) dirigido a pacientes con accidente cerebrovascular (ACV) isquémico reciente, como complemento a la estrategia convencional de rehabilitación. Debido a restricciones operativas relacionadas con la disponibilidad de recursos humanos especializados y capacidad física instalada, la asignación de los pacientes a cada modalidad terapéutica no se realiza mediante un proceso de aleatorización, sino en función de la disponibilidad diaria de cupos y del orden de ingreso al servicio. Esta circunstancia imposibilita la realización de un ensayo clínico aleatorizado, pero configura un escenario idóneo para la aplicación de un diseño cuasiexperimental con grupos no equivalentes, representativo de las condiciones habituales de la práctica clínica. Los participantes son incorporados al programa durante las dos primeras semanas posteriores al evento cerebrovascular y reciben protocolos terapéuticos estandarizados. El grupo de intervención es sometido al PIRN, consistente en tres horas diarias de terapia intensiva durante un período de ocho semanas, mientras que el grupo comparador recibe rehabilitación convencional con una duración de una hora diaria. Ambos grupos continúan recibiendo el mismo tratamiento médico estándar, de acuerdo con las recomendaciones clínicas vigentes para el manejo del ACV isquémico.

Los desenlaces de interés incluyen variables clínicas y funcionales relevantes. Entre los desenlaces principales se encuentran la recuperación motora, evaluada mediante la escala de Fugl-Meyer, y la independencia funcional, medida a través del Índice de Barthel. Como desenlaces secundarios se consideran el grado global de discapacidad, valorado mediante la Escala Modificada de Rankin (mRS), las pruebas de movilidad funcional y la frecuencia de reingresos hospitalarios. Estas variables son evaluadas de manera sistemática en tres momentos predefinidos: al inicio del estudio, a los tres meses y a los seis meses de seguimiento. Aunque la ausencia de asignación aleatoria incrementa el riesgo de sesgo de selección y de confusión, los investigadores pueden minimizar estas amenazas mediante la comparación detallada de las características basales de los grupos y la aplicación de métodos estadísticos de ajuste, tales como modelos multivariantes o técnicas basadas en puntuaciones de propensión. Estas estrategias permiten obtener estimaciones más precisas y válidas del efecto de la intervención. Este escenario ilustra cómo los estudios cuasiexperimentales pueden generar evidencia clínicamente relevante y metodológicamente sólida cuando las limitaciones éticas, logísticas o administrativas impiden la realización de un ensayo clínico aleatorizado, aportando información útil para la toma de decisiones en contextos asistenciales reales.

2. OBJETIVOS

Evaluar la efectividad y seguridad de una intervención en escenarios clínicos reales, cuando la aleatorización no es posible o no es ética. A diferencia de los ensayos clínicos controlados, estos diseños permiten generar evidencia en condiciones que reflejan mejor la práctica cotidiana.

De manera más específica, los estudios cuasiexperimentales buscan:

- Estimar el efecto de la intervención sobre desenlaces clínicos relevantes, ajustando las diferencias iniciales entre los grupos para reducir sesgos.
- Comparar la evolución de los pacientes en el tiempo, identificando si los cambios observados pueden atribuirse a la intervención.
- Explorar la variabilidad de los resultados en subgrupos clave, como edad, severidad de la enfermedad o momento de inicio del tratamiento.
- Describir la seguridad y la adherencia, registrando eventos adversos y el grado en que los pacientes completan la intervención.
- Generar evidencia preliminar que sirva como base para ensayos clínicos aleatorizados futuros o para orientar decisiones en políticas de salud.

En conjunto, estos objetivos muestran cómo los ensayos cuasiexperimentales, aunque más vulnerables a sesgos que los aleatorizados, son herramientas valiosas para responder preguntas relevantes en contextos reales de práctica clínica y salud pública.

3. INTRODUCCIÓN

Los estudios cuasiexperimentales constituyen una alternativa metodológica fundamental dentro de la investigación clínica y la evaluación de intervenciones sanitarias. Estos diseños

ocupan una posición intermedia entre los estudios observacionales y los ensayos clínicos aleatorizados, ya que incorporan una intervención activa por parte del investigador, pero carecen de asignación aleatoria de los participantes a los grupos de estudio (1,2). Aunque presentan una menor capacidad para controlar factores de confusión en comparación con los ensayos clínicos aleatorizados, ofrecen importantes ventajas cuando la aleatorización no es posible, no resulta ética o es inviable debido a limitaciones logísticas, administrativas o de recursos (1,3). En la práctica clínica, la salud pública y la gestión de servicios sanitarios son frecuentes los escenarios en los que deben adoptarse decisiones basadas en evidencia sin disponer de ensayos clínicos aleatorizados. La evaluación de programas de rehabilitación, estrategias de vacunación, intervenciones educativas, modificaciones organizacionales o políticas de control del tabaquismo constituye ejemplos representativos de situaciones en las que los estudios cuasiexperimentales permiten valorar el impacto de una intervención en condiciones reales de implementación (3,4). En este contexto, estos diseños proporcionan evidencia relevante para orientar la toma de decisiones clínicas, epidemiológicas y de salud pública. La principal fortaleza de los estudios cuasiexperimentales radica en la combinación de dos características esenciales: la introducción deliberada de una intervención sobre una población determinada y la ausencia de asignación aleatoria de los participantes a los grupos de comparación (1,2). Esta última característica la diferencia de los ensayos clínicos tradicionales y constituye la principal fuente de vulnerabilidad metodológica, debido al incremento del riesgo de sesgo de selección, confusión y otras amenazas a la validez interna (5,6). Sin embargo, precisamente por desarrollarse en entornos asistenciales reales y bajo las condiciones habituales de atención sanitaria, estos estudios suelen presentar una elevada validez externa y una mayor aplicabilidad de sus resultados. Por tanto, permiten evaluar la efectividad de intervenciones complejas en poblaciones heterogéneas y en escenarios que reflejan de manera más fiel la realidad cotidiana de los sistemas de salud (7,8).

Desde la perspectiva de la Medicina Basada en la Evidencia, la adecuada interpretación de los estudios cuasiexperimentales exige una evaluación crítica rigurosa de su diseño, ejecución, análisis e interpretación. El lector debe ser capaz de identificar posibles fuentes de sesgo, valorar las estrategias empleadas para controlar factores de confusión y juzgar la validez, precisión y relevancia clínica de los resultados obtenidos (1,5). El objetivo de este capítulo es proporcionar al lector los fundamentos conceptuales y metodológicos necesarios para reconocer, analizar e interpretar críticamente los estudios cuasiexperimentales. De esta manera, se pretende desarrollar competencias que permitan evaluar la calidad de la evidencia generada por estos diseños, comprender sus fortalezas y limitaciones, y determinar la aplicabilidad de sus hallazgos en la práctica clínica, la salud pública y la toma de decisiones sanitarias.

4. METODOLOGÍA

Los estudios cuasiexperimentales se caracterizan porque el investigador interviene activamente (introduce un tratamiento, programa o política), pero la asignación de los participantes a los grupos no es aleatoria. Esta característica la diferencia de los ensayos clínicos aleatorizados y explica por qué son más vulnerables a sesgos, en especial el de selección.

Existen diferentes diseños cuasiexperimentales, cada uno con utilidades y limitaciones particulares.

4.1 ESTUDIO CON GRUPOS NO EQUIVALENTES

En este diseño se comparan dos o más grupos de pacientes, pero la asignación no es aleatoria, sino que depende de factores externos. Eso significa que los grupos pueden empezar siendo diferentes en aspectos que influyen directamente en los resultados, lo que se conoce como sesgo de selección. Ejemplo aplicado al caso clínico: algunos pacientes ingresan al Programa Intensivo de Rehabilitación Neuromotora (PIRN) y otros a la rehabilitación convencional, no por decisión clínica o por azar, sino por la disponibilidad de cupos. Esto implica que, al inicio, un grupo puede incluir más pacientes jóvenes, con menor gravedad del ACV o con más apoyo familiar, mientras que el otro concentre pacientes más graves o con más comorbilidades.

Si estas diferencias no se reconocen ni se corrigen, el estudio puede atribuir al programa intensivo un beneficio que en realidad se debe a que sus pacientes estaban “mejor de entrada”.

Las claves para la lectura crítica son:

- Verificar si los autores describen con detalle las características basales de ambos grupos (edad,

sexo, gravedad del ACV, comorbilidades, función previa).

- Evaluar si se usaron ajustes estadísticos (como modelos de regresión o técnicas de emparejamiento por puntuación de propensión) para equilibrar las diferencias iniciales.
- Preguntarse si, aun con los ajustes, pueden existir variables no medidas (p. ej., motivación del paciente, calidad del apoyo familiar) que sigan influyendo en los resultados.

En síntesis, este diseño puede ofrecer información valiosa, pero la validez de sus conclusiones depende de la transparencia en la comparación inicial y de la solidez de los métodos utilizados para reducir las diferencias entre los grupos.

4.2 ESTUDIOS ANTES Y DESPUÉS (PRE-POST)

En este diseño se observa a un mismo grupo de pacientes en dos momentos: antes de recibir la intervención y después de completarla. El análisis se centra en identificar si hubo cambios en los desenlaces tras la exposición al programa o tratamiento. Ejemplo aplicado al caso clínico: los pacientes que ingresan al Programa Intensivo de Rehabilitación Neuromotora (PIRN) se evalúan al inicio y luego nuevamente a los seis meses. Si la mayoría muestra mejoría en la independencia funcional, podría pensarse que el programa es efectivo.

Sin embargo, la principal debilidad de este diseño es que no cuenta con un grupo comparador. La mejoría observada podría deberse no solo al PIRN, sino también a otros factores: la recuperación natural del cerebro tras un ACV, el tratamiento médico estándar, el apoyo familiar o incluso cambios espontáneos en la motivación de los pacientes. Por lo tanto, aunque los resultados sean llamativos, no siempre se puede atribuir el efecto exclusivamente a la intervención.

Las claves de lectura crítica son:

- Verificar si el estudio reconoce esta limitación y discute la posibilidad de factores externos que expliquen el cambio.
- Preguntar si se incluyó un grupo comparador (aunque sea retrospectivo o histórico) para contextualizar los resultados.
- Analizar si se utilizaron varios puntos de medición en el tiempo, en lugar de solo dos, lo cual fortalece la interpretación al mostrar tendencias.

En resumen, los estudios antes y después son útiles como primera aproximación, porque permiten detectar cambios asociados a una intervención de forma rápida y económica. No obstante, sus hallazgos deben interpretarse con cautela, ya que la ausencia de un grupo control debilita la atribución causal directa.

4.3 ESTUDIO DE SERIES TEMPORALES INTERRUPTIDAS

Este diseño analiza la evolución de una variable en múltiples puntos de tiempo antes y después de aplicar una intervención. Su fortaleza es que permite diferenciar los cambios espontáneos de la tendencia de aquellos que aparecen de forma abrupta tras la introducción de la medida.

Ejemplo aplicado al caso clínico: el hospital decide implementar el Programa Intensivo de Rehabilitación Neuromotora (PIRN) en enero de 2024. Para evaluar su impacto, los investigadores revisan de manera mensual, durante todo un año, el porcentaje de pacientes con buena recuperación funcional. Si la tendencia se mantenía estable en los meses previos y se observa un cambio brusco después de la introducción del programa, se puede atribuir con mayor confianza la mejoría al PIRN.

Su principal riesgo radica en que no todos los cambios observados en la tendencia se deben a la intervención. Otros factores externos, como la contratación de más fisioterapeutas, la actualización de protocolos médicos o una campaña comunitaria, pueden coincidir con la implementación y explicar la mejoría.

Las claves de lectura crítica son:

- Verificar si se recolectaron suficientes puntos de tiempo antes y después de la intervención; contar solo con uno o dos registros debilita la evidencia.
- Revisar si los autores describieron posibles factores externos que pudieron influir en la tendencia.
- Observar si se analizaron las series con métodos estadísticos apropiados (p. ej., regresión segmentada) que permitan separar cambios abruptos de variaciones graduales.

En síntesis, las series temporales interrumpidas ofrecen un nivel de evidencia mayor que

los estudios antes y después, porque incorporan la dimensión del tiempo y permiten observar patrones más sólidos. No obstante, su interpretación requiere un análisis cuidadoso de los posibles factores concurrentes.

4.4 DISEÑO CON CONTROLES HISTÓRICOS

En este tipo de estudio, los pacientes que reciben la nueva intervención se comparan con un grupo de referencia que fue tratado en el pasado, antes de que la intervención estuviera disponible. Es decir, el control no es concurrente, sino histórico.

Ejemplo aplicado al caso clínico: los pacientes que participaron en el Programa Intensivo de Rehabilitación Neuromotora (PIRN) en 2024 se comparan con pacientes tratados únicamente con rehabilitación convencional en 2022, antes de que existiera el PIRN en el hospital.

Su principal riesgo es que los grupos pueden diferir no solo por la intervención, sino también por los cambios en la práctica clínica a lo largo del tiempo. Por ejemplo, en 2024 puede haber más acceso a fármacos, nuevas guías de manejo, mayor capacitación del personal o incluso diferencias en la población que llega al hospital. Todo esto puede influir en los resultados y sesgar la comparación.

Las claves de lectura crítica son:

- Verificar si los autores describieron con detalle las diferencias contextuales entre el periodo histórico y el actual.
- Revisar si se aplicaron ajustes estadísticos para reducir las diferencias entre las cohortes de distintos años.
- Preguntarse si los cambios observados pueden explicarse por mejoras generales del sistema de salud y no necesariamente por la intervención evaluada.

En síntesis, los diseños con controles históricos son una alternativa útil cuando no se dispone de un grupo control concurrente, pero sus resultados deben interpretarse con precaución, porque los cambios propios del tiempo pueden distorsionar la atribución causal.

4.5. DISEÑO CON TÉCNICAS ESTADÍSTICAS DE EMPAREJAMIENTO (PROPENSITY SCORE)

Una de las principales debilidades de los estudios cuasiexperimentales es que los grupos suelen empezar siendo distintos. Para reducir esa diferencia se utiliza el propensity score o “puntuación de propensión”. Esta técnica estima la probabilidad de que un paciente reciba la intervención en función de sus características basales (edad, sexo, comorbilidades, severidad de la enfermedad, entre otras). Luego, se pueden emparejar pacientes con puntajes similares de ambos grupos o aplicar ponderaciones estadísticas que equilibran las diferencias. Ejemplo aplicado al caso clínico: supongamos que los pacientes del PIRN eran, en promedio, más jóvenes y con menor severidad del ACV que los de la rehabilitación convencional. Para evitar que estas diferencias distorsionen los resultados, los investigadores calculan la probabilidad de entrar al PIRN según edad, severidad y comorbilidades. Posteriormente, comparan solo a los pacientes que tienen una probabilidad similar de haber sido asignados a cada programa.

El riesgo radica en que, aunque el método es útil, solo puede corregir las diferencias en variables que fueron medidas. Factores no registrados como la motivación del paciente, la calidad del apoyo familiar o la distancia al hospital pueden seguir introduciendo sesgo.

Las claves de lectura crítica son:

- Revisar qué variables se incluyeron en el cálculo de la puntuación y si son clínicamente relevantes.
- Observar si los autores muestran un balance adecuado entre los grupos después del emparejamiento (tablas o gráficos de distribución de covariables).
- Preguntarse si persisten variables importantes no medidas que podrían afectar los resultados.

El uso de técnicas de emparejamiento mediante propensity score hace que los estudios cuasiexperimentales sean más comparables y robustos, aunque nunca alcanzan la solidez de un ensayo aleatorizado.

4.6 DISEÑOS DE REGRESIÓN CONTINUA

Este diseño se utiliza cuando la asignación a la intervención depende de un punto de corte en una variable continua. Permite comparar pacientes muy similares, situados justo por encima

o por debajo del umbral, lo que genera un escenario cercano a la aleatorización. Ejemplo aplicado al caso clínico: el hospital establece que los pacientes con un puntaje de NIHSS ≥ 10 ingresen automáticamente al Programa Intensivo de Rehabilitación Neuromotora (PIRN), mientras que aquellos con puntajes < 10 reciben rehabilitación convencional. En este caso, los pacientes con puntajes de 9 y 10 son prácticamente comparables en severidad, pero fueron asignados a distintos programas por la regla del punto de corte.

El riesgo de este diseño radica en que solo es aplicable cuando existe un criterio de elegibilidad claro y estricto, y sus conclusiones se limitan a los pacientes cercanos al umbral, lo que restringe la generalización.

Las claves de lectura crítica son:

- Confirmar que el umbral de asignación fue respetado y aplicado de forma consistente.
- Evaluar si los pacientes cercanos al punto de corte realmente son comparables.
- Preguntarse si los resultados pueden extrapolarse más allá de los pacientes próximos al umbral.

4.7 DISEÑOS DE ASIGNACIÓN POR ALTERNANCIA O CUOTAS

En lugar de un proceso aleatorio, los pacientes se asignan siguiendo una regla preestablecida o secuencial, como días de la semana, turnos o cuotas de ingreso.

Ejemplo aplicado al caso clínico: los pacientes que ingresan lunes, miércoles y viernes se derivan a PIRN, mientras que los de martes y jueves a rehabilitación convencional. Aunque parece un método ordenado, no es realmente aleatorio, y existe el riesgo de que algunos pacientes elijan acudir en determinado día, introduciendo sesgo.

Su riesgo radica en que, si la asignación sigue un patrón predecible, pueden generarse diferencias sistemáticas entre los grupos, por ejemplo, que pacientes más motivados se presenten en días específicos.

Las claves de lectura crítica son:

- Verificar si la asignación se realizó estrictamente según la regla y sin manipulación por parte de los investigadores o pacientes.
- Analizar si existe algún patrón oculto que pueda explicar diferencias entre los grupos.
- Considerar que este método es más frágil frente a sesgos que la aleatorización verdadera.

Tabla 1. Diseños cuasiexperimentales más utilizados: características, riesgos y claves de lectura crítica

Diseño	Ejemplo aplicado al caso clínico (PIRN vs rehabilitación convencional)	Riesgo principal	Claves de lectura crítica
Grupos no equivalentes	Comparar pacientes asignados a PIRN frente a rehabilitación convencional según cupos disponibles.	Diferencias iniciales entre grupos (edad, severidad, comorbilidades).	Revisar características basales; comprobar si se aplicaron ajustes estadísticos adecuados.
Antes y después (pre-post)	Evaluar independencia funcional en un mismo grupo de pacientes antes y seis meses después de PIRN.	Otros factores (recuperación natural, tratamiento médico estándar) pueden explicar el cambio.	Preguntar si hubo grupo comparador; valorar si se usaron varios puntos de medición.
Series temporales interrumpidas	Medir mensualmente la recuperación funcional antes y después de implementar PIRN en el hospital.	Factores externos coincidentes (contratación de personal, cambios de protocolo).	Verificar número suficiente de mediciones; comprobar si se aplicaron análisis de tendencia apropiados.
Controles históricos	Comparar pacientes tratados con PIRN en 2024 frente a pacientes con rehabilitación convencional en 2022.	Cambios en la práctica clínica y en la población a lo largo del tiempo.	Evaluar si los autores describen diferencias contextuales; revisar si hubo ajustes estadísticos.
Propensity Score (emparejamiento/ponderación)	Emparejar pacientes de PIRN y de rehabilitación convencional con edad, severidad y comorbilidades similares.	Solo ajusta variables medidas; factores no registrados pueden sesgar.	Revisar variables incluidas en el score; comprobar balance entre grupos tras el emparejamiento.
Regresión discontinua	Asignar PIRN a pacientes con NIHSS ≥ 10 y rehabilitación convencional a quienes tienen < 10 .	Conclusiones limitadas a los pacientes cercanos al punto de corte; riesgo de mala aplicación del umbral.	Verificar si el umbral se aplicó de manera estricta; comprobar que los pacientes a ambos lados del corte son comparables; reflexionar sobre la generalización de los resultados.

Diseño	Ejemplo aplicado al caso clínico (PIRN vs rehabilitación convencional)	Riesgo principal	Claves de lectura crítica
Asignación por alternancia o cuotas	Derivar a PIRN a pacientes que ingresan lunes/miércoles/viernes y a convencional los martes/jueves.	Riesgo de sesgo si los pacientes eligen acudir en determinados días o si el patrón de ingreso no es aleatorio.	Revisar si la regla se aplicó sin manipulación; analizar posibles patrones de ingreso; considerar que es un método más frágil que la aleatorización verdadera.

Fuente: Elaboración propia.

5. EVALUACIÓN DE LA VALIDEZ CLÍNICA Y RELEVANCIA

Uno de los mayores desafíos al interpretar un estudio cuasiexperimental es determinar qué tan confiables son sus resultados y si realmente pueden aplicarse a la práctica clínica. A diferencia de los ensayos aleatorizados, estos diseños requieren un análisis más cuidadoso para detectar posibles sesgos y valorar su impacto en las conclusiones. La validez clínica y la relevancia de un estudio cuasiexperimental pueden evaluarse en cinco aspectos fundamentales:

Selección de los grupos. La población suele seleccionarse por “conveniencia” e “idoneidad” por lo tanto, es esencial verificar si los grupos comparados eran similares al inicio del estudio. Si no lo eran, se debe analizar si los autores aplicaron estrategias para equilibrar esas diferencias (ajustes estadísticos, emparejamiento por puntuación de propensión, estratificación). Ejemplo del caso clínico: pacientes asignados a PIRN podrían haber sido, en promedio, más jóvenes y menos graves que los de rehabilitación convencional. La lectura crítica debe comprobar si los investigadores describieron bien estas diferencias y cómo las controlaron. **Medición de las variables.** Los desenlaces deben ser claros, objetivos y medidos de forma estandarizada en todos los grupos. Siempre es deseable que los evaluadores estén cegados a la intervención recibida. *Ejemplo:* la escala Fugl-Meyer o el Índice de Barthel (instrumentos estandarizados) deben aplicarse con el mismo protocolo, sin que el evaluador sepa a qué programa asistió el paciente. **Control de factores de confusión.** Un buen estudio cuasiexperimental no se limita a comparar resultados crudos, sino que ajusta por variables que podrían explicar las diferencias. *Ejemplo:* ajustar por edad, tiempo desde el ACV, severidad inicial y comorbilidades, porque todos estos factores influyen en la recuperación, independientemente del tipo de rehabilitación.

Duración y seguimiento. El tiempo de observación debe ser suficiente para detectar efectos clínicamente relevantes, y las pérdidas de seguimiento deben analizarse con transparencia. Una pérdida elevada o desigual entre grupos puede sesgar los resultados. *Ejemplo:* si en PIRN más pacientes abandonan porque no pueden asistir tres horas al día, los resultados podrían parecer mejores simplemente porque solo permanecieron los más motivados.

Generalización y aplicabilidad. El lector debe preguntarse si los hallazgos se pueden extrapolar a su propio contexto clínico. Un programa exitoso en un hospital de referencia con gran infraestructura puede no tener el mismo impacto en un centro de menor complejidad.

6. INTERPRETACIÓN ESTADÍSTICA

La estadística en un estudio cuasiexperimental no es un fin en sí misma, sino una herramienta que ayuda al lector a distinguir si los cambios observados se deben realmente a la intervención o a factores ajenos. Para la lectura crítica, lo más importante no es memorizar las pruebas utilizadas, sino saber qué significan los resultados y cómo interpretarlos clínicamente.

6.1 COMPARACIONES INICIALES

El primer paso es analizar si los grupos eran comparables antes de la intervención. Esto se suele presentar en la tabla de características basales, donde se describen variables como edad, sexo, gravedad de la enfermedad o presencia de comorbilidades. Si desde el inicio un grupo está compuesto por pacientes más jóvenes, con menos comorbilidades o con una enfermedad menos severa, es muy probable que ese grupo muestre mejores resultados, aun cuando ambos reciban exactamente el mismo tratamiento.

Para entenderlo mejor se presenta la siguiente tabla del estudio:

Tabla 2. Características basales de los pacientes incluidos en el estudio cuasiexperimental sobre rehabilitación post-ACV isquémico en el Hospital de Clínicas de La Paz (n=100).

Variable	PIRN (n=50)	Convencional (n=50)	Diferencia (IC95%)	p-valor
Edad (años, media $\pm$ DE)	62 $\pm$ 10	68 $\pm$ 9	-6 (-10 a -2)	0,01
Sexo masculino (%)	60	58	+2 (-18 a +22)	0,84
NIHSS basal (media $\pm$ DE)	9,2 $\pm$ 2,1	11,0 $\pm$ 2,3	-1,8 (-2,9 a -0,7)	0,002
Tiempo a inicio de rehab. (días)	6 $\pm$ 2	8 $\pm$ 3	-2 (-3,4 a -0,6)	0,03
Hipertensión (%)	70	74	-4 (-22 a +14)	0,65
Diabetes mellitus (%)	30	28	+2 (-15 a +19)	0,82
Fibrilación auricular (%)	18	22	-4 (-18 a +10)	0,6
mRS previo al ACV (0-2, %)	96	94	+2 (-7 a +11)	0,72

La columna de diferencias incluye el intervalo de confianza al 95% (IC95%) para cada comparación. PIRN: Programa Intensivo de Rehabilitación Neuromotora; ACV: accidente cerebrovascular; NIHSS: National Institutes of Health Stroke Scale; mRS: modified Rankin Scale. El valor de p se calculó con la prueba t de Student para variables continuas y la prueba χ^2 para variables categóricas. Un $p < 0,05$ se considera significativo.

La tabla 2 de características basales es el primer filtro que el lector debe revisar en cualquier estudio cuasiexperimental. Allí se observa si los grupos de comparación (en este caso, PIRN vs rehabilitación convencional) eran semejantes al inicio, hay tres diferencias importantes:

- **Edad:** el grupo PIRN tiene una media de 62 años, frente a 68 en el grupo convencional. La diferencia es de -6 años (IC95%: -10 a -2), estadísticamente significativa ($p = 0,01$). Esta diferencia no es trivial: los pacientes más jóvenes suelen recuperarse mejor tras un ACV, lo que podría favorecer artificialmente al PIRN.
- **Severidad inicial del ACV (NIHSS):** el grupo PIRN presenta menor gravedad inicial (9,2 puntos vs 11,0). La diferencia de -1,8 puntos (IC95%: -2,9 a -0,7) también es significativa ($p = 0,002$). Esta es quizá la diferencia más relevante, porque la severidad inicial es uno de los predictores más fuertes de la recuperación neurológica.
- **Tiempo hasta el inicio de la rehabilitación:** los pacientes de PIRN comenzaron en promedio dos días antes que los del grupo convencional (6 vs 8 días, $p = 0,03$). El inicio precoz de la rehabilitación es un factor reconocido que mejora los desenlaces, lo que puede introducir un sesgo a favor de PIRN.

En contraste, otras variables como sexo, hipertensión, diabetes, fibrilación auricular y función previa (mRS antes del ACV) muestran distribuciones similares, sin diferencias significativas. Esto significa que, al menos en esas dimensiones, los grupos son comparables.

Por ello, el lector debe acostumbrarse a revisar con cuidado la tabla basal y preguntarse: ¿los grupos eran realmente comparables? ¿El estudio aplicó algún tipo de ajuste para equilibrar esas diferencias? Esta reflexión es esencial para determinar si los resultados reflejan el efecto de la intervención o si, en realidad, lo que se observa es consecuencia de un sesgo de selección.

6.2 CAMBIOS DENTRO DE CADA GRUPO

Un error frecuente al interpretar estudios cuasiexperimentales es entusiasmarse únicamente con la mejoría observada en el grupo que recibe la intervención. Que un grupo de pacientes mejore no significa necesariamente que el tratamiento sea responsable de ese cambio. La comparación entre grupos es la que otorga validez al hallazgo. Observar mejoría dentro de un solo grupo puede ser interesante, pero no basta para atribuir causalidad.

Para poder entender mejor ello, se presenta la siguiente tabla:

Tabla 3. Resultados clínicos a 6 meses en pacientes con accidente cerebrovascular isquémico tratados con Programa Intensivo de Rehabilitación Neuromotora (PIRN) o rehabilitación convencional (n=100).

Desenlace	PIRN (n=50)	Convencional (n=50)	Diferencia ajustada (IC95%)	p-valor
Fugl-Meyer (cambio, media $\pm$ DE)	+15 $\pm$ 5	+8 $\pm$ 6	+6,5 (3,2 a 9,8)	<0,001
Barthel (cambio, media $\pm$ DE)	+40 $\pm$ 10	+28 $\pm$ 12	+9,0 (4,0 a 14,0)	0,002
mRS 0-2 (%)	60	40	RR 1,5 (1,1 a 2,1)	0,01
Reingresos hospitalarios (%)	10	18	-8 (-20 a +4)	0,18

Desenlace	PIRN (n=50)	Convencional (n=50)	Diferencia ajustada (IC95%)	p-valor
Adherencia ($\geq 80\%$ sesiones, %)	70	78	-8 (-24 a +8)	0,3

La columna de diferencias muestra estimaciones ajustadas por edad, severidad inicial (NIHSS) y tiempo al inicio de rehabilitación. IC95%: intervalo de confianza al 95%; RR: riesgo relativo. Un $p < 0,05$ se considera significativo.

Para su interpretación crítica:

- Escalas motoras y funcionales (Fugl-Meyer y Barthel): los pacientes de PIRN mejoraron más que los de rehabilitación convencional. La diferencia ajustada en Fugl-Meyer fue de 6,5 puntos (IC95%: 3,2 a 9,8), lo que indica no solo significancia estadística ($p < 0,001$), sino también una diferencia clínicamente relevante.
- Discapacidad global (mRS): el 60% de los pacientes en PIRN alcanzó independencia funcional (mRS 0–2), frente al 40% en el grupo convencional. El riesgo relativo de 1,5 (IC95%: 1,1 a 2,1) significa que los pacientes en PIRN tuvieron un 50% más de probabilidad de lograr un buen desenlace.
- Reingresos hospitalarios: fueron menores en PIRN (10% vs 18%), pero la diferencia no fue significativa (IC95% incluye 0, $p = 0,18$). Esto debe interpretarse con cautela: podría ser un hallazgo por azar o un efecto real que el estudio no tuvo poder suficiente para detectar.
- Adherencia: la adherencia fue algo menor en PIRN (70% vs 78%), aunque sin diferencias significativas. Esto sugiere que la intensidad del programa puede haber dificultado la asistencia completa, un aspecto relevante para valorar la aplicabilidad en la práctica clínica.

6.3 COMPARACIÓN ENTRE GRUPOS

El corazón de todo estudio cuasiexperimental está en la comparación entre los grupos. No basta con saber que los pacientes de un programa mejoran, sino determinar si esa mejoría es mayor que la observada en el grupo de control.

Primero, el lector debe preguntarse si esa diferencia es estadísticamente significativa, es decir, si el resultado probablemente no se deba al azar.

Segundo, y aún más importante, se debe valorar si la diferencia es clínicamente relevante.

Para entender ello, se usará la tabla 3 para interpretar la comparación entre grupos donde los pacientes del PIRN mejoraron en promedio 15 puntos en la escala Fugl-Meyer, mientras que los de rehabilitación convencional mejoraron 8 puntos. La diferencia bruta es de 7 puntos a favor del PIRN. A primera vista, este dato suena convincente. Sin embargo, el lector crítico no debe quedarse solo con el número. Lo relevante es analizar dos aspectos:

- La precisión estadística del hallazgo. Esta se refleja en el intervalo de confianza (IC). Si la diferencia fue de 7 puntos con un IC95%: 3 a 11, significa que, incluso en el escenario más conservador, el PIRN sigue ofreciendo al menos 3 puntos adicionales de recuperación frente al tratamiento convencional. (En diferencias de medias, el IC no debe cruzar el 0: si lo hace, podría indicar que la intervención no tiene efecto real, en medidas de asociación como riesgos relativos u odds ratio, el IC no debe cruzar el 1, porque esto indicaría ausencia de efecto).
- La relevancia clínica. Más allá de la estadística, el lector debe preguntarse si el efecto observado tiene un impacto en la vida del paciente. En rehabilitación post-ACV, una ganancia de 7 puntos en Fugl-Meyer puede significar la diferencia entre un paciente que recupera la movilidad funcional del brazo y otro que permanece con limitaciones significativas.

En síntesis, la comparación entre grupos no se limita a “quién mejoró más”, sino que requiere responder dos preguntas clave: ¿La diferencia observada es estadísticamente confiable (IC y pvalor) ?, ¿Esa diferencia es lo suficientemente grande como para tener impacto en la práctica clínica y en la calidad de vida del paciente?

6.4 AJUSTES POR DIFERENCIAS BASALES

Las diferencias entre grupos al inicio pueden sesgar los resultados. Por eso, los análisis ajustados —que corrigen por variables como edad, severidad de la enfermedad o tiempo hasta la intervención— son más confiables que las comparaciones crudas.

Los investigadores aplican métodos estadísticos para corregir esas diferencias basales. La idea es “nivelar el terreno” y que la comparación sea lo más justa posible. Existen varias estrategias:

- **Regresión multivariada:** es la más común. Se introducen en el modelo estadístico las

variables que diferían entre grupos (por ejemplo, edad, severidad inicial, tiempo al inicio de la rehabilitación). Así se calcula el efecto de la intervención corrigiendo por esas variables.

- **Propensity Score (puntuación de propensión):** se estima la probabilidad de que cada paciente reciba la intervención según sus características basales. Luego se emparejan pacientes con puntuaciones similares o se aplica un peso estadístico. Esto ayuda a equilibrar los grupos.
- **Estratificación:** se divide la muestra en subgrupos homogéneos (por ejemplo, pacientes menores de 65 y mayores de 65 años) y se compara dentro de cada estrato.

En nuestro caso se realiza el ajuste con un modelo de regresión multivariada, que incluyó la edad, el puntaje NIHSS basal y el tiempo hasta el inicio de la rehabilitación contando con la siguiente tabla:

Tabla 4. Resultados principales (cambio en Fugl-Meyer a 6 meses) antes y después del ajuste por variables basales

Análisis	PIRN (n=50)	Convencional (n=50)	Diferencia (IC95%)	p-valor
Crudo	+15 ± 5	+8 ± 6	+7,0 (3,5 a 10,5)	<0,001
Ajustado*	+14,5 ± 5	+8,5 ± 6	+6,0 (2,8 a 9,2)	<0,001

*Ajustado por edad, severidad del ACV (NIHSS) y tiempo al inicio de rehabilitación.

Para su interpretación crítica:

- Análisis crudo: la diferencia inicial es de 7 puntos (IC95%: 3,5 a 10,5). Aunque significativa, esta estimación no considera que los pacientes de PIRN eran más jóvenes y menos graves.
- Análisis ajustado: al corregir por edad, severidad y tiempo al inicio, la diferencia se reduce a 6 puntos (IC95%: 2,8 a 9,2), pero sigue siendo estadística y clínicamente significativa. Esto aumenta la confianza en que el efecto observado se debe al programa y no solo a un sesgo de selección.
- Reflexión clave para el lector: si el efecto desaparece después del ajuste, probablemente estaba explicado por diferencias iniciales entre los grupos. En cambio, si el efecto se mantiene (aunque sea un poco menor), es más confiable atribuirlo a la intervención.

6.5 MEDIDAS DE EFECTO MAS COMUNES EN ESTUDIOS CUASIEXPERIMENTALES

Los resultados suelen expresarse mediante tres medidas de efecto: diferencias de medias, riesgos relativos (RR) y odds ratio (OR). La Tabla 2 del caso clínico ofrece ejemplos claros de cómo se reportan y cómo deben interpretarse.

Diferencia de medias: Es la distancia entre los promedios de un desenlace en dos grupos. Se usa cuando la variable es continua (p. ej., puntuaciones de escalas clínicas).

En la escala Fugl-Meyer, los pacientes del PIRN mejoraron 15 puntos, frente a 8 puntos en el grupo convencional. La diferencia ajustada fue de 6,5 puntos (IC95%: 3,2 a 9,8; $p < 0,001$).

- Interpretación: los pacientes en PIRN mejoraron, en promedio, 6,5 puntos más que los de rehabilitación convencional.
- Clave crítica: como el IC no cruza el 0, la diferencia es estadísticamente significativa. Además, una ganancia de más de 5 puntos en Fugl-Meyer se considera clínicamente importante, lo que refuerza la relevancia del hallazgo.

Riesgo relativo (RR): Es la razón de probabilidades absolutas de que ocurra un desenlace en un grupo respecto al otro. Se usa con variables dicotómicas (sí/no).

Al analizar el grado de independencia global (mRS 0–2), el 60% de los pacientes en PIRN alcanzó este desenlace, frente al 40% en el grupo convencional. El RR fue 1,5 (IC95%: 1,1 a 2,1; $p = 0,01$).

- Interpretación: los pacientes en PIRN tuvieron 1,5 veces más probabilidad de lograr independencia funcional que los de rehabilitación convencional.
- Clave crítica: como el IC no cruza el 1, el resultado es significativo. Además, un aumento relativo del 50% en la independencia funcional tiene gran impacto clínico.

Odds ratio (OR): Aunque en esta tabla no se muestra un OR explícito, algunos artículos lo reportan en lugar del RR. La lógica es similar:

- Ejemplo hipotético: un OR de 2,0 para independencia funcional significa que las “odds” (oportunidades) de lograr ese desenlace se duplican con PIRN frente a rehabilitación convencional.
- Clave crítica: igual que con el RR, un OR es significativo si su IC no cruza el 1.

Tabla 5. Medidas de efecto más comunes en estudios cuasiexperimentales. Ejemplo aplicado al caso

Medida de efecto	Definición básica	clínico (PIRN vs rehabilitación convencional)	Clave de interpretación
Diferencia de medias (DM)	Distancia entre los promedios de un desenlace en dos grupos. Se aplica a variables continuas.	Fugl-Meyer: PIRN mejoró 15 puntos, convencional 8. La DM = 7 puntos a favor de PIRN.	El IC no debe incluir 0. Una diferencia distinta de 0 indica un efecto real.
Riesgo relativo (RR)	Razón entre las probabilidades absolutas de un desenlace en dos grupos. Se aplica a variables dicotómicas (sí/no).	Independencia funcional (mRS 0-2): PIRN 60%, convencional 40%. El RR = 1,5.	El IC no debe incluir 1. Un RR > 1 indica mayor probabilidad de desenlace con la intervención.
Odds ratio (OR)	Razón de las "odds" (oportunidades) de un desenlace en un grupo respecto al otro. Similar al RR, pero usa odds en lugar de riesgos absolutos.	Hipotético: OR de independencia funcional con PIRN = 2,0, lo que significa que las oportunidades de lograrla se duplican frente al grupo convencional.	El IC no debe incluir 1. Un OR > 1 indica mayor oportunidad de desenlace con la intervención.

Fuente: Elaboración propia.

6.6 ANÁLISIS DE SENSIBILIDAD

El análisis de sensibilidad es un paso clave en los estudios cuasiexperimentales, porque permite comprobar si los resultados son robustos, es decir, si se mantienen estables cuando se modifican ciertas condiciones del análisis. En otras palabras, es una forma de preguntarse: ¿seguiríamos llegando a la misma conclusión si analizáramos los datos de otra manera?

- El análisis de sensibilidad puede hacerse de distintas maneras:
- Excluyendo pacientes perdidos o con datos incompletos, para ver si los resultados cambian.
- Probando distintos métodos de ajuste (por ejemplo, distintos modelos de regresión o diferentes calipers en el emparejamiento por puntuación de propensión).
- Analizando subgrupos específicos (por ejemplo, solo pacientes con ACV moderado o solo los mayores de 65 años).

Para su interpretación: si los resultados cambian drásticamente en los análisis de sensibilidad, significa que las conclusiones del estudio son frágiles y deben interpretarse con cautela. En cambio, si los resultados se mantienen similares bajo distintos escenarios, el lector puede tener mayor confianza en la solidez de la evidencia.

Para comprenderlo se presenta la siguiente tabla:

Tabla 6. Análisis de sensibilidad para el desenlace primario (cambio en Fugl-Meyer a 6 meses) en pacientes con ACV isquémico (n=100).

Análisis	PIRN (n=50)	Convencional (n=50)	Diferencia (IC95%)	p-valor
Crudo	+15 ± 5	+8 ± 6	+7,0 (3,5 a 10,5)	<0,001
Ajustado*	+14,5 ± 5	+8,5 ± 6	+6,0 (2,8 a 9,2)	<0,001
Excluyendo abandonos	+15 ± 5	+9 ± 6	+6,0 (2,5 a 9,5)	0,001
Propensity score (emparejamiento 1:1)	+14,8 ± 5	+9,1 ± 6	+5,7 (2,3 a 9,1)	0,002
Propensity score (ponderación por IPTW)	+14,6 ± 5	+8,8 ± 6	+5,8 (2,6 a 9,0)	0,001

* Ajustado por edad, severidad inicial (NIHSS) y tiempo al inicio de rehabilitación.

Para su interpretación crítica de la tabla

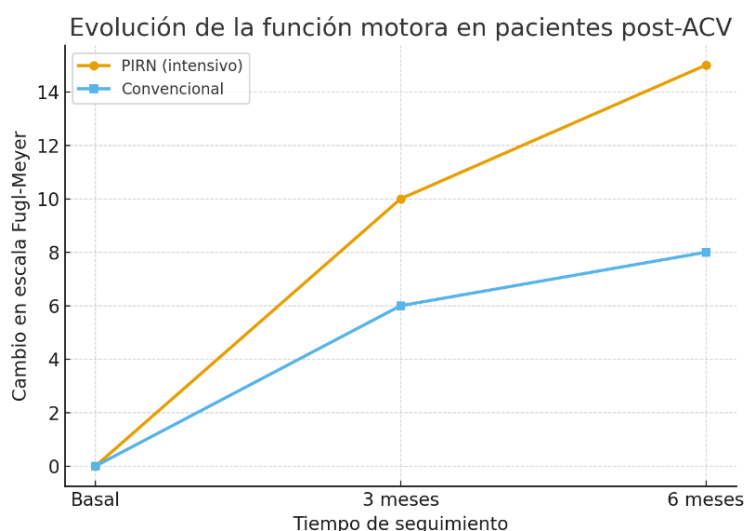
- Consistencia de los resultados: en todos los escenarios (crudo, ajustado, excluyendo abandonos, distintos métodos de propensity score), la diferencia entre PIRN y convencional se mantiene entre 5,7 y 7 puntos, siempre significativa.
- Robustez: como los hallazgos no cambian sustancialmente al modificar la estrategia de análisis, se puede tener mayor confianza en que el beneficio del PIRN es real y no producto de un artefacto estadístico.
- Reflexión crítica para el lector: si en algún escenario los resultados hubieran perdido significancia o se hubieran invertido, la conclusión sería frágil. El lector debe preguntarse: ¿el efecto se mantiene estable en todos los análisis de sensibilidad?

6.7 REPRESENTACIÓN GRÁFICA DE RESULTADOS

Gráfico de evolución en el tiempo (líneas o barras): Los gráficos permiten visualizar la evolución y la magnitud de las diferencias entre grupos de forma más clara que las tablas, donde:

- Los gráficos de líneas se utilizan para mostrar cambios a lo largo del tiempo, permitiendo observar tendencias y la pendiente de mejoría.
- Los gráficos de barras se emplean para comparar proporciones o valores finales entre grupos, facilitando la interpretación de diferencias puntuales.

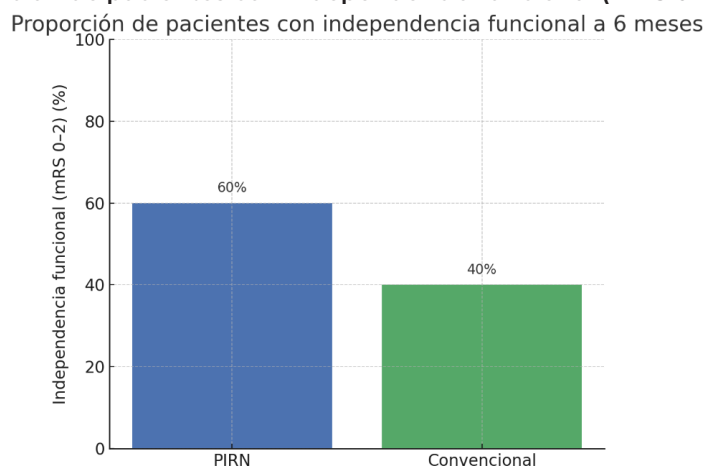
Figura 1: Evolución del cambio en la escala Fugl-Meyer en pacientes post-ACV, medido al inicio, a los 3 y a los 6 meses.



Para su interpretación: ambos grupos mejoran con el tiempo, pero la curva del PIRN es más pronunciada y se separa progresivamente de la convencional. Esto sugiere que el beneficio adicional puede atribuirse al programa intensivo, siempre que la diferencia sea estadísticamente significativa.

Gráfico de barras comparativas: Este tipo de gráfico es muy usado en estudios cuasiexperimentales porque permite comparar proporciones de desenlaces entre dos o más grupos de forma inmediata y visual. Cada barra representa la frecuencia o porcentaje de un desenlace (por ejemplo, el porcentaje de pacientes que logran independencia funcional). Para la Interpretación de este tipo de graficas lo primero es recordar que la diferencia visual entre las barras no garantiza por sí misma un efecto real, es indispensable comprobar si esa diferencia es estadísticamente significativa, además, el lector debe valorar si la diferencia es clínicamente relevante y otro aspecto importante es considerar el tamaño muestral, para entender mejor la interpretación grafica se presenta lo siguiente:

Figura 2: Proporción de pacientes con independencia funcional (mRS 0-2) a los 6 meses.

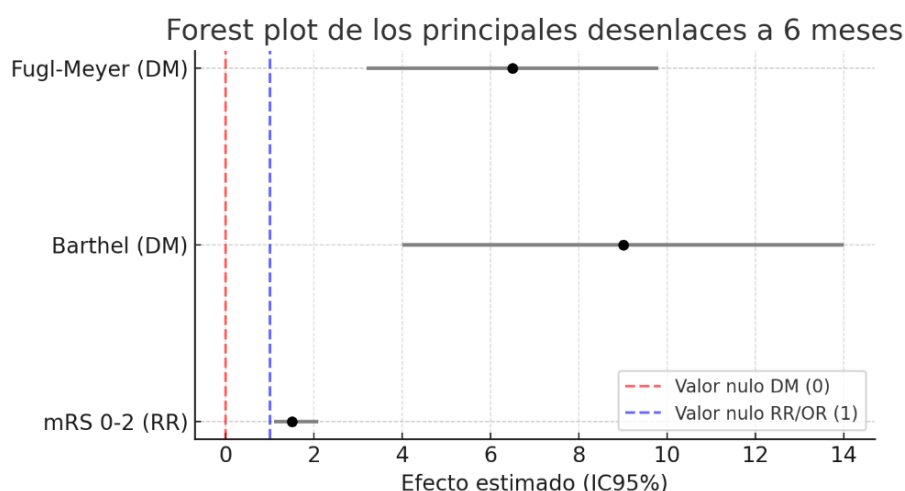


Para su interpretación: el 60% de los pacientes en PIRN alcanzó independencia funcional frente al 40% en la rehabilitación convencional. La diferencia se visualiza de inmediato en la altura de las barras. Lo importante no es solo que una barra sea más alta, sino comprobar si esa diferencia es estadísticamente significativa y si tiene un impacto clínico relevante en la recuperación de los pacientes.

Gráfico de Forest Plot: Es una forma visual de resumir los resultados principales de un estudio. Cada desenlace se representa con un punto central (el efecto estimado) y una línea horizontal (el intervalo de confianza al 95%), los principios generales de interpretación son los mismos:

- Si el desenlace es una diferencia de medias, el valor nulo es 0.
- Si el desenlace es un riesgo relativo (RR) u odds ratio (OR), el valor nulo es 1.
- La regla de oro es que, si la línea del intervalo cruza el valor nulo, el resultado no es estadísticamente significativo.

Figura 3: Principales desenlaces a 6 meses en pacientes post-ACV (PIRN vs rehabilitación convencional).



Para su interpretación crítica:

- Fugl-Meyer (DM 6,5 puntos, IC95%: 3,2 a 9,8): el intervalo no cruza el 0, lo que indica una diferencia real a favor de PIRN.
- Barthel (DM 9 puntos, IC95%: 4,0 a 14,0): el efecto también es estadística y clínicamente significativo.
- Independencia funcional mRS 0-2 (RR 1,5, IC95%: 1,1 a 2,1): como el IC no cruza el 1, los pacientes en PIRN tienen 1,5 veces más probabilidad de lograr independencia funcional.

6.8 OTROS PARÁMETROS A CONSIDERAR

En la práctica de lectura crítica hay otros aspectos adicionales que vale la pena mencionar brevemente para cerrar esta sección:

Tamaño muestral y potencia estadística: Un resultado no significativo ($p > 0,05$) no siempre significa ausencia de efecto: puede deberse a que el estudio tuvo muy pocos pacientes para detectar una diferencia real. Se debe revisar si los autores calcularon el tamaño muestral y si discuten la posibilidad de error tipo II (falso negativo).

- Ejemplo: si solo 20 pacientes completaron PIRN y 20 la rehabilitación convencional, incluso diferencias grandes pueden no alcanzar significancia.

Magnitud del efecto vs. Significancia: Un resultado puede ser estadísticamente significativo, pero clínicamente irrelevante. El lector debe siempre balancear significancia estadística con relevancia clínica.

- Ejemplo: una diferencia de 1 punto en Fugl-Meyer puede ser significativa si el estudio es muy grande, pero no cambia la vida del paciente.

Pérdidas de seguimiento y análisis por protocolo/intención de tratar: En estudios cuasiexperimentales, las pérdidas pueden ser diferentes entre grupos y distorsionar los resultados. Se debe revisar si los autores usaron análisis “por intención de tratar” (todos los pacientes

asignados) o solo “por protocolo” (solo los que completaron).

- Si muchos pacientes abandonan PIRN porque es más exigente, los resultados pueden reflejar solo a los más motivados.

Subgrupos y heterogeneidad de efectos: Algunos estudios reportan resultados diferentes en ciertos subgrupos (ej., mayores vs menores de 65 años). El lector debe ser cauto: si se hacen demasiados subgrupos, puede aparecer significancia por azar. Estos análisis ayudan a entender mejor para quién funciona la intervención.

Tabla 7. Guía de interpretación estadística en estudios cuasiexperimentales

Aspecto a revisar	Qué observar	Ejemplo aplicado al caso PIRN	Claves de interpretación crítica
1. Comparación inicial de grupos	Tabla de características basales	PIRN más jóvenes, menos graves y con inicio más precoz	Si las diferencias no se ajustan, los resultados estarán sesgados.
2. Cambios dentro de cada grupo	Evolución pre-post en cada rama	PIRN mejora 15 puntos en Fugl-Meyer	La mejoría puede deberse a recuperación natural. Se necesita comparar entre grupos.
3. Comparación entre grupos	Diferencia de mejoría entre intervención y control	PIRN +15 vs Convencional +8 → diferencia de 7	Lo relevante es la diferencia y si el IC no cruza el valor nulo.
4. Ajustes por diferencias basales	Análisis crudo vs ajustado	Diferencia ajustada = 6 puntos	Si el efecto se mantiene tras el ajuste, es más confiable.
5. Medidas de efecto	DM, RR, OR según el desenlace	DM = 6,5 puntos; RR = 1,5	En DM el IC no debe cruzar 0; en RR/OR no debe cruzar 1.
6. Intervalos de confianza y p-valor	Precisión y significancia	IC95%: 3,2 a 9,8; p<0,001	El IC informa magnitud y precisión; no basta con el p-valor.
7. Análisis de sensibilidad	Resultados bajo distintos escenarios	Excluyendo abandonos: diferencia se mantiene en 6	Si los resultados son consistentes, el hallazgo es robusto.
8. Representación gráfica	Líneas, barras, forest plots	Líneas divergen; barra PIRN más alta; IC sin cruzar el nulo	Gráficos ayudan a visualizar tendencias y la magnitud del efecto.
9. Tamaño muestral y potencia	Si hubo cálculo de muestra	n=100 puede detectar diferencias claras	Un estudio pequeño puede no mostrar efectos aunque existan.
10. Magnitud del efecto vs relevancia clínica	Diferencia significativa pero pequeña	Ganancia de 1 punto en Fugl-Meyer, aunque p<0,05	Puede ser irrelevante para el paciente. Importa la utilidad clínica.
11. Pérdidas de seguimiento	Porcentaje y manejo de abandonos	PIRN con más pérdidas por intensidad	Pérdidas desiguales sesgan; preferir análisis por intención de tratar.
12. Subgrupos y heterogeneidad	Análisis por edad, severidad, etc.	PIRN más efectivo en <65 años	Interpretar con cautela: muchos subgrupos aumentan riesgo de azar.

Fuente: Elaboración propia.

7. SESGOS Y LIMITACIONES

A diferencia de los ensayos clínicos aleatorizados, los estudios cuasiexperimentales son más vulnerables a sesgos. Estos sesgos no invalidan necesariamente el estudio, pero sí obligan al lector a interpretarlo con cautela. La clave de la lectura crítica está en identificar de qué manera esos sesgos pudieron influir en los resultados y si los autores tomaron medidas para reducirlos.

Uno de los más frecuentes es el sesgo de selección, que ocurre cuando los grupos no son equivalentes desde el inicio. Si un grupo comienza con pacientes más jóvenes o menos graves, sus resultados tenderán a ser mejores independientemente de la intervención. *Ejemplo PIRN:* los pacientes asignados al programa intensivo eran más jóvenes, tenían menor severidad de ACV y comenzaron la rehabilitación antes que los del grupo convencional.

El sesgo de confusión aparece cuando una variable externa, no medida, se asocia tanto con la intervención como con el desenlace, generando una falsa relación de causalidad. *Ejemplo PIRN:* el apoyo familiar podría haber sido mayor en los pacientes que ingresaron al programa intensivo, y este factor por sí solo facilita la recuperación funcional.

También es importante considerar el sesgo de información, que se produce cuando los desenlaces se miden de manera diferente entre grupos. Esto puede suceder si el evaluador conoce la intervención recibida y su expectativa influye en la medición. *Ejemplo PIRN:* si el evaluador sabe que un paciente participó en PIRN, podría calificar con mayor optimismo la escala Fugl-Meyer.

Otro riesgo es el **sesgo por pérdidas de seguimiento**. Si un grupo pierde más pacientes durante el seguimiento, los resultados pueden no reflejar la realidad de toda la cohorte, sino solo de quienes permanecieron. *Ejemplo PIRN:* algunos pacientes pueden abandonar el programa

intensivo por su alta exigencia, quedando en el análisis final solo los más motivados.

Finalmente, se debe reconocer que existen **limitaciones propias del diseño cuasiexperimental**. Estos estudios reflejan mejor la práctica clínica cotidiana, pero no siempre logran controlar por completo los factores externos que ocurren de manera simultánea. *Ejemplo PIRN*: en paralelo a la implementación del programa, el hospital pudo haber contratado más fisioterapeutas o mejorado sus protocolos de atención, lo que también impactaría en los resultados.

8. COMPARACIONES CON OTROS ESTUDIOS

Una parte esencial de la lectura crítica es situar los resultados dentro del marco de la evidencia existente. Un estudio aislado difícilmente puede cambiar la práctica clínica; es al contrastarlo con otros trabajos similares donde se mide su verdadero valor.

Al comparar un estudio cuasiexperimental con otros, el lector debe preguntarse:

- ¿Los hallazgos son coherentes con lo que se ha observado en ensayos clínicos aleatorizados, cohortes o revisiones sistemáticas?
- ¿Existen diferencias importantes y, si es así, pueden explicarse por variaciones en la población estudiada, la intensidad de la intervención o la metodología empleada?
- ¿El contexto del estudio (hospital público, recursos limitados, población local) aporta información complementaria que no estaba presente en la literatura previa?

Un estudio cuasiexperimental bien conducido que coincide con la evidencia de mayor jerarquía puede reforzar la aplicabilidad de esa intervención en escenarios del mundo real. Por el contrario, si muestra resultados muy distintos, obliga al lector a reflexionar sobre posibles sesgos o sobre características propias del contexto que podrían explicar la discrepancia.

En nuestro caso clínico: El programa PIRN mostró una ganancia media de 6 puntos en la escala Fugl-Meyer a los seis meses. Al contrastar este resultado con la literatura internacional, se observa que varios ensayos clínicos controlados han reportado beneficios similares de la rehabilitación intensiva post-ACV, aunque en contextos con más recursos y bajo condiciones muy controladas. El estudio cuasiexperimental aporta así un valor añadido: confirma que un programa intensivo también puede ser efectivo en un hospital público de altura, con limitaciones de personal y cupos.

9. Checklist. En cuasiexperimentales la fortaleza no depende solo de “si hubo diferencia”, sino de cómo se asignaron los grupos, qué tanto se controló la confusión, qué tan bien se describió la intervención (para poder reproducirla) y qué tan transparente es el reporte. Para juzgar esto, conviene combinar estándares de reporte (qué debe describir el artículo: TREND/STROBE/TIDieR) con estándares de validez (qué riesgos de sesgo revisar: ROBINS-I) y criterios específicos por diseño (p. ej., ITS según EPOC/Cochrane).

Tabla 8. Checklist internacional para lectura crítica de estudios cuasiexperimentales

Sección	Pregunta clave	Fuente / Guía	Sí / No / N.A.
A. Identificación del diseño y encuadre	¿Se identifica claramente el diseño cuasiexperimental (grupos no equivalentes, antes-después, CBA, ITS) y la unidad de análisis?	EPOC (Cochrane)	☐
	¿Se justifica por qué no fue posible aleatorizar (factibilidad/ética)?	TREND (CDC)	☐
B. Población, escenario y selección	¿Se describen criterios de inclusión/exclusión, fuente de pacientes, entorno y periodo de estudio?	STROBE (SciELO)	☐
	¿Se explica el proceso de derivación/asignación administrativa y sus implicancias en comparabilidad basal?	TREND / ROBINS-I	☐
C. Intervención y comparador (reproducibilidad)	¿La intervención y el comparador están descritos según estándar TIDieR (qué, quién, cómo, dónde, duración, adaptaciones, fidelidad)?	TIDieR (EQUATOR)	☐
D. Medición de desenlaces y cegamiento	¿Se definen desenlaces primarios/secundarios, instrumentos validados, ventanas temporales y procedimientos de cegamiento de evaluadores?	STROBE / TREND	☐
	¿Se reportan eventos adversos y adherencia/fidelidad de la intervención?	TREND / TIDieR	☐
E. Comparabilidad basal y control de confusión	¿Se presentan características basales completas y balance entre grupos?	STROBE / TREND	☐
	¿Se aplican métodos de control de confusión (regresión, propensity score, DiD, ITS)?	TREND / EPOC / Cochrane	☐

Sección	Pregunta clave	Fuente / Guía	Sí / No / N.A.
F. Supuestos del método y especificaciones por diseño	En DiD: ¿se evalúa la asunción de tendencias paralelas y se discuten common shocks?	SpringerLink	☐
	En ITS: ¿hay suficientes puntos pre y post (≥ 3) y se usa regresión segmentada con autocorrelación/estacionalidad?	EPOC (Cochrane)	☐
G. Estimación y presentación de efectos	¿Se reportan medidas de efecto apropiadas (DM, RR, OR) con IC95% y p-valores, y se discute relevancia clínica?	STROBE / TREND	☐
	¿Se presentan tablas/figuras interpretables (líneas de tiempo, barras, forest plot con valores nulos 0/1)?	STROBE (buenas prácticas)	☐
H. Manejo de datos y análisis complementarios	¿Se explican datos faltantes (método y justificación) y se hacen análisis de sensibilidad?	TREND / ROBINS-I	☐
	¿Se reporta análisis por intención de tratar o por protocolo, y pérdidas de seguimiento cuantificadas?	STROBE / TREND	☐
I. Riesgo de sesgo (ROBINS-I)	¿Se evalúan los siete dominios (confusión, selección, clasificación, desviaciones, datos faltantes, medición de desenlace, reporte selectivo) y se presenta juicio global?	ROBINS-I (BMJ/PMC)	☐
J. Validez externa, equidad y contexto	¿Se discute la generalización, subgrupos relevantes y equidad?	STROBE	☐
K. Transparencia y ética	¿Se informa financiación/conflictos, aprobación ética, protocolo/registro o disponibilidad de datos/código?	STROBE / TREND	☐
L. Coherencia con la evidencia	¿Se comparan los resultados con RCTs, cohortes o revisiones y se discuten discrepancias plausibles?	Buenas prácticas / SQUIRE	☐

Fuente: Elaboración propia.

10. UTILIDAD DE LOS ESTUDIOS CUASIEXPERIMENTALES

La principal fortaleza de los estudios cuasiexperimentales es su aplicabilidad en la práctica clínica real. Aunque no alcanzan el rigor metodológico de los ensayos clínicos aleatorizados, estos estudios son particularmente útiles para:

Generar evidencia en contextos reales. A diferencia de los ensayos clínicos, que se realizan en condiciones altamente controladas, los cuasiexperimentales reflejan las dinámicas cotidianas de hospitales, programas comunitarios o políticas de salud.

Evaluar intervenciones a nivel poblacional o de sistemas de salud. Muchas políticas públicas o programas hospitalarios no pueden asignarse al azar. Los cuasiexperimentales permiten valorar su impacto de manera estructurada.

Proveer evidencia preliminar para futuros ensayos clínicos. Un resultado consistente en un cuasiexperimental puede servir de base para diseñar un ensayo aleatorizado posterior.

Ofrecer respuestas rápidas. En situaciones donde no se puede esperar años para un RCT, un cuasiexperimental bien diseñado puede aportar información útil para orientar decisiones clínicas o de gestión.

En síntesis, la utilidad de los cuasiexperimentales radica en su capacidad de generar evidencia relevante, contextualizada y pragmática, que sirve tanto para la toma de decisiones clínicas y de políticas de salud, como para sentar las bases de investigaciones más robustas.

Tabla 9. Fortalezas y limitaciones de los estudios cuasiexperimentales

Aspecto	Fortalezas	Limitaciones
Aplicabilidad	Reflejan la práctica clínica y los escenarios reales (hospitales públicos, programas comunitarios, políticas de salud).	Resultados menos controlados, con mayor influencia de factores externos.
Factibilidad	Son posibles cuando la aleatorización no es viable por razones éticas, logísticas o administrativas.	Mayor riesgo de sesgo de selección y confusión al no existir asignación aleatoria.
Rapidez y costo	Generalmente más rápidos y menos costosos que los ensayos clínicos aleatorizados.	Pueden tener menor poder estadístico si la muestra es pequeña o las pérdidas son altas.
Generación de hipótesis	Proveen evidencia preliminar para diseñar RCTs o guías clínicas.	No establecen causalidad con la misma solidez que los ensayos aleatorizados.
Impacto en políticas de salud	Permiten evaluar intervenciones poblacionales o cambios de sistema (ej., programas de rehabilitación, leyes de salud pública).	Sus conclusiones suelen estar muy ligadas al contexto, lo que limita la generalización.

Fuente: Elaboración propia

11. PREGUNTAS DE AUTOEVALUACIÓN

PREGUNTAS DE SELECCIÓN MÚLTIPLE

1. ¿Cuál es la principal característica de un estudio cuasiexperimental?
 - a) No interviene el investigador
 - b) La asignación a grupos no es aleatoria
 - c) Solo estudia variables descriptivas
 - d) No necesita grupo comparador
2. ¿Qué sesgo es más frecuente en los estudios cuasiexperimentales con grupos no equivalentes?
 - a) Sesgo de selección
 - b) Sesgo de información
 - c) Sesgo de publicación
 - d) Sesgo de memoria
3. Si la diferencia de medias entre dos grupos es 5 puntos, con un IC95%: -1 a 11, la conclusión correcta es:
 - a) La intervención es efectiva
 - b) El resultado no es significativo
 - c) El resultado es clínicamente relevante
 - d) El estudio demuestra causalidad
4. En el análisis de resultados de un estudio cuasiexperimental, ¿qué significa que un RR = 1,5 con IC95%: 1,1-2,0?
 - a) La intervención aumenta un 50% la probabilidad del desenlace, con efecto significativo
 - b) El desenlace es igual en ambos grupos
 - c) No hay diferencia estadística entre grupos
 - d) El resultado es clínicamente irrelevante

Respuesta correcta: a. El IC no incluye el 1, por lo que es significativo; el RR de 1,5 indica un 50% más de probabilidad.
5. ¿Cuál de las siguientes es una utilidad práctica de los estudios cuasiexperimentales?
 - a) Reemplazar los ensayos clínicos aleatorizados en todas las situaciones
 - b) Evaluar intervenciones en contextos reales donde la aleatorización no es posible
 - c) Confirmar causalidad con el mismo nivel que un ensayo clínico
 - d) Evitar el uso de análisis estadísticos complejos

Respuesta correcta: b. Son valiosos en la práctica real y en políticas de salud, cuando los RCTs no son factibles.

PREGUNTAS ABIERTAS

6. Mencione dos ventajas y dos limitaciones de los estudios cuasiexperimentales. Posible respuesta:
7. Explique con sus palabras por qué es importante revisar las características basales de los grupos en un estudio cuasiexperimental.
8. ¿Cómo debe interpretarse un intervalo de confianza que cruza el valor nulo (0 en diferencias, 1 en riesgos)?
9. Describa una situación clínica en su hospital o servicio donde un estudio cuasiexperimental sería más factible que un ensayo clínico aleatorizado.
10. ¿Por qué los análisis de sensibilidad fortalecen las conclusiones de un estudio cuasiexperimental?

RESPUESTAS:

1. b). La característica principal de un estudio cuasiexperimental es que la asignación a los grupos no se realiza de manera aleatoria. Esto implica que, aunque existe una intervención por parte del investigador, los grupos pueden diferir desde el inicio, lo que introduce riesgos metodológicos que deben ser considerados en la interpretación.
2. a). El sesgo más frecuente en los estudios cuasiexperimentales con grupos no equivalentes es el sesgo de selección. Este ocurre porque los grupos pueden estar formados por pacientes con características distintas desde el inicio, lo que puede influir directamente en los resultados

observados.

3. b). Cuando el intervalo de confianza incluye el valor nulo, en este caso el 0, no se puede afirmar que exista una diferencia real entre los grupos. Por lo tanto, el resultado no es estadísticamente significativo y debe interpretarse con cautela.

4. a). Un riesgo relativo de 1,5 con un intervalo de confianza que no incluye el 1 indica que la intervención aumenta en un 50% la probabilidad del desenlace y que este efecto es estadísticamente significativo. Esto sugiere una asociación real entre la intervención y el resultado observado.

5. b). Los estudios cuasiexperimentales son especialmente útiles para evaluar intervenciones en contextos reales donde la aleatorización no es posible, como en programas hospitalarios o políticas de salud pública, permitiendo generar evidencia aplicable a la práctica cotidiana

6. Los estudios cuasiexperimentales presentan ventajas como su capacidad para reflejar la práctica clínica real y su utilidad en situaciones donde la aleatorización no es factible. Sin embargo, también tienen limitaciones importantes, como un mayor riesgo de sesgos y una menor capacidad para establecer relaciones causales con la misma solidez que un ensayo clínico aleatorizado.

7. Revisar las características basales es fundamental porque permite determinar si los grupos eran comparables al inicio del estudio. Si existen diferencias importantes entre ellos, los resultados pueden estar influenciados por esas desigualdades y no por la intervención en sí.

8. Un intervalo de confianza que cruza el valor nulo indica que no se puede descartar la ausencia de efecto. En el caso de diferencias de medias, si el intervalo incluye el 0, y en el caso de riesgos relativos u odds ratio si incluye el 1, el resultado no es estadísticamente significativo.

9. Un ejemplo de situación en la que un estudio cuasiexperimental sería más factible es la implementación de un nuevo protocolo de tratamiento en un hospital, donde todos los pacientes deben recibir la intervención por razones éticas o administrativas, lo que impide la asignación aleatoria.

10. Los análisis de sensibilidad fortalecen las conclusiones porque permiten evaluar si los resultados se mantienen consistentes bajo diferentes condiciones o supuestos. Si los hallazgos persisten al modificar el análisis, se incrementa la confianza en que el efecto observado es real y no producto de un artefacto metodológico.

REFERENCIAS

- Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2002. Disponible en: <https://books.google.com/books?id=o7jaAAAAAMAAJ>
- Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. *J Am Med Inform Assoc*. 2006;13(1):16-23. <https://doi.org/10.1197/jamia.m1749>
- DesJarlais DC, Lyles C, Crepaz N; TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*. 2004;94(3):361-366. <https://doi.org/10.2105/ajph.94.3.361>
- Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, Ogilvie D, et al. Using natural experiments to evaluate population health interventions: new guidance. *J Epidemiol Community Health*. 2012;66(12):1182-1186. <https://doi.org/10.1136/jech-2011-200375>
- Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919. <https://doi.org/10.1136/bmj.i4919>
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4(10):e296. <https://doi.org/10.1371/journal.pmed.0040296>
- Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014;348:g1687. <https://doi.org/10.1136/bmj.g1687>
- Fuller T, Peters J, Pearson M, Anderson R. Evaluating the impact and use of Transparent Reporting of Evaluations with Non-randomised Designs (TREND) reporting guidelines. *BMJ Open*. 2012;2(6):e002073. doi:10.1136/bmjopen-2012-002073
- Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Chicago: Rand McNally; 1963.
- Wong S, Lee K-P, Wong JY, Chan AW, Chan HY, Pang SM, et al. Applying difference-in-differences design in quality improvement and health systems research. *J Am Geriatr Soc*. 2025;73(4):1035–44. Available from: <https://doi.org/10.1111/jgs.19180>
- Alsaggaf R, O'Hara LM, Stafford KA, Leekha S, Harris AD. Quasi-experimental studies in the fields of infection control and antibiotic resistance, ten years later: A systematic review. *Infect Control Hosp Epidemiol*. 2018;39(2):170–6. Available at: <https://doi.org/10.1017/ice.2017.296>
- Huang HL, Chou CP, Leu S, et al. Effects of a quasi-experimental study of using flipped classroom approach to teach evidence-based medicine to medical technology students. *BMC Med Educ*. 2020;20(1):31. Available at: <https://doi.org/10.1186/s12909-020-1946-7>
- Couto LA, Nieuwlaat R, Nolan R, et al. A quasi-experimental evaluation of a primary care behavioral health integration program based on the chronic care model. *J Gen Intern Med*. 2025. Available at: <https://doi.org/10.1007/s11606-025-09641-0>
- Caniglia EC, Murray EJ. Difference-in-difference in the time of cholera: A gentle introduction for epidemiologists. *Curr Epidemiol Rep*. 2020;7(4):203–11. Available at: <https://doi.org/10.1007/s40471-020-00245-2>