



# Probar hipótesis

## ¡Bienvenido a Probar *Hipótesis!*

A diferencia de los capítulos anteriores, que se centran en pasos y herramientas específicas para aplicar las ciencias del comportamiento, este capítulo está organizado de forma ligeramente diferente. Realizar evaluaciones rigurosas para medir la eficacia de las intervenciones de comportamiento es complejo y va más allá del alcance de esta guía de campo. No obstante, este capítulo ofrece una orientación esencial para ayudar a comprender el valor de la evaluación de impacto, los retos que plantea demostrar la causalidad y los elementos y decisiones clave que intervienen en el diseño y la coordinación de los experimentos. El objetivo de este capítulo es proporcionar una comprensión básica de la comprobación de hipótesis de intervención para permitir una colaboración más eficiente con los especialistas en evaluación.

En esta fase se examina si la intervención seleccionada contribuye a cambios medibles en el comportamiento objetivo al abordar las barreras claves del comportamiento (identificadas durante las fases de *Definición y Exploración y Diagnóstico*). Las perspectivas que se obtienen aquí sirven de guía para decidir si una intervención debe adaptarse, escalarse o interrumpirse, y sientan las bases para la fase de *Escalamiento*, en la que la evidencia se traduce en medidas más amplias.

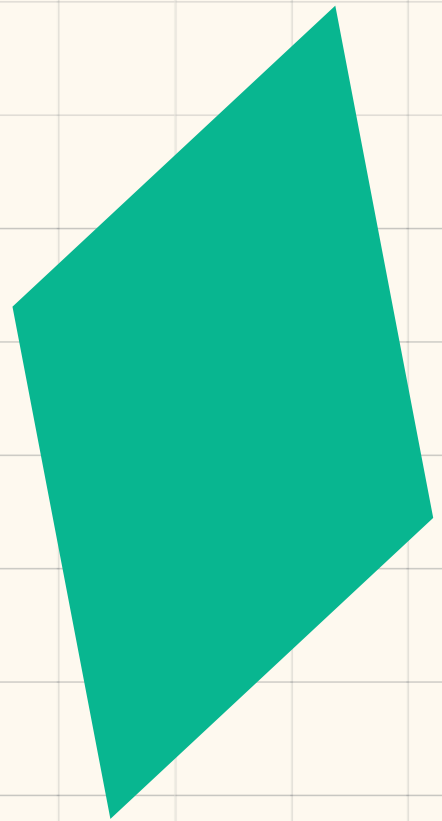
Dada la amplia gama de enfoques de evaluación disponibles, este capítulo describe los métodos más utilizados en las ciencias del comportamiento. Muchas de estas actividades requieren conocimientos técnicos avanzados. Por lo tanto, se recomienda consultar a especialistas en evaluación y utilizar este capítulo para comprender qué conocimientos pueden ser necesarios, qué preguntar a los especialistas en evaluación y qué retos pueden surgir. Para aquellos interesados en realizar evaluaciones directamente, se incluyen manuales adicionales y recursos prácticos a lo largo del capítulo, así como en la sección «Más información» al final de esta fase.

# ¿Por qué poner a prueba las hipótesis?

En esta sección se analizan los cuatro factores siguientes:

1. **La importancia de la evaluación**
2. **El reto de la causalidad**
3. **El marco contrafactual:** comprender qué habría sucedido de otra manera
4. **Cómo la aleatorización crea el estándar de oro para los contrafactuales**

Esta base sienta las bases para seleccionar el método de evaluación adecuado, junto con el uso de evidencia para reforzar las intervenciones.



# 1. La importancia de la evaluación

Imaginemos una situación en la que un paciente está enfermo y un médico le ofrece un nuevo medicamento. Cuando el paciente le pregunta por su eficacia, el médico le responde: «No lo hemos probado formalmente, pero probablemente funcione. Varios pacientes que lo tomaron parecieron mejorar y nuestro equipo confía en él». Pensemos si el paciente tomaría el medicamento. La mayoría de las personas lo rechazarían, y con razón.

Sin embargo, cuando se trata de programas sociales e intervenciones de comportamiento, suele ocurrir lo contrario. Los programas se implementan basándose en buenas intenciones, marcos teóricos prometedores e historias anecdóticas de éxito, pero sin evidencia rigurosa de su impacto. Entonces, ¿por qué las intervenciones sociales se someten a un estándar más bajo que la medicina, cuando ambas tienen como objetivo mejorar el bienestar humano?

## Cuando las historias de éxito engañan

El desarrollo internacional tiene una larga historia de programas bien recibidos que captaron la atención, atrajeron financiación y parecían prometer resultados revolucionarios, hasta que se realizó una evaluación de impacto rigurosa. Uno de los ejemplos más famosos es el de las microfinanzas: aparecida en la década de 1980, fue aclamada como una herramienta transformadora

para la reducción de la pobreza. Al ofrecer pequeños préstamos a personas en entornos de bajos ingresos, las microfinanzas tenían como objetivo fomentar el espíritu emprendedor y el crecimiento económico. El modelo se extendió rápidamente y obtuvo un gran reconocimiento.

Sin embargo, con el tiempo, las evaluaciones rigurosas revelaron una realidad más compleja. Si bien las microfinanzas mejoraron el acceso al crédito, su impacto en la reducción de la pobreza, la movilidad económica y el bienestar a largo plazo era menos claro. Estudios de investigación rigurosos pusieron de relieve el aumento de la carga de la deuda para los prestatarios, la escalabilidad limitada, los modestos resultados empresariales y los escasos avances en la lucha contra la pobreza estructural<sup>1,2,3</sup>. Artículos y libros, como «[Big Money Backs Tiny Loans That Lead to Debt, Despair and Even Suicide](#)» (El gran dinero respalda pequeños préstamos que conducen a la deuda, la desesperación e incluso el suicidio) y [More Than Good Intentions](#) (Más que buenas intenciones), reflejan la desilusión que siguió y el papel crucial que desempeñó la evaluación a la hora de revelar lo que las anécdotas no podían.

Las microfinanzas no son un caso aislado. Otros programas de gran repercusión mediática, como PlayPump<sup>4</sup>, One Laptop Per Child,<sup>5</sup> y el proyecto Millennium Villages<sup>6</sup>, generaron un gran entusiasmo inicial, pero posteriormente su impacto resultó ser menor de lo esperado tras una evaluación minuciosa.

1 John, B. (14 de noviembre de 2024). *Retos y limitaciones de las microfinanzas para lograr la reducción de la pobreza y la creación de empleo a gran escala* [Documento de trabajo].

2 Akbari, M., Nikijoo, I., Khodapanah, B., Foroudi, P. y Padash, H. (2025). *Cuarenta años de investigación sobre microfinanzas y su impacto en los consumidores: una revisión y agenda de investigación utilizando el marco ADO-TCM*. International Journal of Consumer Studies, 49(4), e70101.

3 Blanc, J. (2014). *Microfinanzas, deuda y sobreendeudamiento: malabarismos con el dinero*, Isabelle Guérin, Solène Morvant-Roux y Magdalena Villarreal (dir.). Ediciones Routledge, Londres, Reino Unido, 2014, 316 páginas. Revista internacional de economía social: recma, (334), 122-124.

4 UNICEF. (2007). *Evaluación del sistema de agua PlayPump® como tecnología apropiada para programas de agua, saneamiento e higiene* [https://www-tc.pbs.org/frontlineworld/stories/southernafrica904/flash/pdf/unicef\\_pp\\_report.pdf](https://www-tc.pbs.org/frontlineworld/stories/southernafrica904/flash/pdf/unicef_pp_report.pdf)

5 Cristia, Julian, Ibararan, Pablo, Cueto, Santiago, Ana y Severin, Eugenio, *Tecnología y desarrollo infantil: evidencia del programa One Laptop Per Child (febrero de 2012)*. Documento de trabajo del BID n.º IDB-WP-304, disponible en SSRN: <https://ssrn.com/abstract=2032444>.

6 Mitchell, S., Gelman, A., Ross, R., Chen, J., Bari, S., Huynh, U. K., ... Sachs, J. D. (2018). *El proyecto Millennium Villages: una evaluación retrospectiva, observacional y final*. The Lancet Global Health, 6(5), e500-e513. [https://doi.org/10.1016/S2214-109X\(18\)30065-2](https://doi.org/10.1016/S2214-109X(18)30065-2)

## La trampa de las suposiciones: por qué las buenas intenciones no son suficientes

Incluso con las mejores intenciones, las personas tienden a hacer suposiciones que pueden desviar las intervenciones. Esta «trampa de las suposiciones» opera en múltiples niveles:

- **Suponer que se comprende el problema.** A menudo, los problemas se diagnostican desde perspectivas individuales, en lugar de partir de una comprensión profunda de la experiencia vivida por las comunidades. Lo que parece obvio desde fuera puede carecer o pasar por alto un contexto y una complejidad cruciales.
- **Suponer que se sabe lo que funcionará.** Basándose en su experiencia personal o en experiencias pasadas, las personas pueden llegar a convencerse de que ciertos enfoques tendrán éxito sin evidencia suficiente. Sin embargo, a menudo es posible que aún no se hayan observado los factores que podrían influir en el resultado.
- **Suponer que la implementación se desarrollará según lo previsto.** Es común subestimar los retos prácticos y sobreestimar el grado en que las intervenciones se ajustarán a su diseño cuando se apliquen en entornos reales.
- **Suponer que las anécdotas positivas significan éxito.** Cuando se comparten comentarios favorables o se observan momentos positivos, es común generalizar estas experiencias, dándoles más importancia de la que merecen al evaluar el impacto general.
- **Suponer que la correlación implica causalidad.** Cuando las cosas mejoran después de una intervención, es natural atribuir el cambio a nuestro trabajo, incluso cuando otros factores podrían ser los responsables.

Estas suposiciones no se deben a descuido o incompetencia, sino que son producto del funcionamiento de la cognición humana. Como se ha señalado en módulos anteriores, la mente humana busca patrones, prefiere la evidencia que confirma sus creencias y crea narrativas coherentes, incluso cuando la realidad es más compleja. Si bien estas tendencias son útiles en

muchos contextos, pueden ser engañosas a la hora de evaluar intervenciones sociales complejas.

Sin una evaluación sistemática, estas suposiciones no se cuestionan. Esto puede llevar a invertir en programas que parecen eficaces pero que en realidad no generan un cambio significativo o, lo que es peor, pueden causar daños no deseados. La evaluación proporciona el proceso estructurado necesario para ir más allá de las suposiciones y comprender el verdadero impacto del trabajo.

## La propuesta de valor: por qué vale la pena invertir en la evaluación

Las evaluaciones no son solo ejercicios académicos, sino que aportan un valor concreto:

- **Optimización de recursos:** en entornos con recursos limitados, la evaluación ayuda a dirigir los fondos limitados hacia intervenciones con un impacto demostrado.
- **Corrección del rumbo:** una evaluación oportuna permite identificar y abordar los problemas de implementación antes de escalar, evitando la adopción generalizada de enfoques ineficaces.
- **Confianza de las partes interesadas:** una evaluación rigurosa genera confianza entre los donantes, los gobiernos y las comunidades, lo que facilita las asociaciones y el apoyo a largo plazo.
- **Escalar y replicación:** los programas bien evaluados proporcionan un modelo para la expansión, lo que permite que los enfoques exitosos beneficien a más comunidades.
- **Prevención de daños:** la evaluación puede identificar las consecuencias negativas no deseadas de programas bien intencionados antes de que afecten a grandes poblaciones.

## 2. El reto de la causalidad

### Más allá del «antes y después»

Cuando implementamos un programa y vemos mejoras, es natural suponer que nuestra intervención ha marcado la diferencia. Por ejemplo, se lanza una campaña de vacunación y las tasas de enfermedad disminuyen; se inicia un programa de educación para padres y aumenta la asistencia a la escuela. Estas conexiones parecen obvias, pero pueden ser engañosas.

El reto fundamental de la evaluación es determinar si la intervención *ha causado* realmente los cambios observados o si estos se han producido por otros factores. Esto es más difícil de lo que podría parecer a primera vista.

### Correlación frente a causalidad

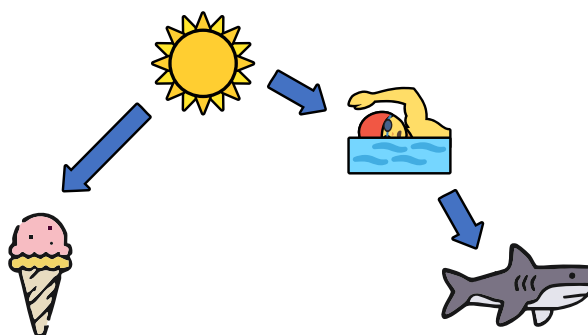
La correlación significa que dos cosas ocurren juntas, mientras que la causalidad significa que una cosa provoca que ocurra la otra. Esta distinción es crucial para la evaluación de impacto de los programas o intervenciones.



Consideremos un ejemplo clásico: los helados y los ataques de tiburones. Los datos muestran que cuando aumentan las ventas de helados, también lo hacen los ataques de tiburones. ¿Las ventas de helados provocan los ataques de tiburones? Por supuesto que no. Esto es lo que los estadísticos denominan «problema de la tercera variable (o confusión por causa común)», cuando un factor oculto influye en ambas variables al mismo tiempo. En este caso, la temperatura es la tercera variable oculta que influye en ambos resultados de forma independiente. Durante los meses de verano.

- Las temperaturas más altas provocan un aumento del consumo de helados.
- Esas mismas temperaturas más altas hacen que más personas se bañen en el mar.
- Un mayor número de nadadores en el agua aumenta la probabilidad de encuentros con tiburones.

Ce phénomène peut être illustré par un diagramme causal simple ou graphe acyclique dirigé. Les flèches bleues représentent les influences causales. Il n'y a pas de flèche reliant les ventes de glaces aux attaques de requins, car il n'existe aucun lien causal direct entre elles. Elles sont corrélées (elles se produisent en même temps), mais non causales (l'une ne provoque pas l'autre).



**Por qué es importante para los programas:** Si no se comprende la diferencia entre correlación y causalidad, se corre el riesgo de sacar conclusiones erróneas, además de diseñar o escalar intervenciones que en realidad no son responsables del cambio observado.

Consideremos un ejemplo de desarrollo en el mundo real. Imaginemos que se implementa un programa de nutrición de UNICEF en varias comunidades. Poco después, los indicadores de crecimiento de los niños comienzan a

mejorar. Podría parecer intuitivo atribuir esta mejora al programa, pero ¿qué más podría estar sucediendo? Consideremos estas explicaciones:

- Quizás sea la temporada de cosecha y la disponibilidad de alimentos haya aumentado de forma natural.
- Quizá otra organización ha comenzado a suministrar agua potable, lo que ha reducido las enfermedades diarreicas.
- O tal vez el gobierno haya implementado una política económica que ha aumentado los ingresos de los hogares al mismo tiempo.

Cualquiera de estos factores podría explicar o contribuir a la mejora observada. Al asumir que el programa provocó el cambio, cuando en realidad se debió a otros factores, se corre el riesgo de invertir o escalar intervenciones que en realidad no funcionan. Peor aún, es posible pasar por alto lo que realmente impulsó el cambio, al tiempo que se descuidan las oportunidades de replicar o reforzar soluciones más eficaces.

Comprender la diferencia entre correlación y causalidad ayuda a evitar estos escollos y fomenta, en cambio, la formulación de mejores preguntas, evaluaciones más inteligentes y decisiones más informadas.

## Factores de confusión: por qué es difícil aislar los efectos

Cuando se implementa una intervención y se miden los resultados, muchos otros factores ajenos al programa pueden influir en los cambios observados. Estos otros factores no relacionados se conocen como variables de confusión. Si no se tienen en cuenta estas variables, se corre el riesgo de atribuir a la intervención efectos que en realidad fueron causados por otra cosa.

Este desafío se denomina endogeneidad: una situación en la que la relación entre una intervención y su resultado se ve distorsionada porque intervienen otras variables. Reconocer esto ayuda a comprender por qué las simples comparaciones antes y después pueden ser engañosas.

A continuación se presentan algunos de los factores de confusión más comunes, ilustrados con ejemplos extraídos de contextos típicos de los programas de UNICEF:

### Factores de confusión basados en el tiempo:

cambios que habrían ocurrido independientemente de nuestra intervención.

- **Variaciones estacionales.** Los indicadores nutricionales mejoran después de que se inicia un programa de alimentación, pero el programa se puso en marcha justo antes de la temporada de cosecha, cuando los alimentos están naturalmente más disponibles.
- **Tendencias preexistentes.** La matriculación escolar aumenta después de una campaña educativa, pero los datos muestran que las tasas ya estaban aumentando de manera constante debido al desarrollo económico a largo plazo.

### Factores de confusión de selección:

Diferencias entre quienes participan en un programa y quienes no lo hacen.

- **Sesgo de autoselección.** Las familias que se unen a un programa de crianza pueden estar ya más comprometidas con el desarrollo de sus hijos, lo que hace que el programa parezca más eficaz de lo que es.
- **Sesgo de selección.** Un programa WASH se dirige a comunidades con altas tasas de enfermedades diarreicas. Incluso sin el programa, estas tasas extremas podrían disminuir con el tiempo simplemente debido a la variación natural.

### Factores de confusión ambientales:

acontecimientos o condiciones externas que se producen al mismo tiempo.

- **Programas concurrentes.** Se lanza una campaña de protección infantil al mismo tiempo que el gobierno comienza a aplicar de forma más estricta las leyes sobre trabajo infantil. Resulta difícil discernir qué iniciativa ha impulsado los cambios observados.
- **Cambios en las políticas.** Se pone en marcha un programa de nutrición infantil justo cuando se introduce un subsidio alimentario nacional. Ambos podrían estar influyendo en la mejora de los indicadores de nutrición.

### Factores de confusión de medición:

cambios en la forma en que se rastrean o detectan los resultados.

- **Mejora del seguimiento.** Tras la introducción de un nuevo sistema de notificación junto con una iniciativa contra la trata de personas, aumenta el número de casos. Esto no se debe a un aumento de la trata, sino a una mejora en el proceso de detección.

## 3. El marco contrafactual: comprender qué habría sucedido de otro modo

### El centro de la inferencia causal se encuentra una pregunta aparentemente sencilla: ¿qué habría pasado si la intervención no se hubiera llevado a cabo?

Este escenario alternativo, en el que el programa no existía, se conoce como contrafactual. Representa el punto de referencia con el que se compara el resultado en el mundo real para determinar si el programa realmente marcó la diferencia.

Pensemos en un niño que recibe una vacuna y no contrae la enfermedad. ¿La vacuna previno la enfermedad o el niño habría seguido estando sano de todos modos? Ambos resultados —ver al niño vacunado y sin vacunar— no pueden observarse a la vez. Este dilema es lo que los estudiosos denominan el «problema fundamental de la inferencia causal». Simplemente, no es posible observar tanto lo real como lo contrafactual para un mismo individuo.

En su lugar, la contrafactual se aproxima buscando o creando un grupo de comparación válido. Este grupo es lo más similar posible al grupo de intervención y experimenta las mismas condiciones externas —como cambios estacionales, variaciones económicas o reformas políticas—, pero no recibe la intervención. Si ambos grupos están expuestos al mismo contexto, cualquier diferencia significativa en los resultados entre ellos puede atribuirse al programa en sí. Esta es la base de un diseño de evaluación creíble. Un contrafactual cuidadosamente construido ayuda a ir más allá de las suposiciones y a responder con confianza a una pregunta importante: ¿La intervención marcó la diferencia o habría ocurrido de todos modos?

Estos ejemplos ilustran por qué es difícil determinar si una intervención fue la causa real de los cambios observados. Cuando múltiples factores influyen simultáneamente en los resultados, ¿cómo se puede aislar el verdadero impacto de un programa?

### Resultados potenciales: una forma formal de pensar en los contrafactuales

Para ayudar a razonar sobre el impacto causal, los estadísticos utilizan lo que se denomina el marco de resultados potenciales. Este proporciona una estructura formal para pensar en la diferencia que supone una intervención mediante la creación de múltiples realidades posibles para cada unidad. Una unidad puede ser una persona, un hogar, una escuela o una comunidad, y para cada unidad hay dos resultados potenciales:

- $Y_1$ : el resultado si la unidad **recibe** el tratamiento/intervención
- $Y_0$ : el resultado si la unidad **no recibe** el tratamiento o la intervención

El efecto causal es la diferencia entre estos dos resultados potenciales:  $Y_1 - Y_0$ .

Aquí está el verdadero desafío: solo se puede observar uno de estos resultados para cada individuo. Cuando un niño recibe una vacuna, es posible observar lo que sucede con la vacunación ( $Y_1$ ). Sin embargo, es imposible observar lo que le habría sucedido a ese mismo niño sin la vacuna ( $Y_0$ ). Esta alternativa no observada, la contrafactual, permanece para siempre desconocida.

Este es el problema fundamental de la inferencia causal: la necesidad de saber tanto lo que sucedió como lo que habría sucedido de otra manera, cuando solo se puede observar una realidad.

Niño A (grupo de tratamiento) → Recibió la intervención → Resultado observado:  $Y_1$

Niño B (grupo de comparación) → NO recibió la intervención → Resultado observado:  $Y_0$

Efecto causal =  $Y_1 - Y_0$

## La solución a los contrafactuales: de los contrafactuales individuales a los grupales

Dado que no se pueden observar ambos resultados para la misma persona (¡todavía no es posible clonar personas ni viajar entre realidades paralelas!), se puede cambiar el enfoque hacia los grupos. Al construir cuidadosamente grupos de comparación muy similares entre sí, es posible aproximarse a lo que habría sucedido con los miembros del grupo que recibieron la intervención y los que no.

En lugar de preguntarnos «¿Este niño en concreto ganó peso gracias al programa?», nos preguntamos «¿Cuánto peso ganan en promedio los niños que participan en el programa de nutrición en comparación con los que no participan?».

Ahora que la atención se centra en los grupos en lugar de en los individuos, es el momento de explorar cómo crear contrafactuales válidos a nivel de grupo. En otras palabras, cómo crear un grupo de comparación que pueda ser similar en tantos factores/variables observables e inobservables como sea posible.

## El reto de construir contrafactuales válidos

Una vez comprendida la necesidad de comparar grupos y no individuos, la siguiente preocupación es cómo crear un grupo que represente con precisión lo que habría sucedido sin la intervención.

La construcción de este grupo de comparación (o contrafactual) es uno de los pasos más importantes y desafiantes de la evaluación causal. Como se ha señalado anteriormente, los factores de confusión pueden distorsionar fácilmente las conclusiones. La calidad de la inferencia causal depende totalmente de hasta qué punto el grupo de comparación seleccionado refleja al

grupo de intervención en todos los aspectos, excepto en uno: no recibieron el tratamiento o la intervención.

Aquí es donde entra en juego el diseño de la evaluación. Existen diferentes enfoques que intentan aproximarse a lo que habría sucedido con nuestro grupo de intervención si no hubiera recibido el tratamiento (el contrafactual). Cada enfoque tiene sus propias ventajas e inconvenientes en cuanto a rigor, viabilidad y riesgo de sesgo. Algunos diseños ofrecen afirmaciones causales más sólidas, pero requieren más control o recursos; otros son más flexibles, pero introducen una mayor incertidumbre.

### 1. Comparación antes y después: contrafactual débil

Un enfoque común pero defectuoso de la evaluación es la comparación antes y después. Este método mide los resultados justo antes de que comience un programa y de nuevo después, atribuyendo cualquier cambio a la intervención. Aunque es sencillo e intuitivo, este enfoque es muy vulnerable a factores de confusión que pueden influir en los resultados a lo largo del tiempo, independientemente del programa en sí. Entre ellos se incluyen:

- **Factores de confusión basados en el tiempo:** variaciones estacionales o tendencias a largo plazo (por ejemplo, un programa de capacitación agrícola muestra un aumento de los rendimientos, pero el periodo de evaluación coincide con la temporada natural de crecimiento).
- **Factores de confusión ambientales:** programas o cambios de política simultáneos (por ejemplo, un programa de nutrición parece tener éxito, pero el gobierno introdujo simultáneamente comidas escolares gratuitas en la misma zona).
- **Factores de confusión relacionados con la medición:** el acto de medir influye en los resultados (por ejemplo, las encuestas repetidas hacen que los hogares sean conscientes del comportamiento «deseado», como lavarse las manos, lo que provoca cambios independientes del programa en sí).

Todos estos factores pueden crear la ilusión de impacto, cuando en realidad el cambio podría haber ocurrido de todos modos.

Por ejemplo, un programa de salud comunitaria puesto en marcha en abril muestra mejores resultados en agosto. Sin embargo, el mes de agosto es también el comienzo de la estación seca, cuando las enfermedades transmitidas por el agua disminuyen de forma natural. Por lo tanto, la mejora observada puede no estar relacionada con la intervención. El enfoque «antes-después» asume esencialmente que «el mismo grupo, en un momento anterior» puede actuar como su propio contrafactual. En entornos dinámicos y reales, esta suposición rara vez se cumple, lo que hace que este diseño tenga una base débil para extraer conclusiones causales.

## 2. Grupo de comparación no equivalente: mejor, pero defectuoso

Un paso adelante con respecto a las comparaciones antes y después es un grupo de comparación no equivalente, es decir, un grupo que no recibe la intervención, pero que se observa durante el mismo periodo de tiempo que el grupo de intervención. Este enfoque ayuda a abordar muchos factores de confusión basados en el tiempo, ya que ambos grupos están expuestos a las mismas condiciones externas (por ejemplo, estaciones, cambios de política o cambios económicos).

Sin embargo, este diseño sigue siendo vulnerable a los factores de confusión de selección: diferencias entre los grupos que pueden afectar a los resultados independientemente de la intervención. Entre ellos se incluyen:

- **Autoselección:** las personas que deciden participar pueden estar ya más motivadas, disponer de más recursos o ser más conscientes de la salud que las que no lo hacen.
- **Selección administrativa:** los programas suelen impartirse intencionadamente en las zonas con mayor necesidad o mayor potencial de éxito, lo que puede sesgar las comparaciones.
- **Diferencias de referencia:** incluso antes de que comience el programa, las comunidades comparadas pueden diferir en aspectos clave, como la infraestructura, los ingresos o la demografía.

Los investigadores suelen intentar emparejar los grupos en función de características observables, pero este enfoque tiene sus limitaciones. Muchos factores importantes (por ejemplo, las actitudes, las aspiraciones, la resiliencia o la genética) no se pueden observar, pero pueden influir en los resultados. Estas diferencias ocultas dificultan atribuir con seguridad los cambios a la intervención.

Aunque es más sólido que el diseño de comparación antes y después, el enfoque de comparación no equivalente sigue sin ser suficiente para producir estimaciones causales de alta confianza, a menos que se apliquen cuidadosamente métodos adicionales (como el ajuste estadístico o los experimentos naturales).

## 4. Cómo la aleatorización crea el estándar de oro para los contrafactuales

### La necesidad de mejores contrafactuales

Como se ha señalado, tanto la comparación antes y después como los diseños de grupos de comparación no equivalente tienen sus limitaciones. Pueden ayudar a observar el cambio, pero les cuesta aislar lo que lo

ha provocado, especialmente cuando hay factores de confusión, tanto conocidos como desconocidos.

Al igual que en el ejemplo del helado y los ataques de tiburones, muchas relaciones del mundo real están determinadas por variables ocultas. En la evaluación de programas, estas variables ocultas suelen ser numerosas, complejas e imposibles de medir en su totalidad.

Por ello, es necesario encontrar una forma de crear grupos de comparación que estén equilibrados en cuanto a las características observadas y no observadas.

Aquí es donde entra en juego el muestreo aleatorio. Al tomar primero un muestreo aleatorio de diferentes grupos (individuos, escuelas o comunidades) de la población objetivo y luego asignar aleatoriamente qué grupos reciben una intervención, se crean grupos estadísticamente equivalentes. En promedio, estos grupos serán similares en todas las características, tanto observadas como no observadas, ya que han sido objeto de muestreo de la misma manera a partir de la misma población subyacente. Esto significa que factores como la motivación, el estado de salud inicial, los ingresos, las creencias no medidas o las normas comunitarias estarán equilibrados de manera aleatoria.

Cuando se aplica correctamente, la aleatorización garantiza que la única diferencia entre los grupos sea si han recibido o no la intervención. Este enfoque hace que sea mucho más probable que cualquier diferencia en los resultados se deba al programa en sí, y no a diferencias externas o preexistentes. Por esta razón, la aleatorización se conoce como el «estándar de oro» en la inferencia causal. Cuando se realiza correctamente y con un tamaño de muestra suficientemente grande, este enfoque proporciona la mayor confianza posible en que la intervención ha provocado el cambio observado.

## La aleatorización como enfoque para crear contrafactuales

Al asignar aleatoriamente qué unidades (es decir, individuos, hogares, escuelas o comunidades) reciben una intervención, se reduce significativamente la influencia de los factores de confusión. Todos los tipos de factores de confusión mencionados anteriormente se distribuyen ahora de manera uniforme entre los grupos por diseño. En la práctica, esto significa:

- Las variaciones estacionales afectan a ambos grupos por igual.

- Se elimina el sesgo de autoselección, ya que la participación se asigna, no se elige
- Las tendencias preexistentes se desarrollan de manera similar en ambos grupos.
- Los programas o cambios de política concurrentes afectan a ambos grupos al mismo tiempo.
- Los efectos relacionados con la medición se aplican por igual en todos los grupos.

Como resultado, la única diferencia sistemática entre los grupos es si reciben o no la intervención. Esto permite atribuir cualquier diferencia en los resultados directamente a la intervención en sí, en lugar de a cualquier otra influencia oculta o externa.

Para dar un ejemplo, imagina que se lleva a cabo un programa para animar a los padres a enviar a sus hijos a la escuela. Hay un gran grupo de padres que cumplen los requisitos, pero se trata de un grupo diverso, en el que cada uno está condicionado por diferentes factores:

- Algunos son ricos, otros se enfrentan a dificultades económicas.
- Algunos viven cerca de la escuela, otros viven lejos
- Algunos han completado la educación superior, otros tienen poca formación académica
- Algunos tienen trabajos flexibles, otros horarios rígidos
- Algunos valoran mucho la educación, otros son escépticos
- Algunos tuvieron buenas experiencias escolares, otros no
- Algunos están muy motivados, otros menos

Al asignar aleatoriamente a los padres para que reciban la intervención o formen parte del grupo de control (es decir, no reciban la intervención), se garantiza que todas las características se distribuyan de manera similar entre los grupos. Esto incluye tanto características observables, como los ingresos y la distancia, como otras no observables, como las creencias y la motivación.

La aleatorización garantiza que haya una distribución equitativa de estos diferentes rasgos entre los dos grupos. En este ejemplo, cualquier diferencia en las tasas de asistencia escolar después de la intervención puede atribuirse a la intervención en sí, y no a diferencias preexistentes entre los padres que la recibieron y los que no.

## El camino de las buenas intenciones al impacto real

Una evaluación rigurosa no solo tiene que ver con la credibilidad académica, sino también con garantizar que los programas realmente mejoren la vida de las personas. Como se ha señalado anteriormente, las intervenciones bien intencionadas pueden no tener el impacto deseado, desperdiciar recursos valiosos o incluso causar daños no deseados cuando se basan en suposiciones en lugar de en evidencia. La diferencia entre correlación y causalidad es importante, ya que determina si escalar lo que realmente funciona o si se está invirtiendo en programas que simplemente coincidieron con un cambio positivo. Al construir contrafactuales válidos, idealmente mediante la aleatorización, la comprensión va más allá de lo que *parece* funcionar y se centra en lo que realmente funciona, para quién y por qué.

Este conocimiento transforma la forma de diseñar programas, asignar recursos y, en última instancia, servir a las comunidades. Si bien una evaluación rigurosa puede parecer desalentadora, continuar con las intervenciones sin conocer su verdadero impacto es mucho más arriesgado.

# Preocupaciones comunes sobre la aleatorización y respuestas

Si bien la aleatorización es uno de los enfoques más rigurosos para abordar el desafío de la causalidad y los contrafactuales, tiende a generar inquietudes. Estas pueden estructurarse en torno a seis categorías diferentes:

## 1. Costo y recursos

**PREOCUPACIÓN:** Las evaluaciones aleatorias son caras y requieren muchos recursos.

**RESPUESTA:**

- Si bien las evaluaciones rigurosas requieren una inversión, el costo debe sopesarse con el valor de la evidencia confiable.
- No todos los ECA tienen que ser a gran escala o costosos: los estudios pequeños y específicos pueden ser rentables.
- El costo de implementar programas ineficaces al escalar es, en última instancia, mucho mayor que los costos de evaluación.
- Las fuentes de datos existentes y los diseños inteligentes pueden, en ocasiones, reducir los costos de manera sustancial.

## 2. Limitaciones de tiempo

**PREOCUPACIÓN:** Las evaluaciones aleatorias llevan demasiado tiempo y UNICEF necesita responder con rapidez.

**RESPUESTA:**

- Las pruebas de ciclo rápido, un proceso estructurado que consiste en probar intervenciones a pequeña escala, medir rápidamente los resultados y repetir el proceso basándose en los comentarios recibidos para perfeccionar las soluciones, pueden proporcionar perspectivas en plazos más cortos.
- La implementación por fases permite tanto la acción inmediata como la evaluación rigurosa.
- El tiempo invertido en la evaluación evita perder años en enfoques ineficaces.
- Algunos resultados pueden medirse a corto plazo (por ejemplo, la adherencia a la medicación, como los antirretrovirales), mientras que otros requieren un seguimiento durante periodos más largos (por ejemplo, la supresión de la carga viral).

### 3. Cuestiones éticas

**PREOCUPACIÓN:** No es ético privar a los grupos de control de programas potencialmente beneficiosos.

**RESPUESTA:**

- Cuando los recursos son limitados, la aleatorización suele ser el método de asignación más justo.
- No hay forma de saber si los programas funcionan sin probarlos, algunos pueden no tener ningún efecto o incluso causar daño.
- La implementación por fases garantiza que todos los participantes reciban finalmente el programa si se demuestra su eficacia.
- La obligación ética de garantizar que los programas realmente ayuden a los niños justifica el proceso de pruebas rigurosas.

### 4. Desafíos políticos y de las partes interesadas

**PREOCUPACIÓN:** Los socios gubernamentales o las comunidades pueden resistirse a la aleatorización.

**RESPUESTA:**

- El encuadre es importante: hay que hacer hincapié en las ventajas de este tipo de evaluación, incluidas las opciones para que sea económica o tenga una duración limitada.
- Involucre a las partes interesadas desde el principio en el proceso de diseño para abordar las preocupaciones y fomentar la apropiación.
- Cree procedimientos claros para detener los ensayos si surge evidencia sólida de sus beneficios.
- Explique que la evaluación refuerza la promoción de los programas exitosos.

### 5. Relevancia contextual y generalización

**PREOCUPACIÓN:** los resultados de un contexto no se aplicarán a otros en los que trabaja UNICEF.

**RESPUESTA:**

- La selección estratégica de los sitios puede mejorar la generalización.

- La medición de los factores de implementación ayuda a comprender qué elementos contextuales son importantes.
- Incluso la evidencia localizada es mejor que la ausencia de evidencia
- Las evaluaciones múltiples en distintos contextos pueden crear una base empírica más amplia.

### 6. Capacidad técnica

**PREOCUPACIÓN:** El personal de UNICEF puede carecer de los conocimientos técnicos necesarios para diseñar y analizar evaluaciones aleatorias.

**RESPUESTA:**

- Las asociaciones con instituciones académicas pueden complementar la capacidad interna
- La inversión en la capacitación del personal desarrolla la capacidad organizativa a largo plazo
- Los diseños aleatorios simples pueden ser más accesibles que los métodos cuasi experimentales complejos
- Los especialistas en evaluación de UNICEF pueden proporcionar apoyo técnico en todos los programas.

# Consideraciones de diseño para la evaluación de impacto

Una vez establecido por qué es importante la evaluación y cómo se puede identificar de manera creíble la causalidad, la atención se centra ahora en diseñar evaluaciones que planteen las preguntas adecuadas, midan los resultados correctos y generen evidencia que sirvan de base directa para la toma de decisiones. En la siguiente sección se describen las consideraciones clave para el diseño de evaluaciones de impacto rigurosas.

## 1. Definición de las preguntas de investigación

Toda evaluación comienza con un reto sencillo: aclarar las preguntas clave que deben responderse. Comience por reunir a las partes interesadas clave y pregúnteles:

- ¿Qué decisiones dependen de esta evaluación?
- ¿Qué se haría de manera diferente si se supiera X, en comparación con Y?

Céntrese en 3-5 preguntas fundamentales que informen directamente sobre acciones concretas. Estas suelen dividirse en tres categorías

- **Eficacia:** ¿Funciona la intervención? ¿Cuánto tiempo tarda en funcionar? ¿Cuándo se observan los efectos?
- **Mecanismo:** ¿Cómo funciona la intervención?
- **Destinatarios:** ¿Para quién funciona mejor?

Para cada pregunta, especifique cómo la respuesta orientará las decisiones. Si un programa de formación profesional aumenta el empleo en un 10 %, ¿se escalará? ¿Y si solo es un 5 %? ¿Y si funciona para los hombres, pero no para las mujeres? Predefinir estos umbrales de decisión evita cualquier racionalización a posteriori y garantiza que la evaluación genere evidencia útil. Recuerde tener en cuenta tanto los escenarios positivos como los nulos: saber que algo no funciona es igualmente valioso para la asignación de recursos.

Una vez definidas las preguntas clave de la investigación, el siguiente paso es determinar *cómo* se responderán esas preguntas. Para ello, es necesario traducir los objetivos generales en resultados precisos y medibles, y seleccionar los indicadores adecuados que reflejen el cambio real.

## 2. Selección de los resultados primarios y la estrategia de medición

### Comenzar con preguntas precisas

El éxito de la evaluación depende de medir los resultados de la manera correcta. Esto comienza con la transformación de objetivos generales en preguntas de evaluación articuladas con precisión. Una pregunta vaga como «¿Funciona este programa de nutrición?» conduce a respuestas vagas y decisiones poco claras. En cambio, compárese con una pregunta como «¿Proporcionar asesoramiento nutricional mensual aumenta las puntuaciones z de altura para la edad en al menos 0,2 desviaciones estándar entre los niños de 6 a 24 meses de comunidades rurales en un plazo de 12 meses?».

En otras palabras: «¿La **[intervención]** conduce a **[un cambio específico medible, con un umbral si es relevante]** en **[una población definida]** dentro de **[un plazo y un contexto]**?». Este nivel de precisión guía todas las decisiones de medición posteriores y garantiza que los resultados sean interpretables y aplicables.

Una vez formuladas las preguntas de evaluación, el siguiente paso es identificar las medidas de resultados que pueden ayudar a responder a las preguntas.

## 3. Elegir qué medir

Elija resultados que muestren claramente lo que el programa intenta lograr. Céntrese en medidas que indiquen directamente si la intervención está logrando el cambio deseado. Estas deben ser lo suficientemente específicas como para medir con precisión y lo suficientemente importantes como para orientar las decisiones del programa.

Considere la proximidad de los resultados a la intervención. Los resultados proximales, aquellos que se producen poco después de la intervención, son más fáciles de cambiar y medir, pero pueden no reflejar el objetivo final. Los resultados distales, aquellos que se encuentran más abajo en la cadena de resultados,

muestran el impacto real, pero suelen requerir un muestreo más amplio y más tiempo para detectarlos.

Por ejemplo, un programa de alimentación escolar podría aumentar rápidamente la asistencia (proximal), pero requeriría datos mucho más amplios y más tiempo para mostrar una mejora en el aprendizaje (distal). Ambos resultados son útiles: los resultados a corto plazo muestran si el programa va por buen camino, mientras que los resultados a largo plazo muestran si realmente está marcando una diferencia.

Una vez que los resultados están claramente definidos, es esencial considerar cómo se medirán. Las diferentes fuentes de datos varían en cuanto a precisión, costo y viabilidad, y reconocer estas compensaciones garantiza que las decisiones de medición refuercen, en lugar de comprometer, la credibilidad de la evaluación.

## Métodos de recopilación de datos y compensaciones

### Ventajas y limitaciones distintivas de las fuentes de datos:

**Los datos de las encuestas** ofrecen flexibilidad para medir exactamente lo que se necesita, pero también plantean varios retos. Los comportamientos autoinformados suelen adolecer de un sesgo de deshabilitación social. Por ejemplo, los padres suelen exagerar o minimizar las medidas disciplinarias severas de las vacunas. Los periodos de referencia son muy importantes; preguntar por acontecimientos de la semana pasada da resultados diferentes en cuanto a precisión que preguntar por los del año pasado. El cansancio por las encuestas también puede comprometer la calidad de los datos en cuestionarios largos. Puede ser más difícil llegar a los grupos marginados a través de encuestas telefónicas, y las personas de comunidades con bajo nivel de alfabetización pueden tener dificultades para comprender ciertas preguntas. Asegúrese de que se realicen entrevistas cognitivas durante la fase piloto para garantizar que las preguntas sean culturalmente apropiadas y se comprendan según lo previsto.

**Los datos administrativos**, como los expedientes escolares, los registros clínicos o las bases de datos de programas, ofrecen una medición objetiva y rentable, pero tienen limitaciones. Los esfuerzos se limitan a lo que ya se ha recopilado, lo que puede no coincidir perfectamente con los resultados de interés. Los datos sobre determinados grupos étnicos o poblaciones marginadas pueden estar incompletos o faltar. La calidad de los datos también varía enormemente: algunas clínicas mantienen registros meticulosos, mientras que otras apenas funcionan. El uso

de datos administrativos a menudo obliga a que la unidad de aleatorización se ajuste a los niveles administrativos de las escuelas o clínicas, en lugar de a las personas.

### La observación directa y las medidas de comportamiento

proporcionan una evaluación objetiva, pero requieren una implementación cuidadosa. Los observadores necesitan una amplia capacitación para garantizar la coherencia. Es importante acercarse a las comunidades con sensibilidad, solicitar el consentimiento con cuidado y evitar invadir espacios privados o culturalmente sensibles. La tecnología permite cada vez más realizar mediciones discretas (seguimiento por GPS, datos de sensores), pero puede que no sea viable en todos los contextos sin necesidad de equipos costosos.

L'identification de sources de données appropriées est indispensable, mais insuffisante à elle seule ; le moment et la fréquence de la collecte sont tout aussi déterminants. Une collecte mal synchronisée peut masquer des effets réels ou déformer l'évaluation des performances d'un programme.

### Los biomarcadores y las medidas antropométricas

ofrecen objetividad para las intervenciones sanitarias, pero requieren formación y equipos especializados. ¿Se pueden mantener las cadenas de frío para las muestras de sangre? ¿Los participantes darán su consentimiento para someterse a procedimientos invasivos? ¿Cómo se gestionarán los errores de medición de los diferentes evaluadores, al tiempo que se garantiza que la recopilación de datos sea respetuosa y suponga una carga mínima para los participantes?

Identificar las fuentes de datos adecuadas es necesario, pero no suficiente por sí solo; el momento y la frecuencia de la medición son igualmente críticos. Una recopilación de datos en un momento inadecuado puede ocultar los efectos reales o tergiversar el rendimiento del programa.

## 4. Momento y frecuencia de la medición

Los resultados se producen en diferentes escalas de tiempo. Los conocimientos pueden cambiar en cuestión de semanas, los comportamientos en meses y los efectos sobre la salud en años. Medir demasiado pronto conlleva el riesgo de encontrar efectos nulos en intervenciones que necesitan tiempo para surtir efecto. Medir demasiado tarde conlleva el riesgo de pasar por alto efectos que se desvanecen o se ven contaminados por otros factores.

Considere múltiples puntos de medición para comprender la dinámica de los efectos. ¿El impacto crece, se estabiliza o disminuye con el tiempo? Un impulso inicial que se desvanece podría sugerir la necesidad de un refuerzo; los efectos de desvanecimiento son comunes en las intervenciones de cambio de comportamiento. Un cambio gradual puede sugerir que los efectos se acumulan con el tiempo y, potencialmente, a partir de diferentes procesos. Lo ideal es intentar presupuestar al menos un seguimiento más allá de la intervención inmediata (por ejemplo, seis meses después) para evaluar la persistencia.

Los patrones estacionales pueden confundir los resultados si no se tienen en cuenta cuidadosamente. Por ejemplo, los resultados agrícolas varían según los ciclos de cosecha, los patrones de enfermedades siguen las tendencias estacionales y las medidas escolares fluctúan con los calendarios académicos. Mida el tiempo para evitar confundir los efectos del programa con la variación estacional, o asegúrese de que tanto el grupo de tratamiento como el de control se miden simultáneamente. Además, intente incluir la recopilación de datos en todos ellos para comprender cómo interactúa el efecto con estas variaciones estacionales.

El momento indica cuándo se produce el cambio, mientras que los mecanismos explican por qué. La medición a lo largo de la cadena causal revela cómo se desarrolla el impacto, dónde fallan los sistemas y cómo se pueden perfeccionar los programas.

## 5. Medición de los mecanismos a lo largo de la cadena causa

Comprender por qué los programas funcionan (o no) es tan importante como saber si funcionan. La medición de los mecanismos tiene múltiples propósitos:

**Validación de la teoría:** ¿funcionan realmente las vías hipotéticas? Un programa de lavado de manos asume lo siguiente: información -> conocimiento -> cambio de actitud -> cambio de comportamiento -> mejora de la salud. La medición de cada paso valida o cuestiona estas suposiciones.

**Diagnóstico de fallos:** cuando los resultados no mejoran, los mecanismos revelan dónde se rompieron las cadenas. ¿Los maestros no recibieron capacitación? ¿La recibieron pero no la entendieron? ¿La entendieron

pero no la implementaron? ¿La implementaron pero los estudiantes no se involucraron? Cada punto de ruptura sugiere diferentes soluciones.

**Perfeccionamiento del programa:** en lugar de abandonar por completo los programas «fallidos», los datos sobre los mecanismos permiten aplicar soluciones específicas. Si los padres recibieron información sobre nutrición pero carecían de recursos para comprar alimentos diversos, la incorporación de vales podría tener un impacto positivo.

Generalización: Los programas que funcionan a través de mecanismos universales, como los recordatorios, probablemente se transfieran mejor

entre contextos que aquellos que dependen de características institucionales específicas.

No se limite a medir los resultados finales, haga un seguimiento de los pasos intermedios. En el caso de un programa de nutrición destinado a reducir la malnutrición infantil, esto podría traducirse en: conocimientos de los cuidadores (inmediatos), prácticas de alimentación (a corto plazo), diversidad alimentaria de los niños (a medio plazo) y estado nutricional (a largo

plazo). Cada uno de ellos proporciona información valiosa sobre el funcionamiento del programa.

Comprender los mecanismos revela cómo se produce el cambio, pero para interpretar esos patrones con precisión, es importante asegurarse de que los datos reflejen fielmente a las poblaciones a las que se atiende. La representatividad y la inclusión en la medición son esenciales para generar evidencia que refleje realidades diversas, y no solo las más fáciles de alcanzar

## 6. Garantizar una medición representativa e inclusiva

A quién se mide es tan importante como qué se mide. Las encuestas escolares excluyen sistemáticamente a los niños que no asisten a la escuela, que suelen ser los más vulnerables. Las encuestas telefónicas excluyen a quienes no tienen teléfono, y los datos de las clínicas no incluyen a quienes no buscan atención médica o no pueden hacerlo.

Elabore estrategias de medición inclusivas desde el principio, asegurándose de que las comunidades contribuyan a definir los indicadores y las fuentes de datos. Utilice múltiples fuentes de datos para captar diferentes poblaciones. Realice un sobremuestreo de los grupos marginados para garantizar una representación adecuada. Traduzca los instrumentos a los idiomas locales y realice pruebas piloto con diversos encuestados. Capacite a los encuestadores de las comunidades que se están estudiando para mejorar la relación y la comprensión.

Tenga en cuenta qué perspectiva se está captando. Los niños, los padres, los maestros y los trabajadores sanitarios pueden informar de manera diferente sobre un mismo fenómeno. Incluso dentro del mismo grupo (por ejemplo, las madres), las experiencias de las personas pueden variar en función de su identidad social, sus antecedentes o su posición dentro de la comunidad. La triangulación entre los informantes puede revelar dinámicas importantes, pero requiere protocolos claros para manejar las discrepancias.

Incluso los instrumentos diseñados con mayor cuidado pueden resultar insuficientes si no se pueden aplicar de manera eficaz en entornos cotidianos reales. Es esencial garantizar que los métodos de medición sean viables, fiables y aptos para su uso sobre el terreno.

## 7. Consideraciones prácticas sobre la medición

Durante el proceso de medición, tenga en cuenta estas consideraciones:

**Realice pruebas piloto exhaustivas.** Nunca dé por sentado que los instrumentos que funcionan en otros lugares también lo harán en otro contexto. Realice pruebas piloto con un grupo reducido de encuestados que imiten las condiciones reales sobre el terreno. Pruebe los patrones de omisión, la sincronización, la traducción y la comprensión. Informe detalladamente a los encuestadores, ya que a menudo detectan problemas que los encuestados no mencionan.

**Equilibre la exhaustividad con la viabilidad.** Los instrumentos largos proporcionan datos ricos, pero

provocan fatiga en los encuestados, aumentan los costos y degradan la calidad. La mayoría de los impactos se pueden detectar con instrumentos específicos. Reserve las mediciones largas para estudios de mecanismos a pequeña escala, en lugar de para evaluación de impacto a gran escala.

**Prevea los errores de medición.** Toda medición contiene errores: la evaluación antropométrica varía entre los evaluadores y las puntuaciones de las pruebas dependen de las condiciones en que se realizan. Incorpore controles de calidad, como ejercicios de estandarización para los evaluadores, medidas repetidas en submuestras y validación con fuentes externas cuando sea posible.

**Documento todo.** Cree protocolos detallados que especifiquen exactamente cómo se mide, codifica y construye cada resultado. Los futuros investigadores deben comprender y, potencialmente, replicar las medidas elegidas. Incluya los instrumentos de encuesta, los materiales de capacitación y el código de construcción de variables en los apéndices.

Las prácticas de medición sólidas garantizan la calidad de los datos, pero su valor depende de la capacidad del estudio para detectar efectos reales. Un poder estadístico adecuado protege contra conclusiones falsas, garantizando que los datos no solo describan lo que se observó, sino que revelen lo que realmente funcionó.

## 8. Determinación del tamaño de la muestra y el poder estadístico

El poder estadístico es, en esencia, la capacidad de la evaluación para detectar un efecto real cuando existe, como tener un radar lo suficientemente sensible como para detectar un avión que se aproxima. Por el contrario, un estudio con poca potencia es como buscar algo con una linterna tenue, incapaz de encontrar nada incluso cuando está ahí. Esto cobra especial importancia a la hora de determinar cuántos participantes se necesitan. Si la muestra es demasiado pequeña, se corre el riesgo de concluir que el programa no ha tenido ningún efecto cuando, en realidad, sí lo ha tenido (lo que se conoce como «falso negativo»). Sin embargo, recopilar datos de más participantes de los necesarios supone un desperdicio de recursos. Los factores clave que afectan a la potencia son, entre otros, los siguientes:

- El tamaño previsto de los efectos del programa: los efectos más grandes son más fáciles de detectar.

- La cantidad de variación natural que existe en la medida de los resultados: una mayor variación requiere muestras más grandes.
- La unidad de aleatorización elegida: la aleatorización a nivel comunitario requiere más esfuerzo que la aleatorización de individuos.

Incorporar supuestos realistas sobre el Pérdida de unidades de estudio (lo habitual es entre un 10 % y un 20 %), el incumplimiento (el grupo de tratamiento no recibe la intervención) y la contaminación (el grupo de control accede a la intervención). Cada uno de ellos reduce el tamaño efectivo de la muestra; es mejor reclutar un 20 % más de participantes que descubrir que el estudio tiene una potencia insuficiente después de la recopilación de datos.

## 9. Estructura de la aleatorización

A la hora de probar una intervención, es fundamental elegir el nivel en el que se llevará a cabo la aleatorización. Esto significa decidir si se asigna la intervención a individuos, grupos, escuelas, comunidades u otra unidad que tenga sentido para el proyecto. El nivel elegido debe coincidir con la forma en que se lleva a cabo la intervención en la vida real. Por ejemplo, si un nuevo plan de estudios solo puede introducirse en toda una clase en lugar de en alumnos individuales, la clase se convierte en la unidad más práctica para la aleatorización.

El nivel de aleatorización también depende del tipo de información que se pueda recopilar. Si los datos solo se pueden medir a nivel de grupo,

como el gasto doméstico o la asistencia escolar, tiene sentido aleatorizar a ese mismo nivel.

TABLA 8. DIFERENTES TIPOS DE UNIDADES DE ALEATORIZACIÓN

UNIDAD DE ALEATORIZACIÓN	VENTAJAS	CONSIDERACIONES	CUÁNDO UTILIZARLA	EJEMPLO
<b>Aleatorización a nivel individual</b>	Requiere un tamaño de muestra más pequeño en comparación con otros niveles; alta eficiencia estadística.	Riesgo de efectos secundarios cuando los individuos interactúan; retos logísticos a la hora de aplicar diferentes intervenciones en el mismo entorno (por ejemplo, comunidad o aula).	Adecuado cuando la interacción entre individuos es mínima y la intervención puede dirigirse fácilmente a individuos específicos.	Recordatorios de vacunación por SMS enviados a cuidadores seleccionados al azar dentro de un gran distrito urbano.
<b>Aleatorización a nivel de hogar</b>	Captura la toma de decisiones a nivel de hogar; se ajusta a la forma en que se configuran muchos comportamientos y resultados.	Los efectos secundarios siguen siendo posibles cuando los vecinos interactúan; se requiere una definición clara de «hogar»; es posible que el análisis deba tener en cuenta las diferencias en el tamaño de los hogares.	Adecuado cuando las intervenciones afectan o involucran a todos los miembros de un hogar (por ejemplo, visitas domiciliarias, transferencias monetarias condicionadas).	Transferencias monetarias condicionadas proporcionadas a hogares asignados al azar con niños menores de cinco años.
<b>Aleatorización a nivel de comunidad o pueblo (aleatorización por conglomerados)</b>	Reduce el riesgo de contaminación o efectos secundarios; a menudo es más fácil de gestionar desde el punto de vista político y operativo.	Requiere más grupos (comunidades) para alcanzar el poder estadístico; la alta variación entre comunidades aumenta la dispersión; la logística de implementación es más compleja a gran escala.	Útil cuando es probable que los individuos dentro de las comunidades se influyan entre sí o cuando las intervenciones se llevan a cabo públicamente (por ejemplo, movilización comunitaria).	Campañas de inmunización dirigidas por trabajadores sanitarios comunitarios probadas en aldeas asignadas al azar.

UNIDAD DE ALEATORIZACIÓN	VENTAJAS	CONSIDERACIONES	CUÁNDO UTILIZARLA	EJEMPLO
<b>Aleatorización a nivel de centro (por ejemplo, escuelas, clínicas)</b>	Práctico para entornos de prestación institucional; se ajusta a las estructuras organizativas existentes; adecuado para intervenciones a nivel del personal o de todo el centro.	Las instalaciones pueden variar en tamaño, calidad o rotación de personal; la superposición de las zonas de influencia puede dar lugar a efectos secundarios; el poder estadístico está limitado por el número de instalaciones disponibles.	Adecuado para evaluar las intervenciones realizadas a través de instituciones, especialmente cuando no es práctico dirigirse a nivel individual.	Formación en comunicación interpersonal impartida al personal de clínicas de salud seleccionadas al azar.

## 10. Cuando no es posible la aleatorización

A veces la aleatorización no es factible debido a limitaciones políticas, éticas o prácticas. Las partes interesadas pueden considerar que la asignación aleatoria es injusta, los programas pueden estar ya en marcha o el tamaño de las muestras puede ser demasiado pequeño para una aleatorización significativa. En estos casos, los métodos cuasi-experimentales intentan crear comparaciones válidas utilizando técnicas estadísticas para aproximarse al contrafactual que habría proporcionado la aleatorización.

Estos enfoques funcionan identificando y controlando los factores que influyen tanto en la participación en el programa como en los resultados, lo que a veces se conoce como «cerrar la ruta trasera». Si bien estos métodos pueden proporcionar evidencia valiosa, requieren hipótesis más sólidas sobre los datos y el contexto. Por lo general, necesitan muestras más grandes, una recopilación de datos más amplia y un análisis más complejo que los ECA. Lo más importante es que siguen siendo vulnerables al sesgo de factores no medidos que la aleatorización habría eliminado.

Para obtener más orientación sobre los enfoques cuasi-experimentales, véase el apéndice 1. Solo se deben utilizar estos métodos con la orientación de expertos, ya que su validez depende en gran medida de supuestos específicos del contexto que a menudo no se pueden comprobar.

# Lista de verificación práctica para la implementación de evaluaciones de impacto

## Protocolo de evaluación

Antes de desarrollar un protocolo de evaluación, es útil esbozar primero los objetivos de aprendizaje, utilizando una herramienta como la [agenda de aprendizaje](#). Esta herramienta ayuda a definir y organizar las preguntas clave que deben responderse en una evaluación de impacto, y su uso se ilustra en el Apéndice 2 a través del **estudio de caso sobre el aumento de la cobertura de vacunación infantil en el Líbano**.

Una vez que se ha establecido la agenda de aprendizaje, se puede elaborar el protocolo de evaluación. El protocolo de evaluación proporciona una guía y un plan estructurados sobre cómo se evaluará la intervención. El desarrollo de un protocolo de evaluación transforma las decisiones de diseño en un documento técnico completo que guía la implementación y el análisis. Considérelo como un contrato con su yo futuro, que evita el análisis selectivo y garantiza la integridad científica. Un protocolo sólido especifica de antemano todas las decisiones analíticas antes de ver los resultados, lo que protege contra los sesgos conscientes e inconscientes hacia la búsqueda de resultados positivos.

Los componentes esenciales del protocolo incluyen:

- Descripción detallada de la intervención: qué se llevará a cabo exactamente, quién lo hará y con qué frecuencia.
- Teoría de cambio con vías causales claras.
- Preguntas de evaluación asignadas a hipótesis específicas.
- Cálculos de potencia con todas las suposiciones explícitas.

- Definiciones precisas de los resultados con preguntas exactas de las encuestas (si las medidas de los resultados se basan en las encuestas).
- Procedimientos de aleatorización, incluidas variables de estratificación.
- Modelos analíticos con ecuaciones de regresión específicas
- Listas de covariables determinadas por la teoría, no por los datos
- Análisis de subgrupos con una justificación clara
- Procedimientos para manejar la pérdida de unidades de estudio y el incumplimiento
- Comprobaciones de solidez para evaluar la sensibilidad de los resultados

Es especialmente importante especificar previamente el plan de análisis en los siguientes casos:

- resultados primarios frente a secundarios, para evitar el cambio de resultados
- análisis de subgrupos, para evitar la búsqueda de efectos significativos
- criterios de inclusión/exclusión para la muestra de análisis
- manejo de valores atípicos y datos faltantes
- ajustes por pruebas múltiples

Cualquier desviación del protocolo debe indicarse claramente como exploratoria en los informes. Como referencia, véase el **estudio de caso en el Líbano**, donde está disponible [el protocolo de evaluación completo](#).

Considere la posibilidad de registrar el protocolo en repositorios públicos (AEA Social Science Registry, RIDIE, ClinicalTrials.gov) antes de comenzar la recopilación de datos. El registro proporciona una marca de tiempo,

## Plan de implementación

Se trata de una hoja de ruta operativa detallada que traduce el diseño de la evaluación en acciones reales. El plan toma el protocolo técnico y lo convierte en una guía de implementación diaria, especificando quién hará qué, cuándo, dónde y con qué recursos, a lo largo de todo el ciclo de vida de la evaluación. El plan incluye lo siguiente:

- cronogramas detallados con fechas e hitos específicos
- asignación clara de funciones con personas responsables de cada tarea
- requisitos de recursos (tiempo del personal, materiales, transporte, tecnología)
- indicadores de seguimiento de la ejecución
- planes de contingencia para problemas comunes
- protocolos de comunicación entre los miembros del equipo y los socios

La herramienta [del plan de implementación](#) puede ser una plantilla sencilla y útil para desarrollar una guía paso a paso sobre cómo se llevará a cabo la intervención. En el apéndice 2 se muestra un ejemplo de cómo se puede utilizar esta herramienta.

### Por qué es importante

El vacío entre un diseño de evaluación brillante y una ejecución fallida suele ser una mala planificación de la implementación. Incluso los diseños de investigación más rigurosos fracasan cuando las tabletas no están cargadas, las encuestas no están traducidas correctamente o los equipos no saben quién es responsable de la selección de los participantes. El plan de implementación evita los miles de pequeños fallos que pueden invalidar una evaluación que, por lo demás, está bien diseñada. Garantiza la coordinación entre los múltiples actores

lo que demuestra la especificación previa y permite a la comunidad investigadora realizar un seguimiento de todos los estudios, no solo de los éxitos publicados. Incluya suficientes detalles para que otro investigador pueda replicar la evaluación, manteniendo al mismo tiempo la flexibilidad operativa para las adaptaciones de campo necesarias que no comprometan el diseño básico.

(equipo de investigación, socios de implementación, funcionarios gubernamentales, líderes comunitarios) que pueden tener diferentes prioridades y estilos de trabajo.

Sin una planificación operativa clara, se corre el riesgo de descubrir demasiado tarde que las campañas de vacunación entran en conflicto con la recopilación de datos, que el personal clave no está disponible durante los periodos cruciales o que los materiales no se han impreso a tiempo. El plan también proporciona una herramienta de gestión para mantener las operaciones complejas por el buen camino e identificar los problemas antes de que se conviertan en un fracaso de la evaluación.

Tenga en cuenta estas consideraciones clave:

- **Planifique en función de las limitaciones fijas.** Trabaje hacia atrás a partir de plazos fijos e inamovibles, como las temporadas agrícolas, los años escolares, los ciclos presupuestarios o las fiestas religiosas, para establecer plazos realistas, y luego añada un margen de tiempo del 20-30 % para los retrasos inevitables. Si se estima que la recopilación de datos llevará tres semanas, planifique cuatro.
- **Asigne una responsabilidad clara y única para cada actividad.** Evite la responsabilidad compartida, que a menudo significa ninguna responsabilidad. Cuando se indique que «el equipo llevará a cabo la capacitación», especifique exactamente quién la dirige, quién la asiste y quién es responsable si una actividad no se lleva a cabo.
- **Haga un seguimiento de la ejecución, no solo de los resultados.** Incluya métricas de ejecución sencillas que puedan seguirse semanalmente: el número de participantes inscritos frente al objetivo, las encuestas completadas por día, las sesiones de intervención realizadas según lo previsto.

Estas métricas son diferentes de las métricas de resultados y se centran exclusivamente en si las actividades se han llevado a cabo según lo previsto.

- **Cree un presupuesto completo.** Un presupuesto detallado es aquel que incluye costos que a menudo se olvidan, como los servicios de traducción, el transporte de los supervisores que realizan controles de calidad, el crédito telefónico para los seguimientos, los refrigerios para las reuniones comunitarias y los materiales de reemplazo para los artículos dañados.
- **Especifique con precisión el flujo de datos.** Anote cualquier consideración relacionada con los datos, como la forma en que se transferirán los datos de los formularios en papel a las bases de datos digitales, quién verifica los datos y tiene acceso a ellos, con qué frecuencia y dónde se almacenarán.
- **Planifique los problemas comunes sobre el terreno con contingencias específicas.** ¿Qué pasa si las fuertes lluvias impiden los desplazamientos durante el periodo de las encuestas? ¿Qué pasaría si el

personal clave se enfermara o renunciara? ¿Qué pasaría si las prioridades del gobierno cambiaran repentinamente y la agencia asociada fuera reasignada? ¿Qué pasaría si los participantes estuvieran ocupados con la cosecha cuando se planeó inicialmente el seguimiento? Utilice [la herramienta riesgos de implementación y mitigación](#) para documentar sistemáticamente cualquier riesgo previsto y las estrategias establecidas para abordarlos. Para ver esta herramienta en acción, consulte el Apéndice 2 para **el estudio de caso en el Líbano**, que ilustra su aplicación práctica.

- **Proporcione herramientas prácticas.** Incluya plantillas y procedimientos operativos estándar como anexos, de modo que los equipos de campo dispongan de herramientas prácticas, y no solo de planes abstractos; esto significa guiones reales para la contratación, guías paso a paso para la introducción de datos y listas de verificación para la ejecución de las intervenciones.

## Obtención de la aprobación ética

La aprobación ética es el proceso formal de obtener la aprobación del comité de ética en investigación (CEI) o del comité de ética, lo que implica cumplir con los requisitos de presentación, los plazos y los procedimientos institucionales. Este proceso es necesario para garantizar que la evaluación pueda llevarse a cabo de forma legal y ética.

Más allá de la mera comprensión de los principios éticos, esto implica gestionar el proceso burocrático práctico, que incluye:

- identificar qué IRB tiene jurisdicción (una universidad, el Ministerio de Salud, UNICEF o varios)
- completar las certificaciones de capacitación requeridas para todos los miembros del equipo
- preparar paquetes de documentación exhaustivos en formatos específicos
- responder a los comentarios y solicitudes de aclaración de los revisores
- mantener el cumplimiento a lo largo del estudio, incluyendo enmiendas, notificación de eventos adversos y renovaciones anuales

### Por qué es importante

La aprobación ética es un requisito legal para las investigaciones con sujetos humanos; proceder sin ella puede invalidar toda la evaluación, exponer a las instituciones a responsabilidades legales y destruir la confianza de la comunidad. Muchas revistas no publican los resultados sin una prueba de aprobación ética, y los donantes exigen cada vez más evidencia de autorización ética antes de liberar los fondos. El reto práctico es que la revisión ética puede llevar de dos a seis meses, con múltiples rondas de revisión, y los retrasos en este proceso se propagan a lo largo de todo el calendario previsto. Una evaluación técnicamente perfecta no tiene ningún valor si no se puede comenzar a recopilar datos porque la aprobación ética está pendiente. Además, mantener el cumplimiento ético durante toda la implementación requiere sistemas para documentar las desviaciones del protocolo, notificar los eventos adversos y garantizar que todos los miembros del equipo sigan los procedimientos aprobados.

Tenga en cuenta estas consideraciones clave:

**Empiece pronto.** Inicie el proceso ético antes de finalizar todos los detalles; las modificaciones menores se pueden presentar más adelante. La obtención de la aprobación inicial pone en marcha el reloj.

**Identifique las autoridades adecuadas.** Determine qué organismos de ética tienen jurisdicción. Los comités de ética de las universidades locales suelen exigir la afiliación, los comités de ética del Ministerio de Salud nacional pueden ser necesarios para la investigación sanitaria, pueden requerirse múltiples aprobaciones para los estudios multinacionales y algunos donantes tienen sus propios requisitos éticos.

**Establezca plazos realistas.** Reserve una cantidad de tiempo considerable para el proceso: de dos a tres semanas para preparar los documentos, de cuatro a ocho semanas para la revisión inicial (más si se requiere una revisión completa por parte de la junta), de dos a tres semanas para responder a los comentarios, de dos a cuatro semanas para la aprobación final y tiempo adicional potencial para las aprobaciones locales o nacionales.

**Prepare una presentación completa.** Reúna un paquete de documentación completo que incluya lo siguiente: protocolo detallado con antecedentes, objetivos y métodos; formularios de consentimiento en todos los idiomas locales con traducciones inversas; instrumentos de encuestas, incluso si aún se están perfeccionando;

currículums y certificados de formación de todo el personal clave; planes de gestión de datos con medidas de seguridad; estructuras de compensación con justificación; procedimientos de mitigación de riesgos y derivación.

**Comprenda el nivel de revisión que probablemente se requiera, ya que esto afecta al calendario.** Una revisión exenta (riesgo mínimo, categorías específicas) tarda entre 2 y 3 semanas; una revisión acelerada (riesgo mínimo, no exenta) tarda entre 4 y 6 semanas; una revisión completa (riesgo superior al mínimo, poblaciones vulnerables) puede tardar entre 2 y 3 meses y solo se realiza una vez al mes.

**Evite los retrasos habituales.** Las razones más comunes de los retrasos son: solicitudes incompletas en las que faltan secciones obligatorias, formularios de consentimiento que utilizan jerga técnica o en los que faltan elementos obligatorios, planes de evaluación o mitigación de riesgos inadecuados, compensaciones que parecen coercitivas, procedimientos de protección de datos poco claros y firmas o aprobaciones institucionales que faltan.

**Después de la aprobación, mantenga el cumplimiento.** Esto implica capacitar a todos los nuevos miembros del equipo sobre el protocolo, documentar y reportar cualquier desviación, presentar enmiendas antes de realizar cambios, presentar informes anuales de revisión continua y cerrar adecuadamente el estudio cuando se haya completado.

## Supervisión de la implementación y evaluación del proceso

Se trata de un sistema integral para hacer un seguimiento de cómo se lleva a cabo realmente la intervención sobre el terreno, que combina rutinas periódicas de supervisión de la implementación con la documentación sistemática del proceso. Esto implica controles estructurados (diarios, semanales o quincenales, dependiendo de la intensidad) utilizando herramientas estandarizadas para evaluar múltiples dimensiones de la ejecución, entre **ellas**:

- **Fidelidad:** ¿se llevó a cabo la intervención tal y como se había diseñado?
- **Alcance:** ¿qué proporción de la población objetivo recibió la intervención?
- **Dosis:** la frecuencia e intensidad de la ejecución.
- **Calidad:** ¿qué tan bien se llevó a cabo la intervención?

- **Participación de los participantes:** ¿los participantes la entendieron y actuaron en consecuencia?
- **Factores contextuales** que afectan a la implementación

Incluye el seguimiento en tiempo real de las tasas de reclutamiento en relación con los objetivos, la supervisión de los patrones de pérdida de unidades de estudio para mantener el poder estadístico, la documentación de todas las adaptaciones realizadas durante la ejecución y la recopilación de comentarios de los ejecutores y los participantes sobre lo que funciona y lo que no.

### Por qué es importante

Muchas evaluaciones no encuentran ningún impacto, no porque la intervención no funcione, sino porque nunca se llevó a cabo correctamente. La falta de un

seguimiento sistemático supone un riesgo: por ejemplo, tras una costosa recopilación de datos finales, se podría descubrir que la mitad del grupo de tratamiento nunca recibió la intervención, que los grupos de control accedieron al tratamiento o que el personal de campo modificó la intervención hasta dejarla irreconocible. La evaluación del proceso distingue entre **el fracaso de la teoría** (la intervención realmente no funciona, incluso cuando se ejecuta bien) y **el fracaso de la implementación** (no se ejecutó correctamente, por lo que no se puede juzgar su eficacia).

Esta información es crucial para interpretar los resultados. Por ejemplo, si no se encontró ningún impacto, ¿fue porque la teoría era errónea o porque solo el 30 % de los participantes recibió la intervención completa? Si se encontraron efectos positivos, comprender lo que realmente se implementó ayuda en los esfuerzos de replicación. El monitoreo en tiempo real permite corregir el rumbo mientras aún hay tiempo. Por ejemplo, si el reclutamiento se está retrasando, se pueden intensificar los esfuerzos antes de que esto ponga en peligro el poder estadístico; si ciertos sitios no están ejecutando correctamente, se puede proporcionar apoyo adicional; si surgen barreras inesperadas, se pueden desarrollar soluciones.

Los datos del proceso también revelan perspectivas críticas para las decisiones de escalar: qué contextos facilitaron una ejecución fluida, qué adaptaciones fueron necesarias, qué retos de implementación es probable que persistan a gran escala y qué nivel de calidad es realista alcanzar en condiciones rutinarias frente a entornos de investigación.

Tenga en cuenta estas consideraciones clave:

- **Establezca rutinas de supervisión que se ajusten a la intensidad de la implementación, pero que no abrumen a los equipos de campo.** Esto puede implicar reuniones diarias para intervenciones intensivas, llamadas semanales para programas estándar o revisiones mensuales para intervenciones ligeras.
- **Cree herramientas sencillas.** Las hojas de monitoreo estandarizadas pueden recopilar información esencial sin crear una carga excesiva de papeleo.
- **Haga un seguimiento de las métricas básicas de implementación por separado de los datos de resultados.** Esto puede implicar el porcentaje de sesiones planificadas impartidas, la asistencia media o las tasas de participación, el porcentaje que recibe la dosis completa de la intervención, el tiempo entre los componentes de la intervención o las calificaciones de calidad de las observaciones estandarizadas.
- **Establezca sistemas para el monitoreo de la calidad de los datos.** Estos pueden incluir verificaciones de alta frecuencia (revisión diaria o semanal automatizada de los datos para detectar valores atípicos, datos faltantes y patrones sospechosos) junto con verificaciones posteriores (volver a entrevistar al 10-20 % de los participantes para verificar la exactitud de los datos y detectar cualquier posible fraude).
- **Supervise continuamente las hipótesis que afectan el poder estadístico:** Tenga en cuenta las tasas de reclutamiento en comparación con los objetivos (¿va el proyecto por buen camino para alcanzar el tamaño de la muestra?), las tasas de pérdida de unidades de estudio generales (¿qué porcentaje se perdió durante el seguimiento?) y la pérdida de unidades de estudio diferencial (¿es mayor la pérdida de unidades de estudio en el grupo de tratamiento o en el de control?), las tasas de cumplimiento (¿qué porcentaje del grupo de tratamiento está recibiendo realmente la intervención?) y la contaminación (¿tiene acceso el grupo de control a la intervención?).
- **Documente cada adaptación utilizando marcos estructurados.** Anote específicamente qué ha cambiado con respecto al protocolo, por qué era necesario el cambio (barrera encontrada, solicitud de las partes interesadas, problema de viabilidad), cuándo se produjo el cambio, quién tomó la decisión, si el cambio fue planificado o reactivo y, por último, si esto afecta a la teoría de intervención central.
- **Cree bucles de retroalimentación rápida con los implementadores.** Utilice grupos sencillos de WhatsApp para resolver problemas en tiempo real, llamadas semanales breves para verificar el estado de la situación, centrándose en los retos y las soluciones, y revisiones mensuales de los datos de seguimiento para identificar patrones.
- **Distinga entre componentes básicos y adaptables.** Diferencie entre los componentes básicos que deben mantenerse para que la teoría de la intervención se mantenga y los elementos periféricos que pueden adaptarse al contexto. Documente ambos tipos, pero trátelos de manera diferente en el análisis.

- **Recopile los comentarios de los participantes.** Incorpore mecanismos de retroalimentación de los participantes mediante breves entrevistas de salida después de las sesiones de intervención, grupos focales periódicos con los participantes y buzones de sugerencias o líneas directas para recibir comentarios de forma continua.
- **Haga un seguimiento de los factores contextuales que podrían afectar a la implementación o a los resultados.** Tome nota de otros programas o políticas

concurrentes que afecten a la población objetivo, factores estacionales (vacaciones, temporadas agrícolas y clima), situaciones políticas o de seguridad, y emergencias sanitarias u otras perturbaciones.

- **Mantenga registros detallados.** Estos son fundamentales para interpretar los resultados, informar las decisiones de escalar y contribuir a una base empírica más amplia sobre los retos y soluciones de la implementación.

## Análisis de costo-beneficio

Se trata de un cálculo sistemático de todos los recursos necesarios para llevar a cabo la intervención y lograr los impactos medidos, lo que da lugar a métricas estandarizadas que permiten comparar diferentes intervenciones, modelos de ejecución u opciones de inversión. Esta contabilidad exhaustiva va más allá de los simples presupuestos de los programas para reflejar el costo económico total de la consecución de los resultados, incluyendo:

- los costos directos del programa (personal, materiales, operaciones)
- los costos indirectos que a menudo se ocultan en otros presupuestos (supervisión, administración, gastos generales)
- costos de oportunidad de los recursos utilizados (tiempo de los voluntarios, personal gubernamental, tiempo de los participantes)
- costos iniciales frente a costos de funcionamiento
- costos marginales de añadir participantes

El análisis genera métricas como el costo por niño vacunado, el costo por punto porcentual de aumento en las calificaciones de las pruebas, el costo por año de vida salvado o los índices de retorno de la inversión que los responsables de la toma de decisiones pueden comparar con puntos de referencia o intervenciones alternativas. Consulte la herramienta [de análisis de costo-beneficio](#) aquí para ayudar a identificar, cuantificar y comparar los costos y beneficios de la intervención. Para ver un ejemplo de cómo aplicar esta herramienta utilizando el **estudio de caso del Líbano**, consulte el Apéndice 2.

### Por qué es importante

Incluso las intervenciones altamente eficaces pueden no merecer la pena si son prohibitivamente caras en comparación con las alternativas. El análisis de costo-beneficio transforma la evaluación en una orientación política aplicable, respondiendo a la pregunta «¿Es rentable?», evidenciando que algo funciona. Esto es cada vez más importante, ya que los donantes exigen evidencia tanto del impacto como de la rentabilidad, los gobiernos con presupuestos limitados necesitan maximizar los resultados por cada dólar gastado y las decisiones de escalar requieren comprender cómo cambian los costos con la escala.

Un programa que logre una mejora del 10 % puede parecer exitoso, hasta que se da cuenta de que cuesta cinco veces más que una alternativa que logra una mejora del 8 %. Comprender las estructuras de costos también revela oportunidades de eficiencia: tal vez el 80 % del impacto podría lograrse con el 50 % del costo mediante la racionalización de ciertos componentes, o los costos fijos que parecen elevados para un programa piloto serían insignificantes a gran escala. Sin un cálculo riguroso de los costos, los programas pueden abandonarse por ser «demasiado caros» basándose en una comprensión incompleta, o escalar con entusiasmo sin reconocer estructuras de costos insostenibles.

Tenga en cuenta estas consideraciones clave:

- **Commencez tôt.** Il est essentiel de collecter les données de coûts dès le premier jour de mise en œuvre : reconstruire les coûts a posteriori est souvent peu fiable et parfois impossible lorsque les

reçus sont égarés, que le personnel oublie comment son temps a été alloué ou que les contributions en nature n'ont pas été documentées.

- **Empiece pronto.** Es importante recopilar datos sobre los costos desde el primer día de la implementación: la reconstrucción retrospectiva de los costos no es fiable y, a menudo, es imposible, ya que se pierden los recibos, el personal olvida la distribución del tiempo y las contribuciones en especie no se documentan.
- **Capture todos los costos, no solo los del presupuesto.** Estos costos pueden incluir el tiempo del personal, incluida la preparación y la capacitación (incluso si lo pagan los socios), el tiempo de los voluntarios valorado según los salarios locales por un trabajo equivalente, el tiempo del personal gubernamental, incluso si no lo paga el proyecto, los costos de los participantes (transporte, salarios perdidos, cuidado de niños), los materiales o lugares donados a su valor de mercado y los gastos generales de la organización que se atribuyen razonablemente al proyecto.
- **Distinga entre las diferentes categorías de costos, ya que se comportan de manera diferente a escala.** Estos incluyen los costos fijos (desarrollo de la capacitación, configuración inicial) que no aumentan con el número de participantes, los costos variables (materiales por participante, incentivos) que escalan linealmente y los costos escalonados (supervisión, nuevos sitios) que aumentan en los umbrales.
- **Realice un seguimiento de los costos desde múltiples perspectivas, ya que las diferentes partes interesadas se preocupan por diferentes cifras.** Considere la perspectiva del implementador (¿cuánto le cuesta a este equipo ponerlo en marcha?), la perspectiva del gobierno (¿cuánto costaría integrarlo en los sistemas existentes?), la perspectiva social (incluidos todos los costos, independientemente de quién los pague) y la perspectiva de los participantes (¿cuánto les cuesta la participación a los beneficiarios?).
- **Calcule múltiples métricas de rentabilidad para permitir diferentes comparaciones.** Esto puede incluir el costo por participante alcanzado/ inscrito/que completa el programa; el costo por unidad de resultado primario alcanzado; el

costo por tamaño del efecto estandarizado para comparaciones académicas; y las relaciones de rentabilidad incremental si se comparan variantes.

- **Compare los resultados con los puntos de referencia pertinentes.** Tome nota de intervenciones similares en el contexto elegido, la disposición revelada por el gobierno a pagar por resultados similares, las normas internacionales (como los umbrales de la OMS para las intervenciones sanitarias) y las formas alternativas de alcanzar objetivos similares.
- **Incluya análisis de sensibilidad que muestren cómo cambia la rentabilidad en función de diferentes supuestos.** Considere si los efectos persisten durante dos años frente a uno, si la intervención es implementada por el gobierno frente a los salarios de las ONG, a diferentes escalas (100, 1000, 10 000 participantes) o con diferentes tasas de cumplimiento o pérdida de unidades de estudio.
- **Documente los costos de los factores impulsores de manera transparente.** ¿Qué hace que esta intervención sea cara o barata? ¿Se podrían modificar componentes específicos para reducir los costos sin sacrificar la eficacia? ¿Qué economías o deseconomías de escala son probables? ¿Qué costos ocultos podrían surgir en la rutina de implementación frente a las condiciones de investigación?
- **Presente los resultados de forma accesible para el público interesado en las políticas.** Esto puede incluir métricas sencillas de costo por resultado, en lugar de modelos económicos complejos, comparaciones visuales con intervenciones conocidas, declaraciones claras sobre los intervalos de confianza e implicaciones prácticas para la planificación presupuestaria.
- **Interprete la rentabilidad con prudencia.** Recuerde que lo más barato no siempre es lo mejor: a veces, las intervenciones de mayor costo merecen la pena si sus repercusiones son proporcionalmente mayores o si llegan a poblaciones a las que no llegan las alternativas más baratas.

Las pruebas rigurosas refuerzan el vínculo entre el diseño y la toma de decisiones. Al basar las conclusiones en evidencia en lugar de en suposiciones, las evaluaciones de impacto permiten a UNICEF y a sus socios empezar a identificar qué intervenciones son eficaces, cuáles requieren adaptación y cuáles deben suspenderse. Mediante mediciones sistemáticas y análisis transparentes,

las perspectivas del comportamiento y las hipótesis de diseño se traducen en pruebas creíbles y aplicables que sirven de base para las decisiones sobre la escala, las políticas y la asignación de recursos.

A medida que los programas pasan a la fase de *escalado* de la prueba de hipótesis DEPTHS, se aplica la evidencia generada a partir de *las hipótesis de prueba*. Los resultados de las pruebas rigurosas orientan la forma en que se perfeccionan las intervenciones, se integran en los sistemas y se amplían de manera responsable. Esto garantiza que las decisiones sobre la ampliación y la replicación de las intervenciones se basen en el impacto demostrado, en lugar de en suposiciones.

# Apéndice

## APÉNDICE 1:

### Alternativas cuasi experimentales cuando la aleatorización no es factible

Este apéndice ofrece orientación técnica adicional para los profesionales que diseñan evaluaciones de impacto en condiciones reales y rutinarias. Si bien el capítulo principal destaca los diseños aleatorios como el método más confiable para establecer la causalidad, las realidades prácticas a veces limitan su uso. En tales casos, los evaluadores pueden necesitar considerar enfoques alternativos a la aleatorización que sigan teniendo como objetivo generar perspectivas creíbles y basadas en evidencia.

La prioridad en el diseño de la evaluación debe ser la aleatorización, ya que sigue siendo el método más confiable para establecer la causalidad. La aleatorización elimina las diferencias sistemáticas entre los grupos, lo que permite una mayor confianza en que los efectos observados son atribuibles a la intervención en sí. Sin embargo, hay circunstancias en las que la aleatorización no es posible debido a restricciones políticas, éticas, logísticas o prácticas. En tales casos, se pueden considerar enfoques alternativos.

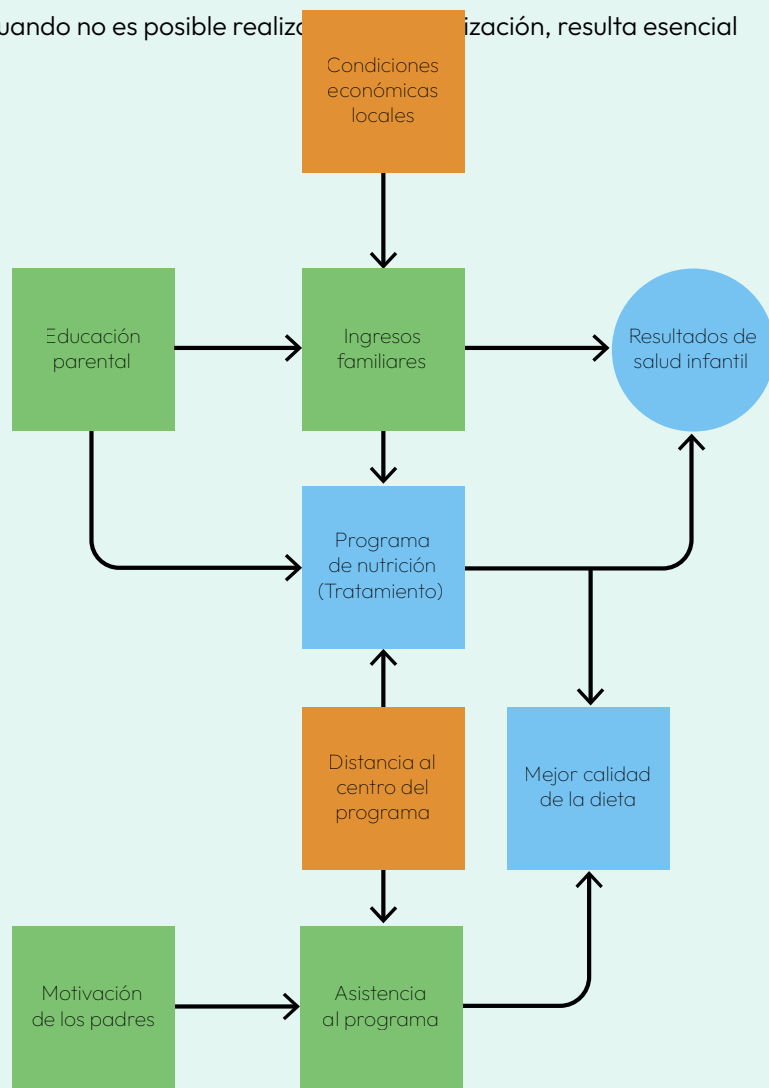
Si bien estas alternativas pueden seguir aportando perspectivas valiosas, introducen una mayor complejidad operativa y mayores exigencias estadísticas. Y lo que es más importante, estos métodos también conllevan un mayor riesgo de sesgo. Se recomienda contar con la participación de expertos en inferencia causal y diseño de evaluaciones cuando se opte por diseños no aleatorios.

1. **Cerrar las rutas traseras: una perspectiva DAG.** Cuando no es posible realizar la randomización, resulta esencial identificar y controlar los factores de confusión, un enfoque conocido en la inferencia causal como «cerrar las rutas traseras». Los gráficos acíclicos dirigidos (DAG) proporcionan un marco útil para comprender este proceso.

Consideremos un DAG que ilustra la evaluación de un programa comunitario de nutrición dirigido a los niños:

En dicho diagrama:

- Las flechas representan efectos de efecto causal directo.
- Las variables (nodos) denotan factores que influyen en los resultados o en la participación en el programa.
- Las rutas entre las variables representan posibles asociaciones causales o no causales.



2. **Identificación de rutas traseras.** Una «ruta trasera» se refiere a cualquier ruta entre el tratamiento (participación en el programa de nutrición) y el resultado (resultados de salud infantil) que no sigue la dirección causal directa. Estas rutas introducen asociaciones que pueden sesgar las estimaciones causales.

En el ejemplo del DAG, pueden existir varias rutas traseras:

- **Programa → Ingresos familiares → Salud infantil:** las familias con ingresos más altos pueden ser más propensas a participar en el programa y a tener hijos más sanos, independientemente del programa.
  - **Programa ← Educación de los padres → Salud infantil:** los padres con mayor nivel educativo pueden inscribir a sus hijos con más frecuencia y también proporcionarles mejores cuidados.
  - **Programa ← Educación de los padres → Ingresos familiares → Salud infantil:** la educación de los padres influye en los ingresos, lo que a su vez afecta tanto a la participación como a los resultados de salud.
3. **El objetivo: cerrar todas las rutas traseras.** Para aislar el efecto causal del programa de nutrición, deben cerrarse todas las rutas traseras. Una vía se considera cerrada cuando:
    - **Se controla (se condiciona) una variable de la vía.** Por ejemplo: el control de los ingresos familiares cierra la vía Programa → Ingresos familiares → Salud infantil.

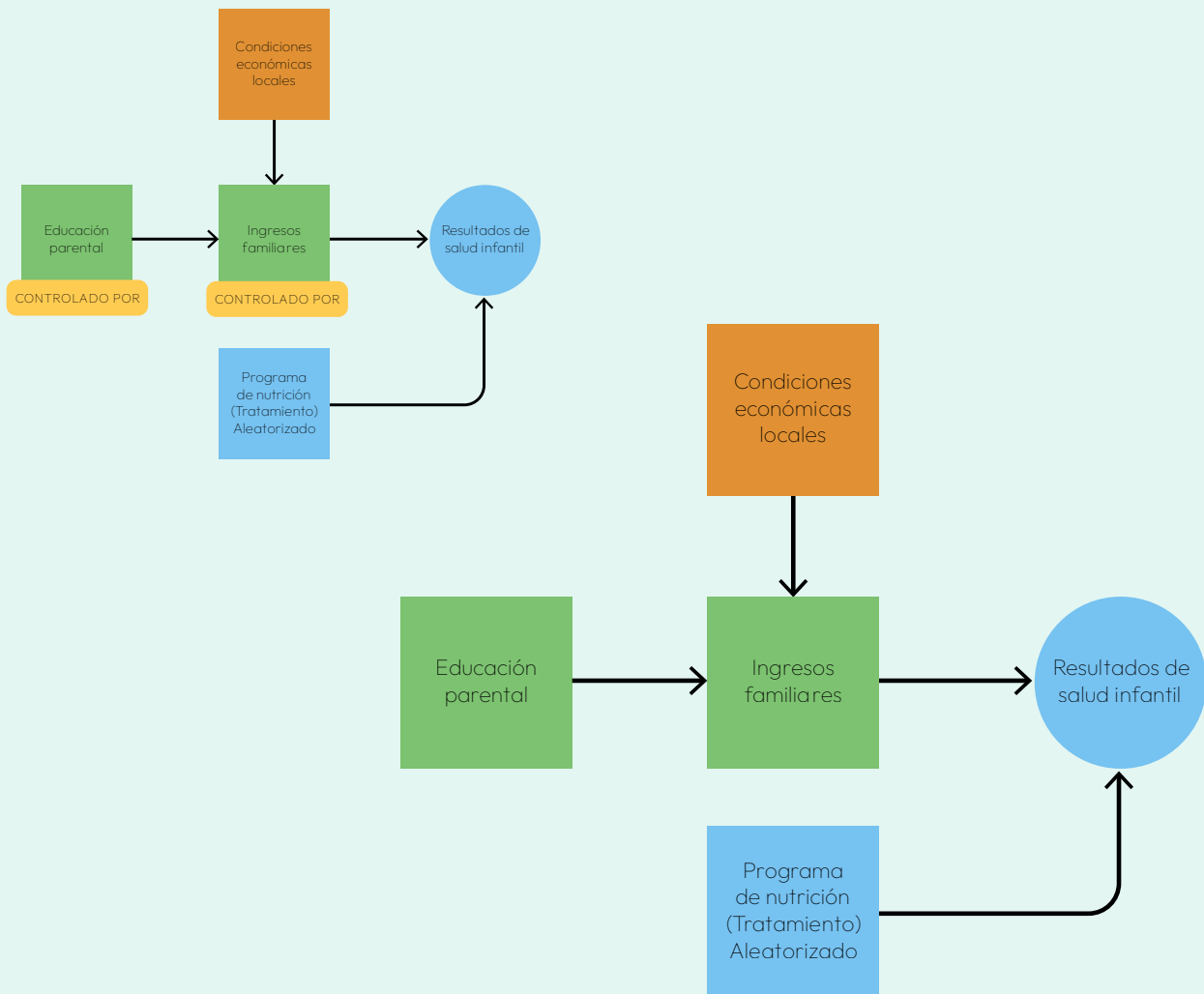
- **La vía contiene un colisionador.** Un colisionador es una variable influenciada por dos o más variables. Por ejemplo, la «asistencia al programa» puede estar influenciada tanto por la «distancia al centro del programa» como por la «motivación de los padres». Esta vía se cierra de forma natural, a menos que el colisionador esté condicionado incorrectamente, lo que volvería a abrir la vía.
- **La ruta incluye un mediador que se deja intencionadamente sin controlar.** Por ejemplo, la «mejora de la calidad de la dieta» se encuentra en la cadena causal entre la participación en el programa y la salud infantil. Si lo que interesa es el efecto total, no se debe controlar esta variable.

Cuando se aplica la aleatorización, todas las flechas que apuntan al nodo «Participación en el programa» se cortan efectivamente. Esto hace que la participación sea independiente de todos los factores de confusión y cierra todas las rutas traseras a la vez. Esta es la ventaja fundamental de la aleatorización: elimina la necesidad de identificar o medir todas las posibles fuentes de sesgo.

4. **Cierre de las rutas traseras sin aleatorización.** En ausencia de aleatorización, se utilizan métodos estadísticos para cerrar las rutas traseras. Dos estrategias comunes son:

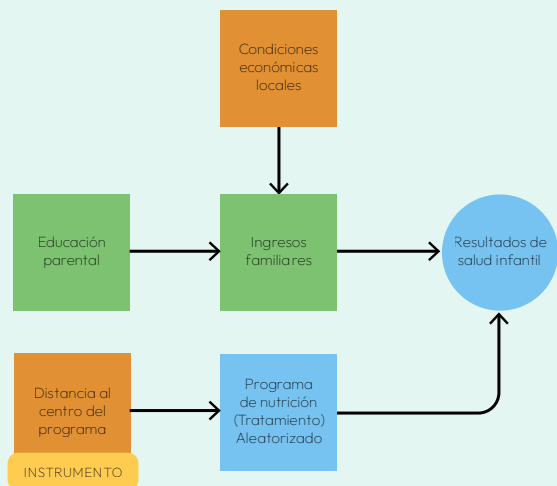
a. **Control de los factores de confusión observados.** Esto implica medir y ajustar variables como la educación de los padres y los ingresos familiares mediante métodos como el ajuste por regresión o el emparejamiento. Este enfoque requiere :

- La identificación de todos los factores de confusión relevantes.
- La medición precisa de esos factores de confusión.
- Modelar correctamente sus relaciones con el tratamiento y los resultados.



**b. El uso de variables instrumentales.** Este método se basa en una variable que influye en la participación en el programa, pero que no está directamente relacionada con el resultado. Por ejemplo, la «distancia al centro del programa» puede determinar la participación sin afectar directamente a la salud infantil. Esta variación puede utilizarse para estimar el efecto causal incluso en presencia de factores de confusión no observados.

Las oportunidades y limitaciones de cada método dependen del contexto y de la disponibilidad de datos. Las diferentes situaciones ofrecerán diferentes puntos de apalancamiento para cerrar las rutas traseras, y la elección del enfoque adecuado requiere un examen cuidadoso tanto del diseño como de las limitaciones de los datos.



**La tabla 3 del paso 1** de este capítulo incluye una descripción de los diferentes enfoques cuando no es posible la aleatorización.

## APÉNDICE 2:

# El caso del aumento de la cobertura de vacunación infantil en el Líbano

*Esta fase también utiliza el estudio de caso presentado en fases anteriores, cuyo objetivo es aumentar la vacunación infantil en entornos de bajos recursos en el Líbano, con el fin de ilustrar las herramientas del proceso DEPTHS.*

## Contexto del estudio de caso:

A través de las fases de *Definición*, *Exploración* y *Diagnóstico*, y *Diseño de Prototipos*, el equipo del Líbano identificó un comportamiento objetivo que cambiar (que los cuidadores regresaran con sus hijos para la siguiente vacunación programada) y sus principales barreras del comportamiento (sobrecarga cognitiva por tareas concurrentes, estrés emocional relacionado con las visitas a la clínica y percepciones sociales moldeadas por las opiniones de la comunidad). También probaron un prototipo de una posible solución: una tarjeta de cita en papel destinada a hacer que la siguiente visita de vacunación fuera más destacada y fácil de recordar.

La intervención se probó mediante un ensayo controlado aleatorio en el que participaron 12,332 niños sin vacunar o con vacunación incompleta de 6,160 hogares. Los hogares se asignaron aleatoriamente a un grupo de tratamiento que recibió la tarjeta de cita durante las visitas de divulgación o a un grupo de control que no la recibió. El resultado principal se definió a nivel de los hogares: si al menos un niño elegible del hogar recibió la siguiente vacuna requerida en los 21 días siguientes a la fecha prevista.

Se seleccionó esta medida binaria a nivel de hogar para reflejar mejor la toma de decisiones de los cuidadores, reconociendo que el costo marginal de vacunar a más niños en el mismo hogar suele ser bajo. El análisis utilizó una regresión logística para estimar el efecto de la intención de tratar, con errores estándar robustos agrupados a nivel de los trabajadores sociales. Las covariables de referencia incluyeron el tamaño del hogar, la nacionalidad y el comportamiento previo en materia de inmunización.

El ensayo reveló que los hogares que recibieron la tarjeta de cita eran significativamente más propensos a volver para la vacunación, con un aumento de 6,7 puntos porcentuales en comparación con el grupo de control. Estos resultados demuestran que una intervención de bajo costo basada en el comportamiento puede mejorar de manera apreciable la aceptación de la inmunización infantil cuando se dirige a puntos clave de decisión. Los resultados han servido de base para los debates políticos en curso y demuestran el valor de las pruebas de hipótesis estructuradas en los programas de salud pública basados en el comportamiento.

De acuerdo con la descripción de la guía de campo, este apéndice incluye un ejemplo completo para ilustrar su aplicación práctica y ayudar a los equipos a utilizar las herramientas mencionadas con facilidad y coherencia. Estas herramientas incluyen:

- Agenda de aprendizaje
- Plan de implementación
- Mitigación de riesgos de implementación
- Análisis de costo-beneficio

## Aplicación de la agenda de aprendizaje

*Esta agenda de aprendizaje no fue desarrollada por el equipo original del proyecto. Se trata de un ejemplo recreado a partir de datos y contexto reales del proyecto.*

A continuación se ilustra cómo aplicar la **agenda de aprendizaje** presentada en la guía de campo, utilizando el estudio de caso del Líbano. Esta herramienta ofrece cómo enmarcar y priorizar las preguntas clave de aprendizaje que guían tanto el diseño de la evaluación como la interpretación de los resultados. El ejemplo destaca cómo una agenda de aprendizaje clara, centrada en lo que el equipo necesita aprender, por qué es importante y cómo se utilizarán los resultados, garantiza que la evaluación sea útil, viable y adaptable.

### Agenda de aprendizaje

Utilice esta hoja de trabajo para definir los elementos clave de su plan de evaluación utilizando el marco PICOS.

Intervención: Tarjeta de recordatorio de cita

**Problema** Los cuidadores de niños pequeños, especialmente en comunidades sirias y libanesas de bajos ingresos, a menudo olvidan o desconocen cuándo deben acudir a la siguiente cita de vacunación de sus hijos. Esto hace que se pierdan o se retrasen las vacunaciones rutinarias, dejando a los niños vulnerables a enfermedades prevenibles mediante vacunación y aumentando los abandonos del calendario de vacunación.

#### PICOS

**Población** Cuidadores primarios de niños de 0 a 15 meses que acuden a clínicas de salud pública en el Líbano, especialmente los de hogares sirios y libaneses que residen en entornos de bajos recursos.

**Intervención** Una tarjeta de recordatorio de la cita con información sobre el comportamiento (tamaño tarjeta postal), proporcionada a los cuidadores después de una visita de vacunación. La tarjeta incluye: (a) La fecha de la próxima cita de vacunación del niño, (b) un sello del Ministerio de Salud Pública; (c) señales visuales para aumentar la prominencia; y (d) un formato sencillo, de bajo costo y baja tecnología diseñado para actuar como dispositivo de compromiso y recordatorio ambiental.

#### Comparación

Grupo de control aleatorizado

#### Resultado

1. Resultado primario: Tasa de regreso a la cita programada para la vacuna de seguimiento (binario: regresó frente a no regresó).
2. Resultados secundarios:
  - a. Intención de regresar autoinformada por los cuidadores.
  - b. Capacidad de los cuidadores para recordar la fecha exacta de seguimiento
  - c. Percepciones autodeclaradas de la utilidad de la tarjeta recordatoria

#### Estudio y tipo de evaluación

Impact

#### Lecciones clave

##### Si los resultados son positivos

La intervención puede escalarse como una solución de bajo coste, basada en el comportamiento, para aumentar las tasas de seguimiento de la vacunación en entornos de bajos recursos. Las lecciones aprendidas guiarán a los equipos de los programas del Ministerio de Salud Pública y AIA en la institucionalización de la tarjeta postal como parte de la prestación de servicios estándar. Podrían probarse otras mejoras del comportamiento (por ejemplo, mensajes, color, ubicación en el hogar).

##### Si los resultados son nulos o no concluyentes:

Puede ser necesario adaptar el diseño de la postal, el proceso de entrega o los mecanismos de apoyo. Entre las posibles mejoras cabe citar la explicación verbal de la fecha de regreso, la orientación a otros miembros de la familia (por ejemplo, las abuelas) o la incorporación de empujones adicionales. Los comentarios cualitativos de los cuidadores y el personal de primera línea deben orientar las adaptaciones.

##### Si los resultados son negativos (la intervención es contraproducente):

Investigar los efectos no deseados (por ejemplo, confusión, desconfianza o percepción de obligación). Analizar si alguna dinámica operativa o social interfirió en la intervención (por ejemplo, desinformación, mala comunicación, estigma). Puede que sea necesario rediseñar la postal o sustituirla por otro empujón mejor adaptado a las rutinas y normas locales.

El equipo del Líbano elaboró primero una agenda de aprendizaje sencilla para aclarar qué esperaban aprender con la evaluación y por qué. Identificaron un problema específico: los cuidadores de las comunidades libanesas y sirias no llevaban a cabo las vacunaciones rutinarias de sus hijos, a menudo porque se olvidaban o no entendían cuándo debían volver. La intervención, una sencilla postal diseñada para influir en el comportamiento, tenía por objeto abordar este problema sirviendo de recordatorio físico.

La pregunta de aprendizaje principal que plantearon fue: «¿Proporcionar una tarjeta personalizada de recordatorio de citas aumenta la aceptación oportuna de la vacunación infantil entre los cuidadores de las comunidades libanesas y sirias de bajos ingresos?».

Las preguntas secundarias incluían cómo influía la intervención en la intención de los cuidadores de volver y en su recuerdo de la fecha de seguimiento. También se trazaron las interpretaciones previstas de los resultados.

Por ejemplo, si aumentaba la aceptación, se podría escalar la intervención. Si los resultados eran mixtos o nulos, podría ser necesario realizar ajustes en el diseño y más pruebas. Si se producía un efecto contrario al deseado, el equipo investigaría los efectos no deseados, como la desconfianza o la mala interpretación.

# Aplicación del plan de implementación

Este plan de implementación no fue desarrollado por el equipo original del proyecto. Se trata de un ejemplo recreado basado en datos y contexto reales del proyecto.

A continuación se ilustra cómo aplicar el **plan de implementación** presentado en la descripción de la guía de campo, utilizando el estudio de caso del Líbano. Esta herramienta muestra cómo traducir el diseño de la evaluación en una planificación operativa estructurada, definiendo las actividades clave, las funciones, los plazos y los indicadores de seguimiento. El ejemplo proporciona una referencia práctica para desarrollar planes de implementación que promuevan la coordinación, la transparencia y la rendición de cuentas a lo largo del proceso de evaluación.

## Elaboración de un plan de implementación

Para poner en marcha la evaluación, el equipo del Líbano también desarrolló un plan de implementación claro en el que se esbozaban los pasos prácticos, las responsabilidades y el calendario de ejecución. Este plan detallaba cada fase, desde la finalización del diseño de la postal y la realización de una pequeña prueba previa, hasta la recopilación de datos de referencia y la asignación de los participantes a los grupos de tratamiento o control. Cada etapa se asignó a los responsables correspondientes (por ejemplo, el equipo del proyecto de UNICEF, J-PAL MENA, el Ministerio de Salud Pública), con un calendario definido e indicadores específicos para el seguimiento de los progresos, como las tasas de inscripción, la cobertura de la distribución y los controles de calidad de los datos. Este plan sirvió no solo como herramienta de coordinación, sino también como base para la transparencia y la rendición de cuentas a lo largo de la implementación.

Equipado con el plan de evaluación y el plan de implementación, el equipo estaba listo para someterlo a una junta de revisión institucional para su revisión y aprobación ética antes de implementar y evaluar la intervención.

### Plan de implementación

Intervención: Tarjeta de recordatorio de cita

Utilice esta hoja de trabajo para dividir su intervención en pasos prácticos. Para cada área prioritaria, defina las actividades clave, asigne responsabilidades, establezca un calendario e identifique qué recursos e indicadores de seguimiento serán necesarios para seguir los avances.

Área prioritaria <i>[Inserte la prioridad]</i>	Actividades clave <i>[Enumere 1-2 acciones]</i>	Responsable <i>[Responsable y colaboradores]</i>	Calendario <i>[MM/AA]</i>	Recursos y presupuesto <i>[Insumos necesarios]</i>	Indicadores de seguimiento <i>[Por ejemplo, % de actividades realizadas]</i>
Piloto realizado	Finalizar el diseño de la postal y realizar una pequeña prueba previa en 1-2 clínicas	Equipo del programa AIA, unidad de diseño del Ministerio de Sanidad, equipo del proyecto de UNICEF	03/23	Consultor de diseño, transporte	Finalización del diseño de la postal e integración de los comentarios
Recopilación de datos de referencia	Recopilación de datos de referencia sobre las tasas de retorno de los cuidadores y el rendimiento de las clínicas	Equipo de evaluación (J-PAL MENA), Ministerio de Sanidad, equipo del proyecto de UNICEF	03-04/23	Tiempo de los encuestadores, viajes, tabletas	% de encuestas de referencia completadas, comprobaciones de la calidad de los datos
Reclutamiento realizado	Identificar e inscribir a los cuidadores elegibles durante las visitas clínicas rutinarias	Trabajadores sanitarios, personal administrativo de la clínica, equipo del proyecto de UNICEF	05-07/23	Orientación del personal, listas clínicas	# N° de cuidadores inscritos; tasa de consentimiento
Asignación realizada	Asignación aleatoria de los cuidadores a la postal (tratamiento) o a la atención estándar (control)	Equipo de evaluación, apoyo a la aplicación de la AIA, equipo del proyecto de UNICEF	05-07/23	Protocolo de aleatorización, formularios de datos	Aleatorización completada, comprobaciones de equilibrio verificadas
Comienzo de la intervención	Comienzo de la distribución de postales a los cuidadores del grupo de tratamiento después de la vacunación	Personal del dispensario, supervisado por el equipo de campo de AIA, equipo del proyecto de UNICEF	05/23	Postales impresas, formularios de seguimiento	Porcentaje de cuidadores elegibles que reciben la postal
Comprobación de la ejecución	Realización de llamados de supervisión intermedias y visitas sobre el terreno para evaluar la fidelidad y el alcance	Equipo de campo de AIA, puntos focales de distrito del MSP, equipo de proyecto de UNICEF	06/23	Teléfono/datos, costes de las visitas sobre el terreno	% de clínicas que informan sobre la rutina, desviaciones observadas
Fin de la intervención	Concluir la fase de distribución y detener la inscripción de nuevos participantes	Personal del dispensario, notificación al equipo de evaluación, equipo del proyecto de UNICEF	07/23	N/A	Fecha límite aplicada en todos los centros piloto
Datos recopilados	Administración de encuestas de seguimiento a los cuidadores; extracción de los registros de devolución del dispensario.	Encuestadores de J-PAL, personal de seguimiento y evaluación de las clínicas, equipo del proyecto de UNICEF	08-09/23	Herramientas de encuesta, transporte, incentivos	% de encuestas de seguimiento completadas, registros extraídos
Depuración y análisis de datos	Limpieza de los conjuntos de datos, análisis estadístico de los resultados primarios y secundarios.	Analistas de evaluación (J-PAL MENA), equipo del proyecto de UNICEF	10-11/23	Tiempo de los analistas, software	Plan de análisis final completado; resultados validados

## Aplicación de la herramienta de mitigación de riesgos

A continuación se ilustra cómo aplicar la **herramienta de riesgos de implementación y mitigación** presentada en la guía de campo principal, utilizando el estudio de caso del Líbano. Esta herramienta muestra cómo se pueden identificar y gestionar sistemáticamente los posibles riesgos operativos, contextuales y de comportamiento a lo largo del proceso de evaluación. El ejemplo proporciona una referencia práctica para los equipos que desean anticipar y abordar los retos de la implementación antes de que amenacen la integridad de la evaluación o el éxito del programa.

### Gestión de los riesgos durante la puesta en marcha

El equipo utilizó la herramienta de riesgos de implementación y mitigación de forma dinámica, no solo en la fase de planificación, sino también como parte de la implementación activa. Surgieron tres riesgos clave :

RIESGOS	PROBABILIDAD	IMPACTO	MITIGACIÓN
<b>Las postales no llegan a tiempo</b>	media	alto	El equipo de UNICEF trabajó con un servicio de mensajería para escalonar los horarios de entrega y proporcionar un stock de reserva.
<b>Las enfermeras no entregaban las postales con la explicación adecuada</b>	alta	Medio	Se impartió formación mediante notas de voz y se distribuyeron hojas de referencia plastificadas.
<b>Los cuidadores extravían la postal</b>	media	Medio	Se utilizaron fundas de plástico transparentes y se recordó que se pegara la postal cerca del calendario en casa.
<b>Los registros de datos no se actualizan de manera constante</b>	media	alto	Se integró un recordatorio rápido para el registro al final de cada turno a través de avisos de WhatsApp.
<b>Interrupciones contextuales</b> (por ejemplo, huelga de transporte, condiciones meteorológicas)	baja a media	alto	Se asignaron puntos focales locales para la reasignación flexible de las rutas de entrega.

Cada riesgo se asignó al equipo de UNICEF o a un supervisor del Ministerio de Salud Pública para su seguimiento, con plazos vinculados al calendario de implementación.

# Aplicación del análisis de costo-beneficio

A continuación se ilustra cómo aplicar la **herramienta de análisis de costo-beneficio** presentada en la descripción de la guía de campo, utilizando el estudio de caso del Líbano. Esta herramienta muestra cómo se pueden identificar, cuantificar y comparar sistemáticamente los costos y beneficios de un programa para evaluar su rentabilidad. El ejemplo proporciona una referencia práctica para los equipos que desean estimar la eficiencia económica de las intervenciones de comportamiento y tomar decisiones informadas sobre escalar, la adaptación o la asignación de recursos.

## Análisis de costo-beneficio

El objetivo principal de la intervención era aumentar la cobertura de vacunación oportuna entre los cuidadores de las clínicas urbanas de bajos recursos del Líbano. En ausencia de la intervención, los datos sugerían que muchos cuidadores retrasarían las visitas de seguimiento, con el riesgo de que se perdieran o se completaran las vacunas. Con un sencillo recordatorio de comportamiento, la postal tenía como objetivo cambiar esa trayectoria.

Los costos fueron relativamente modestos. Aunque el estudio no informó directamente sobre los datos financieros, una estimación razonable situó el

### Análisis de costo-beneficio (ACB)

Utilice esta hoja de trabajo para sopesar los costes y beneficios de su intervención y decidir si merece la pena escalarla, adaptarla o interrumpirla.

Intervención: **Tarjeta de recordatorio de citas**

#### Definir el objetivo

El objetivo del proyecto es mejorar la puntualidad de la vacunación infantil mediante una sencilla tarjeta postal recordatoria de bajo costo. Los cuidadores –especialmente las familias refugiadas y con bajo nivel de alfabetización– se beneficiaron de recibir un recordatorio claro de la próxima visita. Sin la intervención, las tasas de abandono y de no vacunación seguían siendo elevadas.

#### Lista de costes

- Directos: Diseño, impresión, formación y distribución de tarjetas (~0,20 \$/tarjeta).
- Indirectos: Tiempo del personal, supervisión y seguimiento.
- Oportunidad: Mínima, ya que la entrega se realizó durante las visitas existentes.
- Total: Estimado <\$10.000 para la fase piloto.

#### Lista de todos los beneficios

- Directos: Más niños vacunados a tiempo.
- Indirectos: Menos abandonos, menos citas perdidas, mejor compromiso de los cuidadores.
- Equidad: Mayor impacto entre los grupos sirios y con bajo nivel de alfabetización.

#### Asignar valores

- Costes valorados a través de las tasas de adquisición/personal.
- Beneficios: Aumento de 7 puntos porcentuales en las devoluciones a tiempo, ~350 niños vacunados adicionales.
- Valor aproximado por vacunación a tiempo: ~\$50-\$150 (basado en estimaciones de la OMS).

#### Comparación de costes y beneficios

- Coste por vacunación puntual adicional: ~\$28
- BCR estimado: ~3:1
- Fuerte retorno de la inversión.

#### Comprobar los supuestos

Las comprobaciones de sensibilidad mostraron que los resultados se mantenían:

- Menores tamaños del efecto (3-5%)
- Mayores costes de las postales
- Incluso en el peor de los casos, los beneficios superaron a los costes.

#### Formular un juicio

La intervención es rentable, ampliable y equitativa. Se recomienda escalar en entornos similares de bajos recursos, con circuitos de retroalimentación continuos para optimizar la prestación.

costo por postal impresa en menos de 0,20 dólares, incluyendo el diseño y la distribución. Suponiendo un alcance piloto modesto (por ejemplo, menos de 10,000 cuidadores), el costo total probablemente se mantuvo por debajo de los 10,000 dólares, incluyendo los materiales, la supervisión y el tiempo del personal.

En cuanto a los beneficios, el estudio demostró que la postal aumentó las tasas de retorno en siete puntos porcentuales. La literatura sobre salud pública sugiere que cada vacuna adicional administrada a tiempo contribuye a obtener beneficios a largo plazo en la prevención de enfermedades, la reducción de la mortalidad infantil y la disminución de los costos de atención médica. Según las estimaciones de la OMS, una valoración conservadora de cada vacuna administrada a tiempo podría suponer entre 50 y 150 dólares en beneficios sociales. Si se aplica a todo el grupo de intervención, esto se traduce en ganancias agregadas significativas, que probablemente arrojen una relación beneficio-costos de entre 3:1 y 5:1.

Para comprobar la solidez, el equipo consideró una serie de hipótesis. Incluso en escenarios pesimistas (por ejemplo, con costos más elevados o efectos más débiles), la intervención seguía pareciendo rentable, debido en gran medida a su bajo costo unitario y a su diseño escalable.

Es importante destacar que el equipo también tuvo en cuenta la equidad y la inclusión. La intervención benefició de manera desproporcionada a los cuidadores marginados, especialmente a las familias sirias, lo que pone de relieve su potencial como estrategia de bajo costo para reducir las disparidades en el acceso a la salud. La postal recordatoria no sólo era rentable, sino que también mejoraba la equidad, una consideración clave para su futuro escalamiento.

# Más información

Esta guía de campo ofrece herramientas prácticas, marcos y hojas de trabajo para ayudar a los equipos a aplicar las ciencias del comportamiento a los retos de desarrollo del mundo real. Sin embargo, ninguna guía puede abarcarlo todo. Las ciencias del comportamiento se encuentran en la intersección de múltiples disciplinas, que van desde el diseño centrado en el ser humano y la ciencia de la implementación, hasta la ética, la medición y la evaluación. Por eso hemos incluido esta sección, para aquellos que sienten curiosidad por profundizar, perfeccionar su práctica ética, fortalecer el diseño de la implementación o explorar cómo seleccionar mejores medidas de resultados. Los recursos que se presentan a continuación ofrecen puntos de partida seleccionados para un viaje de aprendizaje a su propio ritmo.

## «Quiero una guía paso a paso más detallada sobre cómo realizar experimentos».

Existen múltiples manuales, recursos y cursos sobre cómo realizar experimentos para programas sociales. Algunos recursos gratuitos útiles son el [sitio web de J-PAL sobre Introducción a las evaluaciones aleatorias](#) y la [Guía práctica de evaluación de impacto del Banco Mundial](#).

## «Quiero mejorar mi enfoque de la ética en las ciencias del comportamiento aplicada».

La ética es fundamental en cualquier investigación o proyecto de comportamiento que involucre a personas. Ya sea que la tarea en cuestión sea redactar formularios de consentimiento, evaluar riesgos o navegar por las dinámicas de poder, estos recursos brindan un apoyo accesible y práctico:

- [El kit de herramientas éticas para proyectos de ciencias del comportamiento aplicadas de UNICEF](#) ayuda a los equipos a reflexionar sobre los riesgos éticos desde el principio e integrar medidas de protección a lo largo de la implementación.
- [El Procedimiento sobre normas éticas en la investigación, la evaluación, la recopilación de datos y el análisis de UNICEF](#) describe los protocolos y expectativas oficiales de la organización.
- [La lista de verificación del consentimiento informado \(J-PAL\)](#) es una plantilla comentada con orientación sobre lo que debe incluirse en los formularios de consentimiento de los participantes.
- [Las plantillas de consentimiento de UNICEF](#) (véase la página 41) incluyen ejemplos editables para participantes, cuidadores y guardianes.

## «Necesito hacer un curso de ética para el IRB».

Hay múltiples sesiones de capacitación disponibles, y algunas organizaciones incluso ofrecen su propia capacitación interna en ética con un certificado reconocido por diferentes IRB. Para quienes busquen capacitación externa, exploren los siguientes recursos:

- [Formación en protección de la investigación en seres humanos del HHS](#) (con sede en EE. UU., certificación gratuita, ~5-6 horas)
- [Curso sobre política de los tres consejos sobre ética de la investigación](#) (con sede en Canadá, certificación gratuita, ~4 horas)

## «Quiero explorar cómo diseñar y medir la implementación de manera más eficaz».

Comprender *qué* se ha logrado exactamente, junto con *cómo*, es esencial para reconocer si una intervención de comportamiento ha funcionado. El [Repositorio de Resultados de Implementación](#) ofrece orientación y ejemplos para medir constructos como la viabilidad, la fidelidad y la aceptabilidad.

### «Quiero mejorar la forma en que selecciono o adapto las medidas de resultados».

Una medida de resultados sólida no solo evalúa la eficacia, sino que captura el comportamiento adecuado de la manera correcta. Si planea evaluar el cambio de comportamiento o los resultados indirectos, las [Propiedades psicométricas de las medidas de implementación](#) revisan la validez y fiabilidad de las herramientas más utilizadas en la ciencia de la implementación.

### «Quiero evaluar la calidad y el rigor de los informes y estudios de evaluación».

Si es responsable de revisar, encargar o interpretar estudios, es importante comprender no solo lo que dice un informe, sino también la fiabilidad de sus conclusiones. Estas herramientas y artículos ayudan a evaluar la calidad de los estudios, ya sea al revisar una evaluación de impacto, un informe de implementación o un artículo académico.

### Evaluación del diseño general y el rigor de los informes

- [La herramienta de evaluación DAC \(DAT\) de Gates](#) ayuda a evaluar si un estudio está bien diseñado, bien analizado y se comunica con claridad. Creada originalmente para ensayos clínicos, es aplicable a todos los sectores.
- El artículo [«Publicar artículos cuantitativos con rigor y transparencia»](#) ofrece una guía fácil de entender sobre la transparencia y la solidez para los equipos que redactan los resultados.

### Revisión de revisiones sistemáticas

[La herramienta de la evidencia Project Risk of Bias Tool](#) evalúa el rigor tanto en estudios aleatorios como no aleatorios. Es especialmente útil cuando se leen revisiones sistemáticas o síntesis de métodos mixtos.

### Evaluación del rigor de la investigación cualitativa

- [Indicadores de rigor en la investigación cualitativa](#) explican cómo juzgar la credibilidad, la transferibilidad y la fiabilidad de los estudios cualitativos.
- [El poder de la información en el muestreo cualitativo](#) ofrece una alternativa útil a la idea de «saturación» para justificar el tamaño de las muestras en las entrevistas.

## Recursos:

1. Akbari, M., Nikijoo, I., Khodapanah, B., Foroudi, P., & Padash, H. (2025). Forty Years of Microfinance Research and Its Impact on Consumers: A Review and Research Agenda Using the ADO-TCM Framework. *International Journal of Consumer Studies*, 49(4), e70101.
2. Behavioural Insights Team. How to Run Simple Behavioural Insight Projects. 2022. <https://www.bi.team/wp-content/uploads/2022/11/BIT-Handbook-How-to-run-simple-BI-projects.pdf>.
3. Blanc, J. (2014). *Microfinance, Debt and Over-Indebtedness: Juggling with Money*, Isabelle Guérin, Solène Morvant-Roux et Magdalena Villarreal (dir.). Editions Routledge, Londres, Royaume-Uni, 2014, 316 pages. *Revue internationale de l'économie sociale: recma*, (334), 122-124.
4. Bloomberg. "Big Money Backs Tiny Loans That Lead to Debt, Despair and Even Suicide." Bloomberg.com, May 3, 2022. <https://www.bloomberg.com/graphics/2022-microfinance-banks-profit-off-developing-world/>.
5. Clemens, Michael A., and Gabriel Demombynes. "When Does Rigorous Impact Evaluation Make a Difference? The Case of the Millennium Villages." *Journal of Development Effectiveness* 3, no. 3 (2011): 305–339. <https://doi.org/10.1080/19439342.2011.587017>.
6. Cristia, Julian, Pablo Ibararán, Santiago Cueto, Ana Santiago, and Eugenio Severín. "Technology and Child Development: Evidence from the One Laptop per Child Program." *American Economic Journal: Applied Economics* 9, no. 3 (2017): 295–320. <https://doi.org/10.1257/app.20150385>.
7. Evaluation Hub. "Run Evaluations." <https://www.bitevaluationhub.com/run-evaluations>.
8. John, B. (2024, November 14). Challenges and limitations of microfinance in achieving large-scale poverty reduction and job creation [Working paper].
9. J-PAL. "Design and Iterate the Implementation Strategy." <https://www.povertyactionlab.org/resource/design-and-iterate-implementation-strategy>.
10. J-PAL. "Ethical Conduct of Randomized Evaluations." <https://www.povertyactionlab.org/resource/ethical-conduct-randomized-evaluations>.
11. J-PAL. "Impact Evaluation Methods Table." <https://www.povertyactionlab.org/sites/default/files/research-resources/impact-evaluation-methods-table.pdf>.
12. J-PAL. "Power Calculations Exercise." [https://www.povertyactionlab.org/sites/default/files/Exercise-PowerCalcs\\_0.pdf](https://www.povertyactionlab.org/sites/default/files/Exercise-PowerCalcs_0.pdf).
13. J-PAL. "Questionnaire Piloting." <https://www.povertyactionlab.org/resource/questionnaire-piloting>.
14. J-PAL. "Data Security Procedures for Researchers." <https://www.povertyactionlab.org/resource/data-security-procedures-researchers>.
15. J-PAL and IPA. *Implementing Impact Evaluations: Case Study*. 2023. <https://poverty-action.org/sites/default/files/2023-03/Case-Study-Implementing-Impact-Evaluations.pdf>.
16. Karlan, Dean, and Jacob Appel. *More Than Good Intentions: Improving the Ways the World's Poor Borrow, Save, Farm, Learn, and Stay Healthy*. Penguin, 2011.
17. NYU Office of Research. "IRB Decision Tree." <https://www.nyu.edu/content/dam/nyu/research/documents/IRB/IRBDecisionTree.pdf>.

18. Shelly, Sarah, et al. "Improving Communication with Participants in Behavioural Trials." 2023.
19. UNICEF. An Evaluation of the PlayPump® Water System as an Appropriate Technology for Water, Sanitation and Hygiene Programmes. 2007. [http://www-tc.pbs.org/frontlineworld/stories/southernafrica904/flash/pdf/unicef\\_pp\\_report.pdf](http://www-tc.pbs.org/frontlineworld/stories/southernafrica904/flash/pdf/unicef_pp_report.pdf).
20. UNICEF. Ethical Considerations When Applying Behavioural Science to Programmes with Children. Innocenti, 2021. <https://www.unicef.org/innocenti/media/5186/file/UNICEF-Ethical-Considerations-Behavioural-Science-Children-2021.pdf>.
21. UNICEF. "UNICEF Procedure for Ethical Standards in Research, Evaluation, Data Collection and Analysis." <https://www.unicef.org/evaluation/documents/unicef-procedure-ethical-standards-research-evaluation-data-collection-and-analysis>.
22. UK What Works Evaluation Hub. "Pilot Impact Studies." <https://evaluationhub.eif.org.uk/pilot-impact-studies/>.
23. University of California Santa Barbara Library. "Data Evaluation Checklist." <https://www.library.ucsb.edu/sites/default/files/attachments/data-curation/resources/DataEvaluationChecklist.pdf>.
24. World Health Organization. Monitoring the Building Blocks of Health Systems: A Handbook of Indicators and Their Measurement Strategies. 2010. [https://iris.who.int/bitstream/handle/10665/44708/9789241502320\\_eng.pdf](https://iris.who.int/bitstream/handle/10665/44708/9789241502320_eng.pdf).
25. Australian Institute of Family Studies. Process Evaluation. 2025. [https://aifs.gov.au/sites/default/files/2025-03/2502%20EES%20process%20evaluation\\_1.pdf](https://aifs.gov.au/sites/default/files/2025-03/2502%20EES%20process%20evaluation_1.pdf).
26. Tableau. "What Is Data Cleaning?" <https://www.tableau.com/learn/articles/what-is-data-cleaning>.
27. BetterEvaluation. "Data Cleaning." <https://www.betterevaluation.org/methods-approaches/methods/data-cleaning>.