



EHU

Encuentro Besaldi | Evaluación de políticas públicas

Del ciclo de la evaluación a la inferencia causal

Ainhoa Vega

2026-09-01

¿Por qué evaluar?

Evaluar: una herramienta de buen gobierno

La **evaluación integral** de políticas públicas es un proceso sistemático y razonado de generación de conocimiento sobre el **diseño**, la **implementación** y los **efectos** de una intervención pública, encaminado a alcanzar un juicio valorativo basado en evidencias.

Ciclo de la evaluación

1. Pre-análisis y diagnóstico
2. Diseño de la intervención
3. Implementación
4. Medición de efectos
5. Aprendizaje y ajuste



Idea clave

La evaluación es una herramienta de mejora y de buen gobierno.
Evaluar también es aprender.

Para qué evaluamos

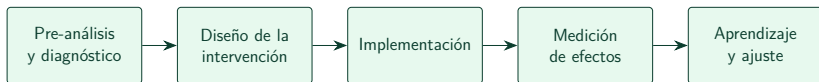
- ✓ Tomar decisiones basadas en evidencia
- € Usar los recursos de forma eficiente
- ⚖ Garantizar la rendición de cuentas
- 🔧 Escalar lo que funciona y corregir lo que no
- 🏛 Fortalecer la capacidad institucional de gestión



El dilema del decisor

Ante la escasez de tiempo y recursos, la respuesta pasa por revisar la literatura, estudiar el contexto de implementación y comparar opciones con un pre-análisis de coste-efectividad.

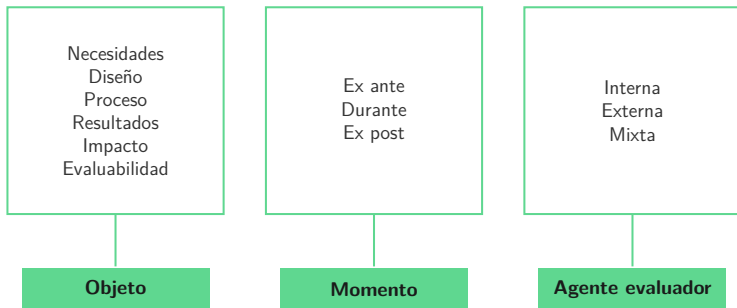
El ciclo de la evaluación



La evaluación integral **combina todos los tipos** de evaluación a lo largo del ciclo completo de la intervención.

Tipos y criterios

Tres ejes de la evaluación



Los criterios de la OCDE

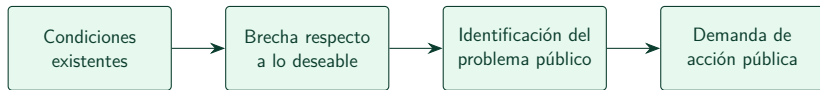
Criterio	Pregunta que responde
Pertinencia	¿El diseño responde al problema real?
Coherencia	¿Es compatible con otras intervenciones?
Eficacia	¿Se alcanzan los objetivos previstos?
Eficiencia	¿Se logran los resultados al menor coste posible?
Impacto	¿Qué cambios reales y duraderos produce?
Sostenibilidad	¿Los beneficios persisten cuando termina el apoyo?

! Distinción clave

Eficacia: ¿funciona? Eficiencia: ¿al coste adecuado? Impacto: ¿qué cambios duraderos produce?

Del problema a la intervención

Del problema público a la acción pública



Caracterizar bien el problema exige analizar sus **causas** (¿es multicausal?), la **población objetivo**, la **magnitud**, los **actores** y su **duración**.



Error frecuente

Aprobar medidas sin un diagnóstico fino lleva a políticas poco ajustadas al problema real.

Teoría del cambio y cadena causal

La **teoría del cambio** explica por qué y cómo se espera que una intervención produzca los cambios deseados. Se construye de atrás hacia adelante, partiendo del impacto final.

Insumos → Actividades →
Productos → Resultados → Impacto

Preguntas clave

- ¿Cuál es el impacto final deseado?
- ¿Qué resultados intermedios son necesarios?
- ¿Qué actividades los producen?
- ¿Qué supuestos deben cumplirse?
- ¿Cómo mediremos cada eslabón?

El **marco lógico** formaliza la teoría del cambio con indicadores verificables para cada nivel.

Indicadores: medir lo que importa

SMART

- ✓ **S**pecific
- ✓ **M**easurable
- ✓ **A**chievable
- ✓ **R**elevant
- ✓ **T**ime-bound

Tipos de indicadores

- **Insumo**: recursos utilizados
- **Proceso**: actividades ejecutadas
- **Producto**: bienes y servicios entregados
- **Resultado**: cambios a corto plazo
- **Impacto**: cambios estructurales



Precaución

Sin un sistema de indicadores definido antes de la intervención, la evaluación de resultados e impacto es muy difícil o imposible.

Los datos: sin datos no hay evaluación

Datos existentes

- Registros administrativos (MCVL, registros fiscales)
- Estadísticas oficiales (INE, Eustat, Eurostat)
- Seguimiento interno del programa

Datos nuevos

- Encuestas (pre/post, panel)
- Entrevistas y grupos focales
- Observación directa
- Experimentos de campo

! Regla de oro

Planificar la recogida de datos antes de la intervención es imprescindible para poder construir un contrafactual válido.

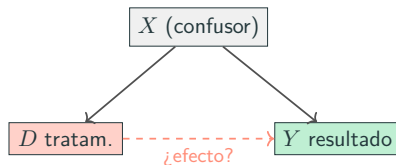
La evaluación causal de impacto

¿Qué es la inferencia causal?

La pregunta es si la política D **ha causado** el resultado Y , o si ambos simplemente están correlacionados.

El problema

La correlación no implica causalidad. Las comparaciones simples están contaminadas por el sesgo de selección.



El marco de resultados potenciales

Cada unidad i tiene dos resultados potenciales: Y_i^1 si recibe el tratamiento e Y_i^0 si no lo recibe. El efecto causal individual es $\tau_i = Y_i^1 - Y_i^0$ (Rubin, 1974).

Unidad	D_i	Y_i^0	Y_i^1	τ_i	Y_i^{obs}
Eneko	1	?	8	?	8
Miren	0	5	?	?	5
Julen	1	?	10	?	10
Jone	0	3	?	?	3

El problema fundamental

Solo observamos uno de los dos resultados potenciales. El contrafactual es siempre no observable.

El sesgo de selección

La diferencia de medias entre participantes y no participantes no identifica el efecto causal:

$$\underbrace{\bar{Y}_1 - \bar{Y}_0}_{\text{comparación}} = \underbrace{\tau}_{\text{efecto}} + \underbrace{\text{diferencias previas}}_{\text{sesgo de selección}}$$

Tres falsos contrafactuales

- ⚠ **Antes/después:** confunde tendencia temporal con efecto.
- ⚠ **Con/sin programa:** quien participa difiere de quien no participa.
- ⚠ **Predicción sin tratamiento:** depende por completo de los supuestos del modelo.

¿Qué podemos estimar?

ATE

$$\mathbb{E}[Y^1 - Y^0]$$

Efecto para una unidad aleatoria de la población.

ATT

$$\mathbb{E}[Y^1 - Y^0 \mid D = 1]$$

Efecto para quienes reciben el tratamiento.

LATE

$$\mathbb{E}[Y^1 - Y^0 \mid C]$$

Efecto local, para los cumplidores.

La **pregunta política** decide el estimando: ATE si el programa es de acceso general, ATT si evaluamos un programa ya implantado.

Panorama de métodos contrafactuales

Método	Supuesto clave	Cuándo usarlo
RCT	Asignación aleatoria	Cuando es posible aleatorizar
Regresión	No hay confusores inobservables (CIA)	Datos ricos en covariables
Matching / IPW	CIA y soporte común	Observacional, CIA plausible
RDD	Continuidad en el umbral	Política asignada por umbral
Instrumentos	Relevancia y exclusión	Experimento natural disponible
DiD	Tendencias paralelas	Panel con cambio de política
Control sintético	Tendencias pre-tratamiento	Un solo tratado, muchos controles

Idea clave

Cada método compra identificación a cambio de un supuesto. El supuesto es el que hay que discutir, no el estimador.

Diferencias en diferencias

Compara el **cambio** en el resultado del grupo tratado con el **cambio** en el grupo de control a lo largo del tiempo.

$$DiD = (\Delta Y^T) - (\Delta Y^C)$$

Requiere datos **longitudinales**, con observaciones pre y post-tratamiento.

Supuesto clave

Tendencias paralelas: en ausencia de tratamiento, la diferencia entre los grupos se mantiene constante en el tiempo.

Emparejamiento (matching)

Empareja a cada individuo tratado con individuos no tratados de características observables X similares, estimando la probabilidad de recibir el tratamiento (**propensity score**) y emparejando en base a ella. Se puede aplicar con datos de **sección cruzada**.

Supuestos clave

Independencia condicional: la selección al tratamiento depende solo de las X observables.

Soporte común: deben existir controles similares a los tratados.

DiD y matching: qué controla cada uno

	Dif. en diferencias	Matching (PSM)
Tipo de datos	Longitudinales (pre y post)	Sección cruzada o longitudinales
Qué controla	Diferencias inobservables constantes en el tiempo	Diferencias observables que causan la selección
Supuesto clave	Tendencias paralelas	Independencia condicional
Contrafactual	Tendencia temporal del grupo de control	Individuos de control más similares a los tratados



En la práctica

A menudo se combinan (DiD-matching): igualar primero las características observables y aplicar después DiD sobre los grupos emparejados.

De la evidencia a la decisión

De la evaluación a la decisión

La evaluación solo tiene valor si sus resultados se usan:

-  Informar el rediseño del programa
-  Justificar continuación, escala o cierre
-  Orientar la asignación presupuestaria
-  Alimentar la rendición de cuentas
-  Generar evidencia transferible a otros contextos

Marco institucional

La Ley 27/2022 institucionaliza la evaluación, crea el IEPP y obliga a evaluación ex ante y ex post en las políticas de especial impacto.

Desafío central (OCDE, 2023)

La brecha entre producir evaluaciones y utilizarlas en la toma de decisiones sigue siendo el principal reto de los sistemas de evaluación.

El ecosistema vasco: universidad y Besaldi

La alianza entre la universidad pública y el órgano de evaluación aporta lo que ninguna de las dos partes tiene por separado: **independencia técnica y capacidad metodológica** de un lado, **conocimiento operativo y acceso a la información** del otro.

-  **Acceso protocolizado a los microdatos:** vías estables y seguras a los registros administrativos enlazados, con las garantías de confidencialidad correspondientes.
-  **Evaluabilidad desde el diseño:** definir cómo se evaluará una política antes de aprobarla, con grupo de comparación y línea de base.
-  **Cultura de uso y de publicación:** publicar también los resultados nulos o incómodos, porque informan la decisión igual que los favorables.

La evaluación en la inclusión y el empleo

- ✂ **Heterogeneidad:** detrás de las cifras agregadas hay trayectorias muy distintas, y la media esconde a quienes peor están.
- 🔍 **Acceso:** el diseño formal de una prestación y su alcance real difieren, y las barreras administrativas explican buena parte de esa distancia.
- 🔧 **Brecha entre evaluar y decidir:** el principal reto de los sistemas de evaluación.

! Cuantitativo y cualitativo

Lo cuantitativo mide cuánto y para quién. Lo cualitativo explica por qué y cómo. La voz de las personas usuarias, expertas por experiencia, constituye evidencia y debe tratarse como tal.

Gracias

Ainhoa Vega
ainhoa.vega@ehu.eus

Euskal Herriko Unibertsitatea (EHU)

Referencias

Referencias

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics*. Princeton University Press.

Card, D., & Krueger, A. B. (1994). Minimum Wages and Employment. *American Economic Review*, 84(4), 772–793.

Ministerio de Hacienda. Definición de evaluación integral de políticas públicas.

OCDE (2023). *Evaluation systems and the use of evidence in decision making*.

Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score. *Biometrika*, 70(1), 41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701.

Vega, A., & Artabe, A. (2026). *Introducción a la evaluación cuantitativa de políticas públicas*. Euskal Herriko Unibertsitatea.