



Universidad Autónoma  
de Madrid



# Jornada sobre Evaluación del Impacto Social: Introducción

**Miguel A. Sorrel** ([miguel.sorrel@uam.es](mailto:miguel.sorrel@uam.es))  
Universidad Autónoma de Madrid,

25/septiembre/2025  
Instituto de Mayores y Servicios Sociales



# Objetivos de la jornada

**Capacitar** de forma **introductoria y básica** en la **comprensión de conceptos clave y aplicación práctica** de los **métodos** para la **Evaluación del Impacto Social** en proyectos, programas y políticas

## Ejemplo práctico para el curso:

### Programa de Termalismo Social del IMSERSO 2024

**Contexto:** En 2024, el IMSERSO ofertó unas **192.000 plazas de termalismo social** en cerca de 90 balnearios de toda España. Este programa está dirigido a personas mayores, con el objetivo de **mejorar la salud, prevenir la dependencia y favorecer la socialización.**

En el [informe relativo a la revisión del Programa de termalismo del Imsero realizado en 2023](#) se identifican como ámbitos de mejora:

- 1. Carencia de un marco teórico/conceptual** que sustente el Programa
- 2. Existencia de una alta carga de gestión** para la implementación del Programa

1

## Introducción

10:00-11:00

Definiciones y marco

2

## Métodos

11:00-12:00

Diseño e indicadores

3

## Análisis

12:30-13:30

Análisis de datos

4

## Presentación

13:30-14:30

Informes técnicos



Coffee break: 12:00-12:30

1

Introducción

2

Métodos

3

Análisis

4

Presentación

## ¿Por qué es necesaria la **Evaluación del Impacto Social**?

*“Lo que no se mide no se puede mejorar...  
y lo que no se mejora se deteriora”*

Algunas respuestas que podemos dar:

1. Demostrar el **valor real** de los proyectos y políticas: Evitar medir sólo lo económico (#plazas, #usuarios), ¿qué ha cambiado en la vida de las personas (calidad de vida, inclusión)?
2. Rendición de cuentas y **transparencia**: Compromiso con la ciudadanía y los financiadores
3. Fortalecer la **participación** y el compromiso social: Motiva a los actores sociales (usuarios, familias, asociaciones) a involucrarse activamente
4. Aprendizaje y **mejora**: No se trata solo de “justificar”, sino de **aprender qué funciona y qué no**

# ¿Por qué es necesaria la **Evaluación del Impacto Social**?

## **Ejemplo práctico: Programa de Termalismo\***

### **1 Demostrar el valor real de los proyectos y políticas**

Si solo contamos “número de plazas ocupadas” o “número de balnearios participantes”, sabemos cuánto se hizo... **pero no si mejoró la vida de la gente.**

Al medir impacto (encuestas de salud, soledad percibida, bienestar emocional), descubrimos que:

**El 65 % de participantes reporta mejoras en su salud física.**

**El 70 % dice haberse sentido acompañado y menos solo** tras la experiencia.

 Esto demuestra el **valor real del programa**, más allá de lo económico.

*\*Datos ficticios*

# ¿Por qué es necesaria la **Evaluación del Impacto Social**?

## **Ejemplo práctico: Programa de Termalismo\***

### **2 Rendición de cuentas y transparencia**

El programa implica **fondos públicos**, y la ciudadanía tiene derecho a saber:

- ✓ ¿En qué se gasta?
- ✓ ¿A quién beneficia?
- ✓ ¿Qué cambios genera en la vida de las personas mayores?

Al presentar resultados claros (“7 de cada 10 participantes mejoraron su bienestar”), **rendimos cuentas de forma transparente** y refuerza la confianza social en las políticas públicas.

*\*Datos ficticios*

# ¿Por qué es necesaria la **Evaluación del Impacto Social**?

## **Ejemplo práctico: Programa de Termalismo\***

### **3 Fortalecer la participación y el compromiso social**

Si se mide el impacto, también se puede **dar voz a los mayores beneficiarios**.

Ejemplo: incluir testimonios cualitativos como *“En el balneario hice amistades que mantengo todo el año”*.

Esto hace que los mayores se sientan **parte activa** del programa, no solo receptores. Además, el sector turístico local (balnearios, hoteles, transporte) se ve implicado y se compromete más al ver que el impacto es **visible y positivo**.

# ¿Por qué es necesaria la **Evaluación del Impacto Social**?

## **Ejemplo práctico: Programa de Termalismo\***

### **4** **Aprendizaje y mejora**

La medición permite detectar **qué funciona y qué no**.

Ejemplo:

- Se observa que los participantes con movilidad reducida **aprovechan menos las actividades** → Plan de acción: adaptar mejor las instalaciones o diseñar talleres específicos.
- En algunas zonas con baja ocupación → se puede rediseñar la oferta o reforzar la comunicación.

Así, **aprendemos y mejoramos** el programa año tras año.

*\*Datos ficticios*

## Definiciones

**Valoración-Assessment** (personas) / **Evaluación-Evaluation** (programas):

*“Proceso de determinar en qué medida los **objetivos** de un programa se han alcanzado”*

*“Examen de los **efectos**-outputs de un programa”*

*“investigación sistemática, científica”*

**Programa:** *“los **sistemáticos** esfuerzos realizados para lograr objetivos **preplanificados** con el fin de mejorar la salud, el conocimiento, las actitudes y la práctica”*

## Definiciones

La **Evaluación del Impacto Social** (SIA por sus siglas en inglés) es...

- un **proceso sistemático y participativo**, ...
- que identifica, **analiza y valora los efectos**, tanto positivos como negativos, que un proyecto, política o programa puede generar **sobre las comunidades, grupos sociales y su entorno**, ...
- con el fin de informar la **toma de decisiones** y diseñar medidas para maximizar beneficios y minimizar impactos adversos.

1

## Observación / Identificación del proyecto y contexto

Igual que en el método científico, se parte de observar el “fenómeno”: se describe el proyecto, su entorno y los grupos sociales afectados.

2

## Formulación de hipótesis / Identificación de impactos

Se anticipan los posibles impactos sociales, positivos y negativos, como si fueran hipótesis sobre cómo afectará el proyecto a la comunidad.

3

## Experimentación / Análisis y predicción

Se recogen datos mediante encuestas, entrevistas o estadísticas para evaluar la magnitud y relevancia de los impactos, similar a probar la hipótesis con experimentos.

4

## Conclusiones y gestión / Mitigación y monitoreo

Se diseñan medidas para maximizar beneficios y minimizar daños, y se realiza seguimiento para ajustar estrategias, equivalente a sacar conclusiones y replicar experimentos si es necesario.

**LISTADO DE CUESTIONES RELEVANTES EN EVALUACIÓN DE PROGRAMAS**

- I. Evaluación de necesidades** (ej., *Se han realizado una serie de **sondeos de opinión** para consultar a la población objetivo y a la población general acerca de cuáles son sus necesidades*)
- II. Definición de objetivos** (ej., *Los objetivos se expresan en criterios e indicadores **cuantificables***)
- III. Criterios de selección del programa** (ej., *Existen **datos empíricos** que avalan la validez del programa en el **contexto** de aplicación*)
- IV. Definición del programa** (ej., *La **estructura material** concerniente al programa está definida*)
- V. Grado de implantación** (ej., *Se sigue la **temporalización** fijada en las unidades programadas*)
- VI. Diseño de la evaluación** (ej., *En la evaluación del programa se puede utilizar un **diseño experimental***)
- VII. Calidad de las operaciones a observar** (ej., *Las operaciones que van a permitir la evaluación del programa pueden serlo mediante **instrumentos fiables y válidos** ya existentes*)
- VIII. Recogida de información sobre el contexto** (ej., *Es posible acudir a **datos** procedentes de procedimientos retrospectivos sobre el contexto donde se aplica el programa*)
- IX. Aceptabilidad de la evaluación** (ej., *Antes de empezar la evaluación, los **implicados** en el programa han colaborado con el equipo evaluador*)
- X. Implicación del evaluador** (ej., *El evaluador es **externo** a la Organización que ha planificado e implementado el programa*)
- XI. Finalidad de la evaluación** (ej., *Está claramente establecido cuál es el **propósito** de la evaluación*)
- XII. Costes de la evaluación** (ej., *Se puede determinar la **duración** de la evaluación del programa*)

Ballesteros (1995)

## 📌 Ejemplo práctico: Programa de Termalismo\*

Dimensión LCREP	Situación en el programa
<b>I. Evaluación de necesidades</b>	<ul style="list-style-type: none"> <li>✓ Se analizaron datos epidemiológicos sobre artrosis y dolencias crónicas.</li> <li>✗ Pero no se consultó directamente a la población objetivo mediante encuestas o grupos focales.</li> </ul>
<b>II. Definición de objetivos</b>	<ul style="list-style-type: none"> <li>✓ Hay metas sobre acceso a plazas.</li> <li>✗ Los objetivos se plantean de forma general (“mejorar la salud”), sin indicadores cuantificables.</li> </ul>
<b>III. Criterios de selección del programa</b>	<ul style="list-style-type: none"> <li>✓ Experiencias previas en otros contextos avalan la validez del programa.</li> <li>✗ Faltan estudios actualizados en España tras la pandemia.</li> </ul>
<b>IV. Definición del programa</b>	<ul style="list-style-type: none"> <li>✓ Acciones y temporalización bien descritas.</li> <li>✗ Recursos humanos no están especificados con detalle (ej. ratio profesionales/usuarios).</li> </ul>
<b>V. Grado de implantación</b>	<ul style="list-style-type: none"> <li>✓ La mayoría sigue la planificación anual.</li> <li>✗ En algunos balnearios se reportan retrasos y cancelaciones.</li> </ul>
<b>VI. Diseño de la evaluación</b>	<ul style="list-style-type: none"> <li>✓ Solo cuestionarios de satisfacción.</li> <li>✗ No hay diseño preestablecido. No se plantean grupos de comparación.</li> </ul>

\*Datos ficticios

## Ejemplo práctico: Programa de Termalismo\*

Dimensión LCREP	Situación en el programa
<b>VII. Calidad de las operaciones a observar</b>	<ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Existen indicadores (dolor percibido, movilidad).</li> <li><input type="checkbox"/> Pero los instrumentos no son siempre validados ni aplicados de forma homogénea.</li> </ul>
<b>VIII. Recogida de información sobre el contexto</b>	<ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Se tienen datos de plazas ocupadas por región.</li> <li><input type="checkbox"/> No se recoge información sistemática del perfil socioeconómico de los participantes.</li> </ul>
<b>IX. Aceptabilidad de la evaluación</b>	<ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Los usuarios valoran positivamente dar feedback.</li> <li><input type="checkbox"/> Los gestores muestran cierta resistencia a evaluaciones externas.</li> </ul>
<b>X. Implicación del evaluador</b>	<ul style="list-style-type: none"> <li><input type="checkbox"/> La evaluación está a cargo de personal interno del IMSERSO, no independiente.</li> </ul>
<b>XI. Finalidad de la evaluación</b>	<ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Se menciona de forma genérica la “mejora del programa”.</li> <li><input type="checkbox"/> No está claramente definido cómo se usarán los resultados.</li> </ul>
<b>XII. Coste de la evaluación</b>	<ul style="list-style-type: none"> <li><input checked="" type="checkbox"/> Se ajusta a presupuesto global del IMSERSO.</li> <li><input type="checkbox"/> No se definen costes específicos.</li> </ul>

\*Datos ficticios

## Conceptos

### **Parte 2: Métodos:**

*Evaluación de necesidades, Definición de objetivos, Definición del programa (acciones, RRHH, materiales, temporalización), Diseño de la evaluación, Calidad de las observaciones (fiabilidad y validez)*

### **Parte 3: Análisis:**

*Análisis de datos cuantitativos, Análisis de datos cualitativos*

### **Parte 4: Presentación:**

*Prácticas cuestionables*

1

Introducción

2

Métodos

3

Análisis

4

Presentación



## Quando evaluamos...

- ...**hacemos supuestos** (¿Están justificados o pueden sesgar la mirada?)
- ...**miramos desde un punto de vista** (¿Gestores, usuarios? ¿Faltan voces?)
- ...**tenemos un propósito** (¿Rendición de cuentas, mejora, aprendizaje?)
- ...**respondemos preguntas clave** (¿Impacto, eficacia, eficiencia, satisfacción?)
- ...**usamos información** (¿Es fiable, suficiente, con alternativas explicativas?)
- ...**traducimos en indicadores y conceptos** (¿bien definidos o forzados?)
- ...**interpretamos resultados** (¿Apoyan hipótesis? ¿Reconocemos límites?)
- ...**con consecuencias reales** (¿Qué cambios o decisiones se derivan?)

## Conceptos clave en diseño

- **Teoría:** Explicación amplia que orienta cómo y por qué un programa debería producir cambios (ej. el envejecimiento activo mejora bienestar y participación social)
- **Hipótesis:** Supuesto concreto que podemos comprobar en la evaluación (ej. la participación en programas de termalismo reducirá la soledad percibida en mayores)
- **Variable:** Característica que podemos medir y que puede cambiar (ej. nivel de satisfacción, frecuencia de actividades sociales, estado de salud)
- **Variable dependiente (VD):** Lo que queremos explicar o mejorar con el programa (ej. reducción de la soledad, mejora de la calidad de vida)
- **Variable independiente (VI):** Lo que el programa introduce y que pensamos que influirá en la VD (ej. número de sesiones de termalismo, actividades de ocio, apoyo profesional)

## Conceptos clave en diseño

- *Variables que influyen en la relación VI → VD*

### Variable extraña / Confound

Factores externos que también pueden afectar a la VD y deben ser controlados.

*Ejemplo:* el estado de salud previo puede influir tanto en la participación en el programa (VI) como en la calidad de vida (VD).

### Mediador (M)

Variable que explica el mecanismo por el que la VI afecta a la VD:

**VI → M → VD**

*Ejemplo:*

- VI: Participación en talleres de envejecimiento activo
- M: Aumento de redes sociales
- VD: Reducción de la soledad percibida

## Conceptos clave en diseño

- ***Variables que influyen en la relación VI → VD***

### **Moderador**

Factor que modifica la dirección o magnitud de la relación entre VI y VD.

*Ejemplo:*

- VI: Participación en actividades culturales
- Mod: Nivel socioeconómico
- VD: Bienestar percibido

*(El impacto puede ser mayor en personas con menos recursos previos)*

- **Causalidad**

Podemos hablar de causalidad si:

- VI y VD covarían (se mueven juntas)
- La VI antecede en el tiempo a la VD
- Se descartan causas alternativas (controlando variables extrañas)

## Fuentes documentales

- **Fuentes Generales:** Bases de datos (e.g., PsychInfo, Web of Knowledge)
- **Fuentes Secundarias:** Revistas que publiquen revisiones (e.g., Annual Review of Psychology)
- **Fuentes primarias:** Revistas que incluyen artículos originales (e.g., Psicothema)

Web of Science 

Herramientas ▾ Búsquedas y alertas ▾ Historial de búsqueda Lista de registros marcados

Seleccionar una base de datos Colección principal de Web of Science ▾

Búsqueda básica Búsqueda de autores <sup>BETA</sup> Búsqueda de referencia citada Búsqueda avanzada Búsqueda de estructura

Burnout × Título ▾

And ▾ 2018-2020 × Autor ▾ **Buscar** Sugerencias de búsqueda

Seleccionar del índice + Agregar fila | Restablecer

ANÁLISIS | 1

## Un premio Nobel contra la falacia que una correlación y causalidad

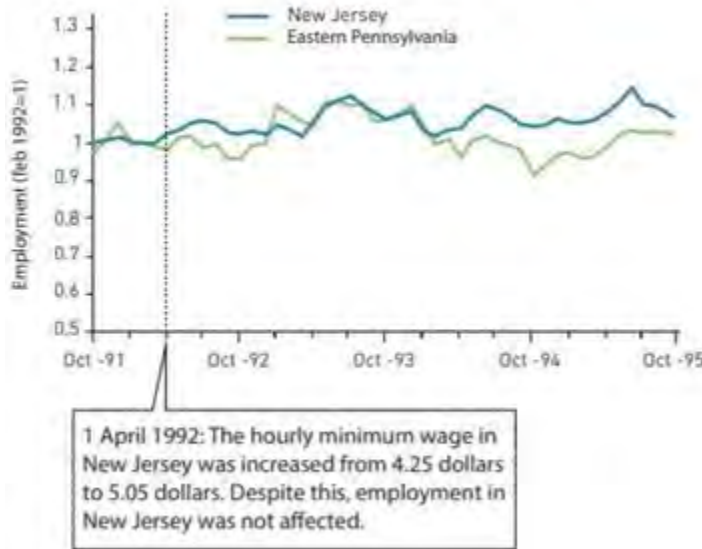
Los economistas David Card, Joshua Angrist y Guido Imbens han obtenido el galardón por hacer más científica la economía, inventando trucos para atacar su pregunta más difícil: qué causa qué



“Si ocurre después de esto, por tanto ocurre a causa de esto” (post hoc ergo propter hoc)

...o no

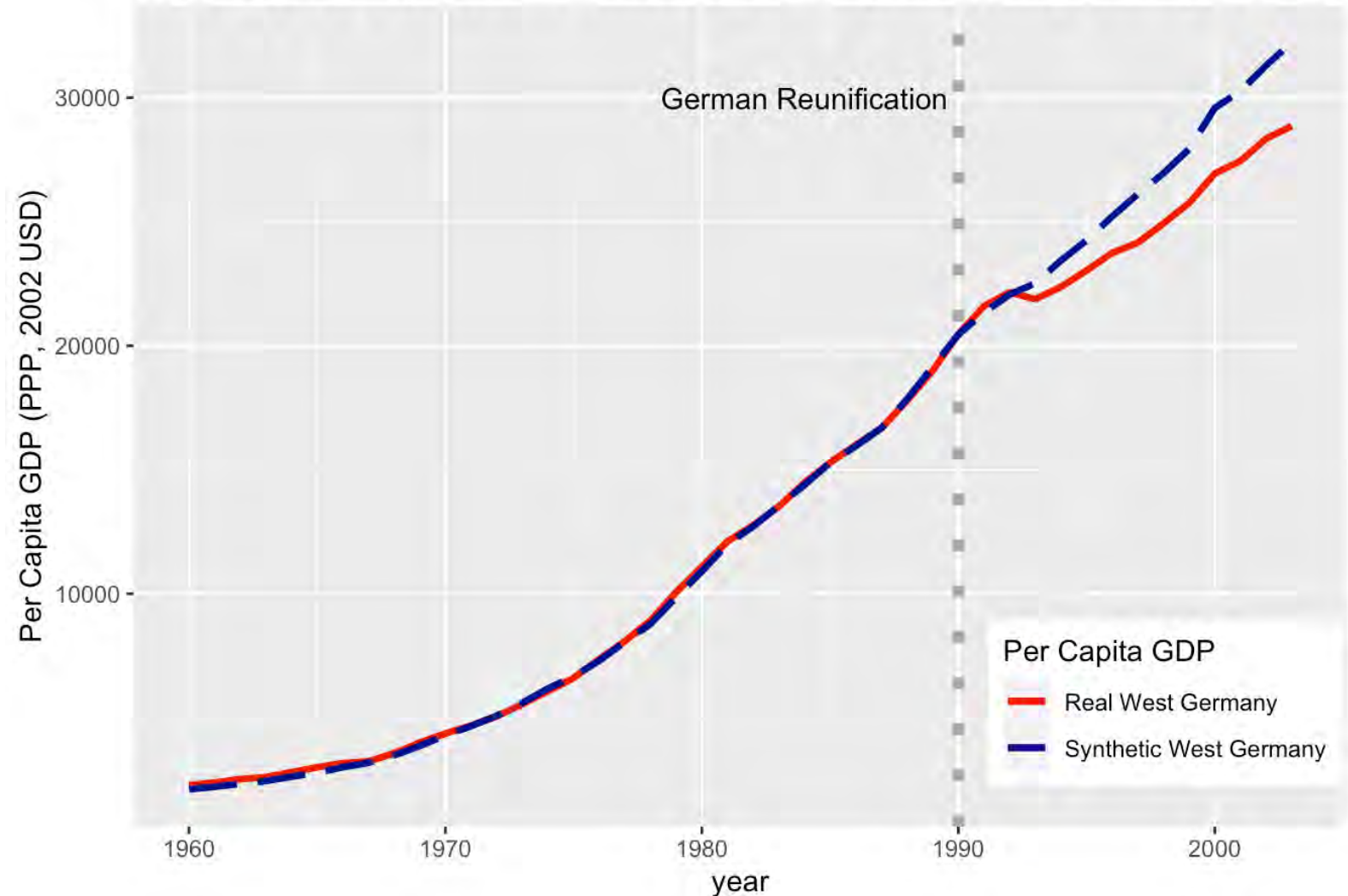
¿Qué efecto tiene vacunar a una persona?  
 ¿Alguna carrera elegida lleva a más años en el paro?  
 ¿Subir el salario mínimo destruye empleos?



Fuente: <https://www.nobelprize.org/prizes/economic-sciences/2021/popular-information/>

Exploratory Data Analysis V - Reproducing German Reunification with Generalized Synthetic Control Method (GSCM)  
<https://rpubs.com/jiangzm/1053384>

Trends in Per Capita GDP: Real West Germany and West Germany under GS



Metodologías cualitativas	Dimensión	Metodologías cuantitativas
Subjetivismo, constructivismo	Epistemología	Objetivismo, positivismo
Describir (desde dentro): Etnografía Comprender: Estudio de casos Transformar: Investigación-Acción En todos los casos, Interpretar	Objetivos / Planes de investigación	Describir (desde fuera): Observación y encuesta Clasificar: "ex post facto" retrospectivo Predecir: "ex post facto" prospectivo Explicar: experimentos Intervenir: cuasi-experimentos
Observación participante Entrevistas no estructuradas Cuestionarios abiertos Narraciones personales Análisis de documentos	Recogida de evidencia empírica	Observación sistemática Entrevistas estructuradas Cuestionarios estandarizados Tests Tareas estructuradas
Muestras no probabilísticas Análisis cualitativo de datos Rechazo de la estadística	Selección y análisis de datos	Muestras probabilísticas Análisis cuantitativo de datos Técnicas estadísticas de análisis
Implicación del investigador Confianza en los resultados Seguridad en las conclusiones Triangulación	Valoración de la investigación	Fiabilidad Validez interna y externa Sensibilidad Generalizabilidad

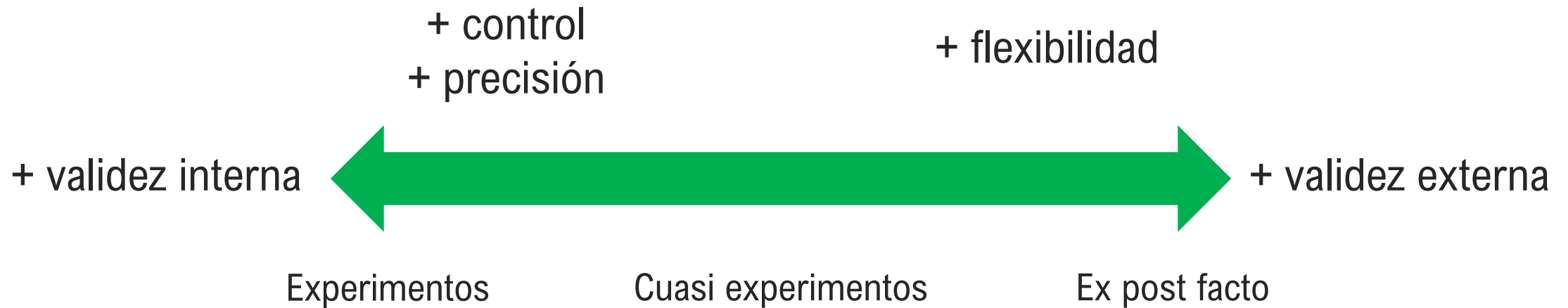
## Validez interna

- Se refiere a la seguridad que tenemos acerca de que los cambios en VD se deben a VI. Hace referencia al grado de **control** que se establece en el diseño
- Una alta validez interna permite hacer inferencias de los resultados en términos de **causalidad** con un alto grado de seguridad. Esto requiere además:
  - **Covariación**: un cambio en la VI se corresponde con un cambio en la VD.
  - **Antecesión**: la VI se produce antes de la VD.
  - **Rechazo de explicaciones alternativas plausibles**: las diferentes variables extrañas y enmascaradas han sido controladas

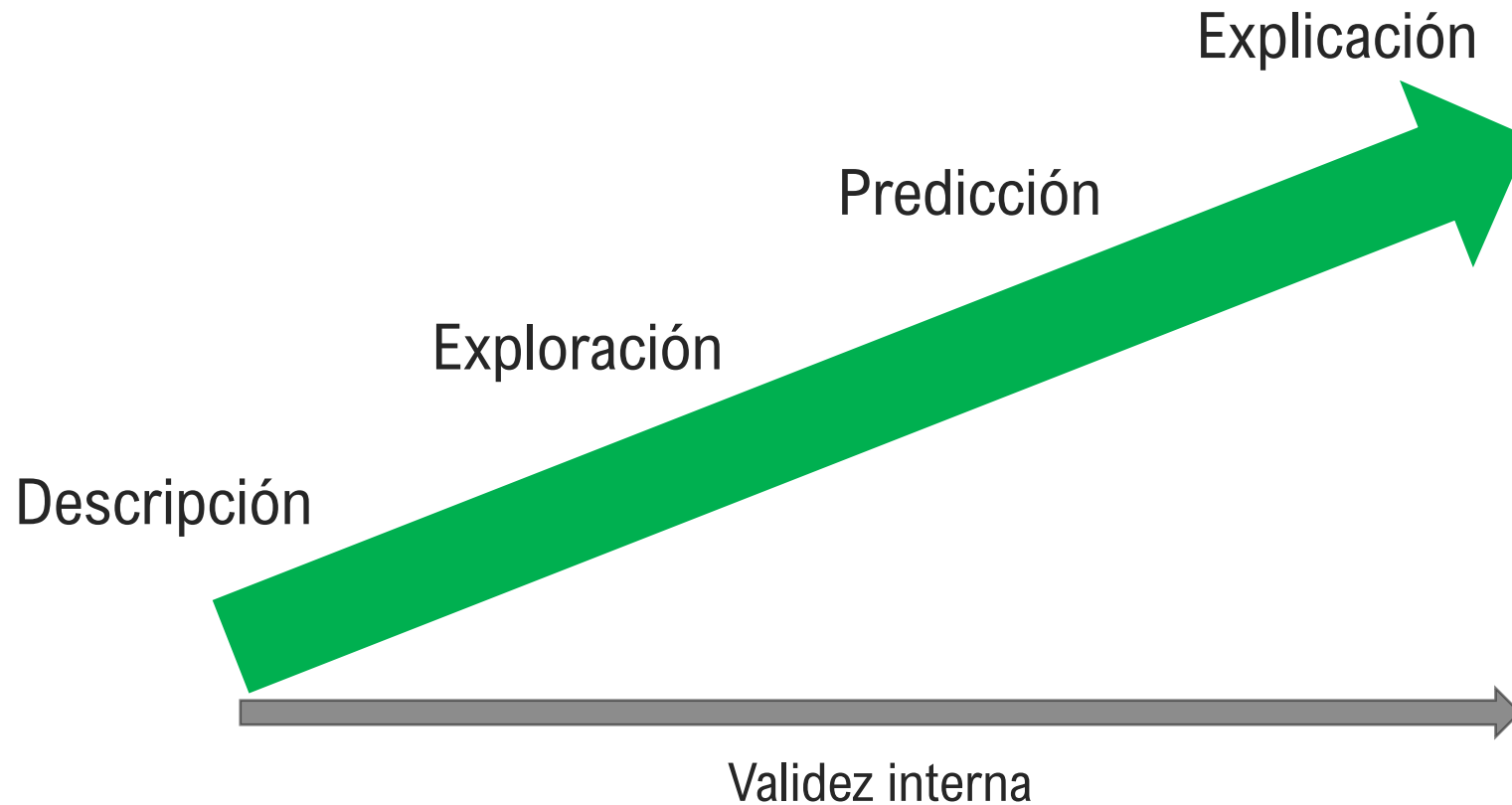
## Validez externa

- La validez externa hace referencia al poder de **generalización** de los resultados de la investigación (a más valores de la VI, más personas, más circunstancias y más momentos).

# Compromiso entre validez interna y validez externa



# Las conclusiones de la investigación



# Tipos de diseño

	¿MANIPULACIÓN DE VARIABLES?	¿ALEATORIAMENTE?
1. Diseño experimental	<b>SÍ</b>	<b>SI</b>
2. Diseño cuasi-experimental	<b>SÍ</b>	NO
3. Diseño ex post facto	NO	NO
4. Observación	NO	NO

# 1. Diseño experimental

## CONTROL

- Los experimentos se realizan en **situaciones controladas**
- Se detectan y controlan variables extrañas

## MANIPULACIÓN

- El investigador **manipula la VI** para contrastar si influye en la VD
- Distintos niveles de la VI se aplican a diferentes grupos (inter grupos)
- **Grupo experimental:** recibe el efecto de un nivel de la VI. Puede haber varios grupos
- **Grupo control:** no recibe el efecto de la VI. Funciona como línea base.

## CAUSALIDAD

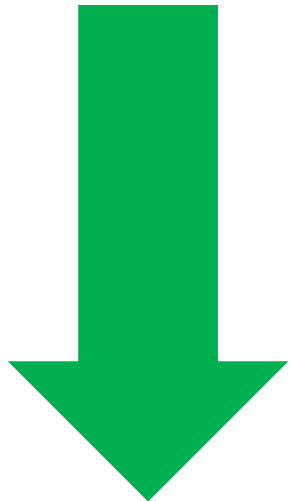
- Un experimento proporciona un grupo de mediciones de la VD bajo cada una de las condiciones de la VI
- Gracias al control podría afirmarse causalidad

Inconvenientes: Baja validez externa y en ocasiones difíciles de aplicar

# 1. Diseño experimental (Cont.)

**Para controlar las variables extrañas, ¿mismos sujetos o distintos?**

Grado de control:



(1) Diseños intra-sujetos

(2) Diseños con sujetos genéticamente iguales (gemelos, mellizos)

(3) Diseños inter-sujeto (con bloqueo)

# 1. Diseño experimental (Cont.)

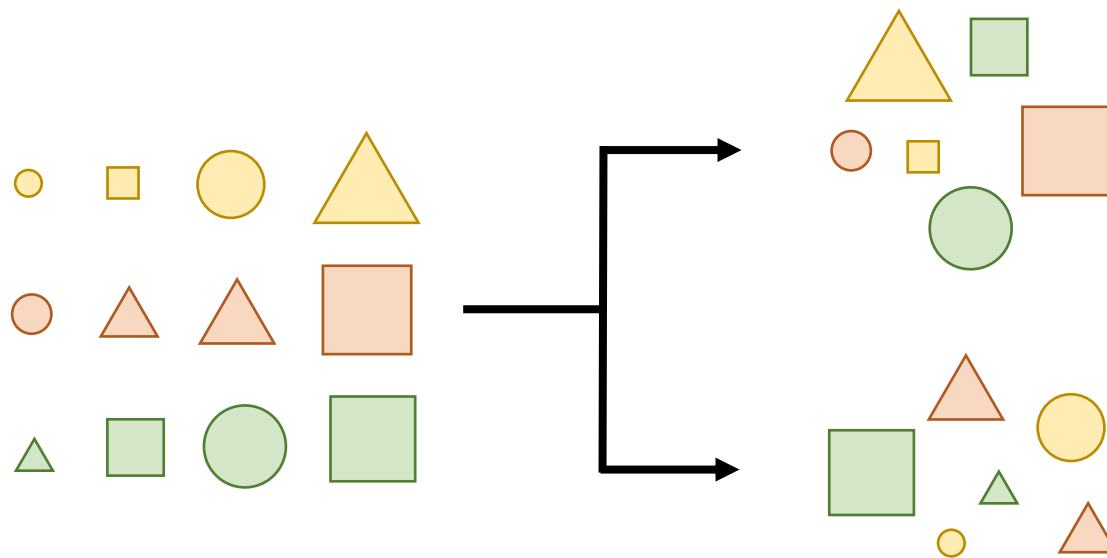
## (1) Intra-sujeto

- Todos los niveles de la VI se aplican a todos los participantes (**intra grupos**)
- Los grupos de **medidas** son **repetidas** (un mismo individuo aporta valores en los diferentes niveles de la VI)
- Debe controlarse:
  - **Efecto del aprendizaje** (dependerá de la VD)
  - **Efecto de la fatiga** (longitud del experimento)
  - **Efecto de la motivación** (tareas repetitivas)
  - **Efecto de la práctica** (orden de presentación: aleatorizar, reequilibrado [completo: ABBA; incompleto: grupo 1: AB; grupo 2: BA])
  - **Efecto de la persistencia** (establecer periodos de extención)

# 1. Diseño experimental (Cont.)

## (3) Intersujeto: Diseño de grupos aleatorios

Los individuos son **asignados al azar** a cada uno de los niveles de la VI para controlar por equilibrado las variables extrañas que están presentes en distinto grado en los participantes



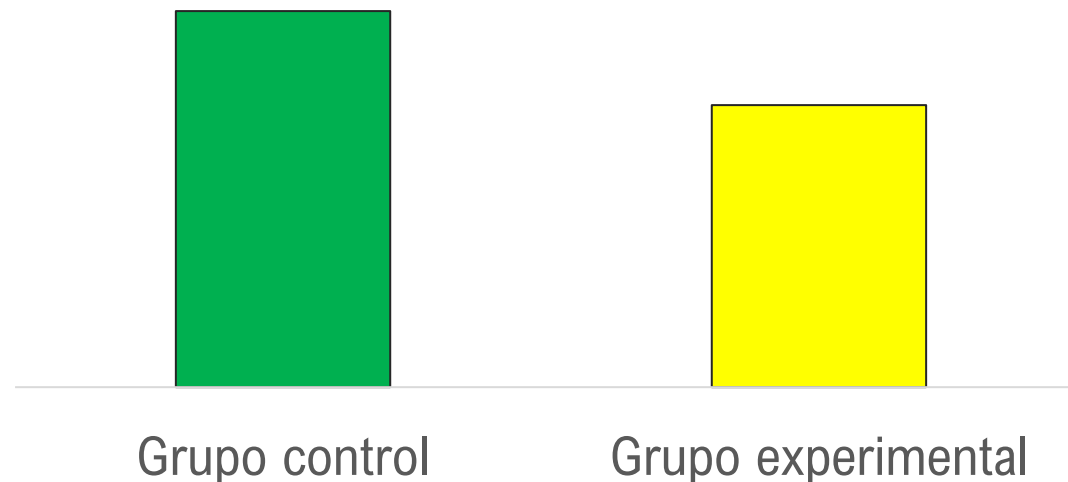
# 1. Diseño experimental (Cont.)

## (3) Intersujeto: Diseño de grupos aleatorios con bloques

- Son interesantes cuando la muestra es pequeña y conocemos alguna **variable extraña** que pueda estar relacionada con la VD para asegurar el equilibrado de dicha variable en los grupos experimental y control
- Procedimiento:
  1. Se mide la variable extraña que deseamos controlar
  2. Se realizan tantos bloques como se quiera de participantes con similares valores en la variable que se bloquea (e.g., bajo, medio, alto)
  3. Se asignan al azar a los participantes de cada bloque a cada uno de los grupos

# 1. Diseño experimental (Cont.)

- Efecto placebo: efecto de mejora que se da siempre por el mero hecho de someterse a un tratamiento (sugestión). Es por ello que es fundamental incluir un **grupo control** como línea base para valorar cualquier intervención
- Efecto nocebo: ¡el grupo control podría empeorar! Debe ser un grupo control **activo**



# 1. Diseño experimental (Cont.)

## Diseños factoriales

- Son diseños que incluyen más de una VI. Se crean tantos grupos como posibles combinaciones de las Vis: 2x2, 3x2, 4x2x2, ...
- En estos casos además del **efecto principal** de cada VI, interesa explorar si existe **interacción** entre las VIs
- Un diseño 2x2 sería este:

		VI 2	
		Nivel 1 VI 2	Nivel 1 VI 2
VI 1	Nivel 1 VI 1	Grupo 1	Grupo 2
	Nivel 1 VI 2	Grupo 3	Grupo 4

# 1. Diseño experimental (Cont.)

## 📌 Ejemplo práctico: Programa de Termalismo\*

### Variables

• **Variable Independiente (VI):** Participación en actividades sociales complementarias (talleres culturales, grupos de ejercicio)

- Niveles: Sí / No (grupo de control: charla de ecologismo)

• **Variable Dependiente (VD):** Nivel de bienestar subjetivo / satisfacción y reducción de soledad

### Cómo usarlo en evaluación

**Asignación aleatoria** de participantes a los 4 grupos (idealmente).

**Medición antes y después** del programa del bienestar y la soledad.

**Análisis de interacción:** permite ver si combinar termalismo + actividades sociales potencia los efectos sobre la VD más que cada VI por separado.



## 2. Diseño cuasi-experimental

### CONTROL

- Los cuasi experimentos son investigaciones en las que el investigador **no** puede establecer controles en relación con la **asignación aleatoria de participantes** (para diseños con distintos grupos) o con la **inversión del orden de aplicación** (para diseños con el mismo grupo)

### MANIPULACIÓN

- Sí hay **manipulación** de los **niveles de la VI**

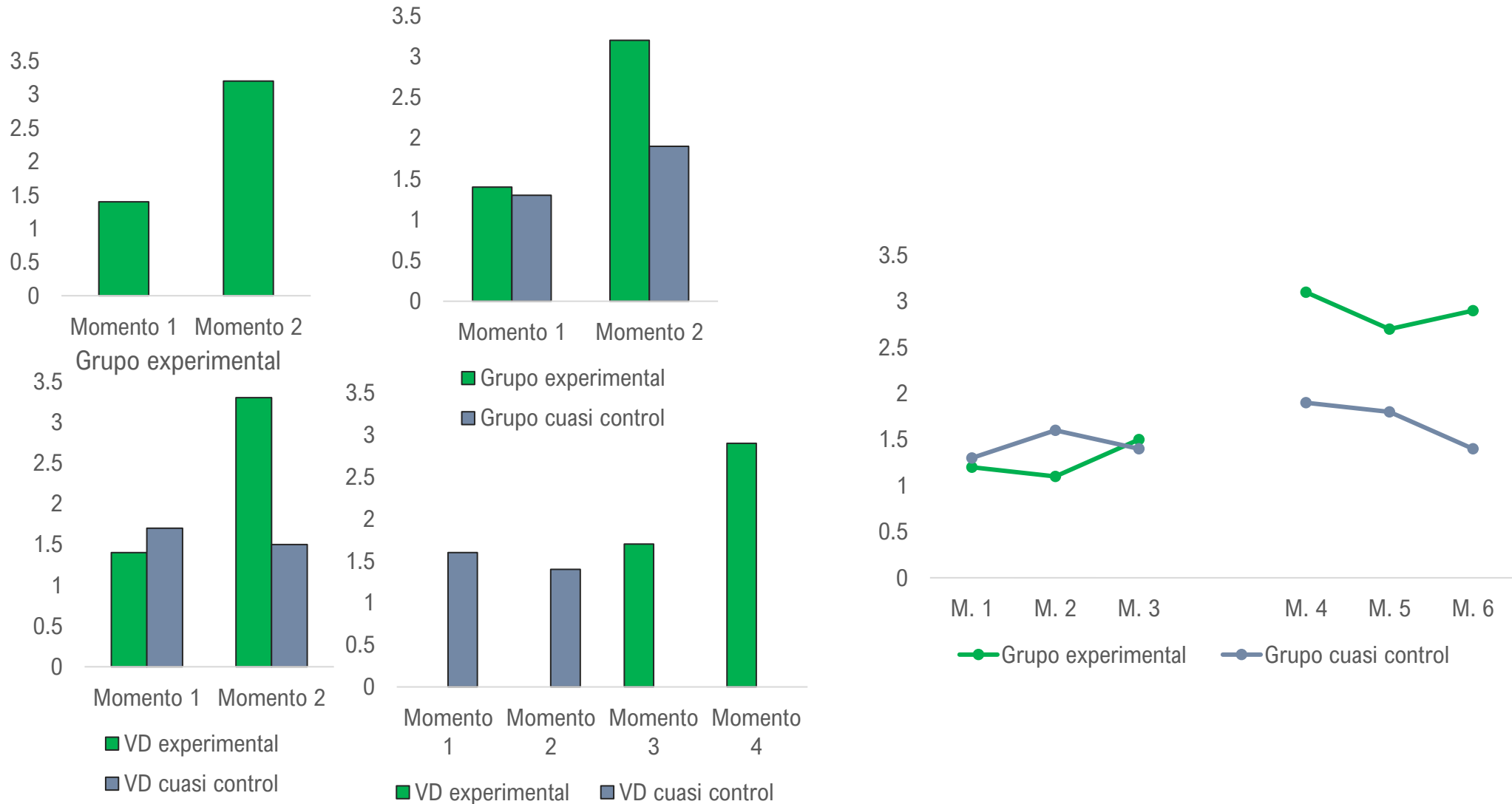
### CAUSALIDAD

- Podríamos llegar a hablar de causalidad, pero con menor seguridad que en un diseño experimental (incluir grupo cuasi-control)

Ventaja: Puede ganarse en validez externa. Son más comunes en la investigación aplicada, están encaminados a probar que algo *funciona*. A veces no se asigna a las personas a los distintos niveles de la VI por motivos prácticos y/o éticos.

...Amenazas a la validez externa: Selección de los participantes (e.g., voluntarios).

## 2. Diseño cuasi-experimental (Cont.)



## 2. Diseño cuasi-experimental (Cont.)

### 📌 Ejemplo práctico: Programa de Termalismo\*

#### Variables

• **Variable Independiente 1 (VI1):** Participación en el programa de termalismo

- Niveles: Sí / No

• **Variable Independiente 2 (VI2):** Participación en actividades sociales complementarias (talleres culturales, grupos de ejercicio)

- Niveles: Sí / No (grupo de control: charla de ecologismo)

Nota: Hay manipulación de las Vis, pero no podemos asignar a los grupos

• **Variable Dependiente (VD):** Nivel de bienestar subjetivo / satisfacción y reducción de soledad

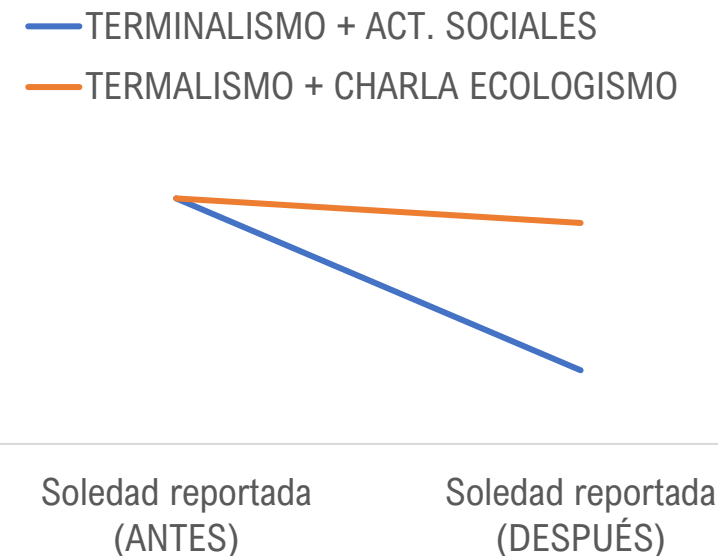
#### Cómo usarlo en evaluación

**Asignación aleatoria** de participantes a los 4 grupos (idealmente).

**Medición antes y después** del programa del bienestar y la soledad.

**Análisis de interacción:** permite ver si combinar termalismo + actividades sociales potencia los efectos sobre la VD más que cada VI por separado.

Nota: Los resultados pueden estar influidos por variables externas. Medición previa recomendada.



## 3. Diseño ex post facto (correlacional)

### CONTROL

- No existe control

### MANIPULACIÓN

- Los niveles de las Vis no son manipulables sino que son características que los sujetos poseen ANTES de comenzar el experimento
- Mayor facilidad para variables extrañas y enmascaradas.

### CAUSALIDAD

- Debido al escaso control que existe sobre las variables extrañas (no hay antecesión ni asignación aleatoria). Sólo se puede hablar de **covariación** (siempre y cuando no sea un diseño retrospectivo simple).

Inconvenientes: Problemas de validez externa debido al sesgo de selección

Ventajas: Son fácilmente aplicables (a veces la única opción).

## 3. Diseño exp post facto (correlacional)

📌 Ejemplo práctico: Programa de Termalismo\*

### Diseño ex post facto retrospectivo

- **VI:** Tipo de programa IMSERSO realizado hace 3 años
  - Niveles: Termalismo + talleres culturales / Termalismo solo / No participó
- **VD:** Estado de salud mental y física actual
  - Medidas: Escalas de depresión, ansiedad, movilidad, presión arterial, satisfacción con la vida
- **Amenazas a la validez interna:** no puedo descargar sesgo retrospectivo (descartar explicaciones alternativas) y enmascaramientos

### Posibles mediadoras (explican el mecanismo por el cual la VI influye en la VD):

**Actividad social:** nivel de interacción social durante el programa, **Nivel de ejercicio físico:** frecuencia y duración de la actividad física realizada, **Adherencia al programa:** asistencia a sesiones programadas, ...

### Posibles moderadoras (alteran la fuerza o dirección del efecto de la VI sobre la VD):

**Edad:** podría modificar cuánto beneficio obtiene cada participante, **Sexo:** hombres y mujeres podrían responder distinto a los programas, **Condición de salud inicial:** personas con enfermedades crónicas podrían mejorar menos que las saludables, ...

## 3. Diseño exp post facto (correlacional)

 Ejemplo práctico: Programa de Termalismo\*

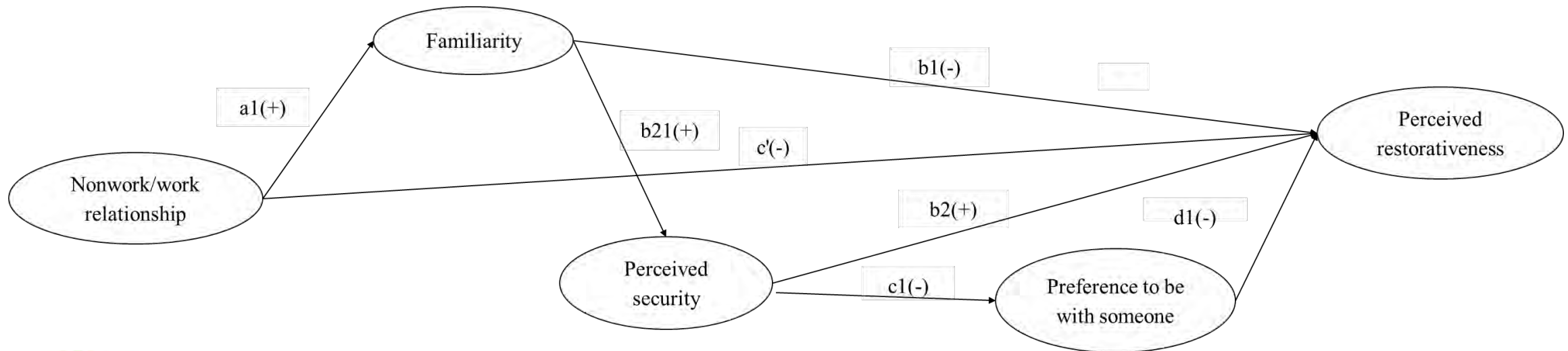
### Prospectivo:

VI participación en programas que se iniciará próximamente y VD estado de salud durante los próximos 3 años.

• **Amenazas a la validez interna:** pérdida de participantes, cambios en condiciones externas, ...

### 3. Diseño exp post facto (correlacional)

Diseños ex post facto: De más de un eslabón causal



Aunque nunca va a haber unas máximas garantías de causalidad en un diseño ex post facto, lo idóneo es que se mida temporalmente de acuerdo a las relaciones establecidas en la cadena



# Propuesta de un estudio para estudiar los beneficios del programa de termalismo

## (1) Propuesta de diseño

- Diseño (experimental, quasi-experimental, correlacional)
- Variable Independiente(s) (VI) y sus niveles
- Variable Dependiente (VD)
- ¿Mediadoras? ¿Moderadoras?
- Hipótesis

## 4. La observación

CONTROL

MANIPULACIÓN

CAUSALIDAD

- **Fiabilidad:** acuerdo interjueces ( $\kappa$  de Cohen)
- Amenazas a la **validez interna:**
  - **Reactividad:** la mera presencia del observador altera el comportamiento usual de la gente. Emplear un periodo de desinsibilización en el que no se registra
  - **Sesgo del observador** a favor de sus expectativas: Usar doble ciego
- A pesar de esto, gracias a su alta **validez externa** puede ayudar a orientar un problema de investigación o establecer qué variables deben tenerse en cuenta

## 4. La observación

Los tipos de diseños en observación son:

- **Observación natural:** El observador es un mero espectador de la situación, sin intervenir en los acontecimientos observados. Se produce en el contexto natural.
- **Observación estructurada:** Los investigadores están en el contexto natural, pero interactúan de alguna manera con la situación. Es un equilibrio entre la observación natural y el experimento de campo, con un grado de control intermedio sobre la situación.
- **Observación de campo:** La observación se realiza en el contexto natural, pero el nivel de estructuración de la situación es mucho mayor. Se sustenta en teorías. Hay manipulación de la VI y control de posibles variables extrañas.
- **Observación participante:** Está más ligado a la metodología cualitativa y a los procesos de investigación-acción.

## 4. La observación

- La elección de la medida de observación depende de la naturaleza del fenómeno observado y de los intereses de la persona que observa:
  - **Ocurrencia:** aparición o no del fenómeno
  - **Frecuencia:** número de veces que aparece el fenómeno
  - **Latencia:** tiempo que transcurre entre la aparición del estímulo y la aparición de la reacción
  - **Duración:** tiempo que transcurre entre la aparición de la reacción y su extinción
  - **Intensidad:** fuerza con la que aparece el fenómeno
- El **muestreo de tiempo** hace referencia a la sistematización de los periodos de observación a lo largo de una jornada: Registro continuo o por intervalos.

## 4. La observación

Verdugo, M. A., Monjas, M. I., & Arias, B. (1992). Intervención sobre la competencia social de los alumnos con necesidades educativas especiales en Educación Infantil y Primaria. *Madrid: MEC*.

- A) Categorías no interactivas:
  - (1) Permanece solo en inactividad
  - (2) Permanece solo en actividad
  - (3) Permanece solo como espectador
- B) Categorías interactivas:
  - (a) Tipo de interacción
    - (4) Interacción cooperativa
    - (5) Interacción social positiva
    - (6) Interacción agresiva
    - (7) Interacción social negativa
  - (b) Modalidad de interacción
    - (8) Iniciación
    - (9) Respuesta
  - (c) Con quién se produce la interacción
    - (10) Niño individual
    - (11) Niño en grupo
    - (12) Adulto
  - (d) Modo de interacción
    - (13) Verbal
    - (14) Física
    - (15) Mixta

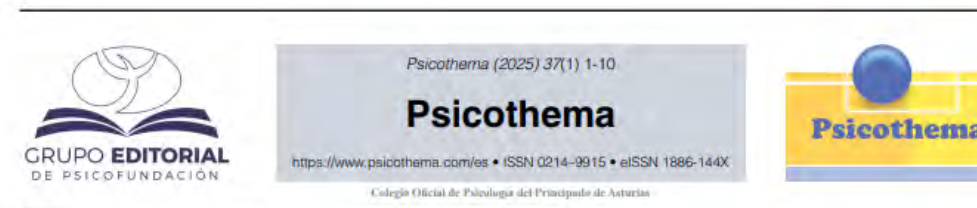
## 4. La observación



### Methodological quality checklist for studies based on observational methodology (MQCOM)

Salvador Chacón-Moscoso<sup>1,2</sup>, M. Teresa Anguera<sup>3</sup>, Susana Sanduvete-Chaves<sup>1</sup>, José L. Losada<sup>3</sup>,  
José A. Lozano-Lozano<sup>2</sup>, and Mariona Portell<sup>4</sup>

<sup>1</sup> Universidad de Sevilla, <sup>2</sup> Universidad Autónoma de Chile (Chile), <sup>3</sup> Universidad de Barcelona, and <sup>4</sup> Universitat Autònoma de Barcelona



Article

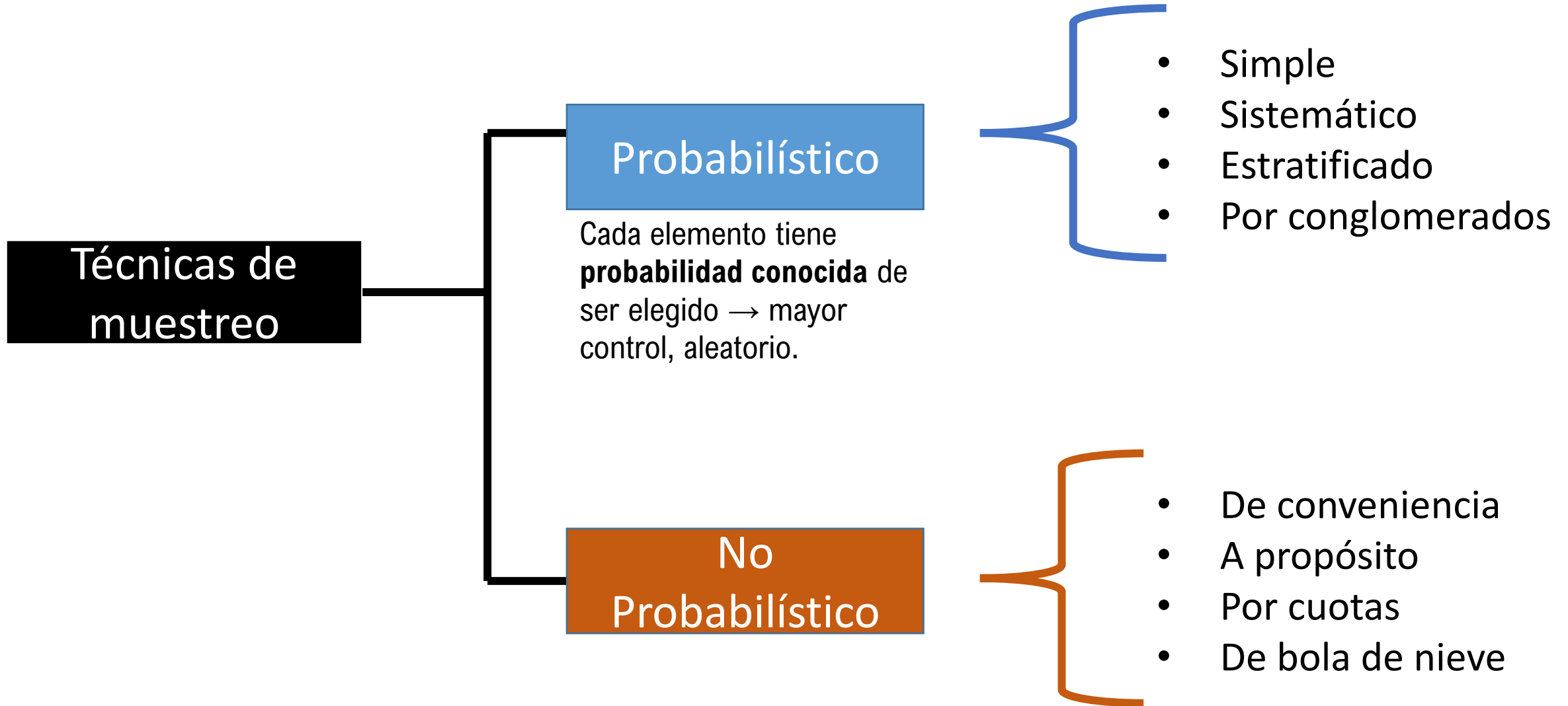
### A Scale for Evaluating the Methodological Quality of Studies Based on Observational Methodology

Susana Sanduvete-Chaves<sup>1</sup>, Daniel López-Arenas<sup>1</sup>, M. Teresa Anguera<sup>2</sup> and Salvador Chacón-Moscoso<sup>1,3</sup>

<sup>1</sup> Universidad de Sevilla (Spain)  
<sup>2</sup> Universidad de Barcelona (Spain)  
<sup>3</sup> Universidad Autónoma de Chile (Chile)

Aspecto	MQCOM (2019)	MQSOM (2025)
Objetivo	Checklist para calidad metodológica	Escala validada para evaluar calidad
Ítems / Dimensiones	16 ítems / 11 criterios	11 ítems / 2 dimensiones
Validación	Expertos + fiabilidad intercodificador	Análisis factorial + fiabilidad inter/intra
Uso	Guía práctica para investigadores	Evaluación cuantitativa y detallada

# Muestreo y Potencia estadística





Hipótesis nula: La persona está embarazada

## Error Tipo I (Falso Positivo)



## Error Tipo II (Falso Negativo)



## Ejemplo de contraste unilateral en programa de termalismo

### Contexto:

- Objetivo: Evaluar si el programa de termalismo **reduce el dolor muscular**.

### Hipótesis:

- H0 (nula)**: La media de dolor muscular  $\geq 5$  (sin mejora).
- H1 (alternativa, unilateral)**: La media de dolor muscular  $< 5$  (el programa reduce el dolor).

### Resultado del contraste:

- Rechazo H0  $\rightarrow$  concluyo que el programa **reduce el dolor**.

### Realidad:

- La media de dolor muscular **es 5**, es decir, **no hubo reducción real**.

### Interpretación:

- Se comete un error tipo I unilateral: se afirma una reducción que en realidad no ocurrió.**

### Error Tipo I (Alfa)

Ocurre cuando un experimentador cree que tiene un resultado significativo, pero en realidad se debe al azar. El riesgo de cometer un error tipo I es igual al nivel de significación (i.e.,  $p < 0.05$ )

### Soluciones:

- ✓ Evitar problemas de **validez interna** (e.g., variables extrañas)
- ✓ Usar un **nivel más exigente de significación** (e.g., corrección por comparaciones múltiples [Bonferroni])
- ✓ **Replicar** el resultado en otra muestra

## Ejemplo de contraste unilateral en programa de termalismo

### Contexto:

- Objetivo: Evaluar si el programa de termalismo **reduce el dolor muscular**.

### Hipótesis:

- H0 (nula)**: La media de dolor muscular  $\geq 5$  (sin mejora).
- H1 (alternativa, unilateral)**: La media de dolor muscular  $< 5$  (el programa reduce el dolor).

### Resultado del contraste:

- No rechazo H0  $\rightarrow$  concluyo que el programa **no reduce el dolor**.

### Realidad:

- La media de dolor muscular **es 4**, es decir, **sí hubo reducción real**.

### Interpretación:

- Se comete un **error tipo II**: no detectamos el efecto real del programa

### Error Tipo II (Beta)

Ocurre cuando un experimentador no encuentra un resultado significativo cuando, de hecho, el efecto es significativo

Debe calcularse con un test de potencia (esto es, dado el tamaño muestral, ¿cómo de grande debería ser un efecto para que fuese capaz de detectarlo?)

Soluciones:

- ✓ **Aumentar el tamaño muestral**
- ✓ Usar medidas **más precisas**
- ✓ **Replicar** el resultado en otra muestra

# Muestreo y Potencia estadística

- Power analysis permite determinar el tamaño muestra requerido para **detectar un efecto dado un cierto grado de confianza** (i.e.,  $1 - \beta$  o potencia; la probabilidad de rechazar una hipótesis nula falsa)
- Se hace normalmente **ANTES de recoger los datos** y se recomienda que la potencia se sitúe entre **0.80 y 0.90**
- Es necesario hipotetizar el **tamaño del efecto** (effect size) definido como la desviación mínima frente a la hipótesis nula que uno espera detectar
  - Ejemplos de medidas de tamaño del efecto son la correlación, el coeficiente de regresión, la diferencia media estandarizada
- Tiene ventajas frente al uso de test estadísticos únicamente (medida de **relevancia**)

# Effect size converter (<https://www.escal.site/>)

## Convert between different effect sizes

By convention, Cohen's  $d$  of 0.2, 0.5, 0.8 are considered small, medium and large effect sizes respectively.

Cohen's $d$	0.300
Pearson's correlation $r$	0.148
R-squared	0.022
Cohen's $f$	0.150
Odds ratio (OR)	1.723
Log odds ratio	0.544
Area-under-curve (AUC)*	0.584
Fisher's $z$ ( $z'$ )	0.149

# G\*Power ([www.gpower.hhu.de](http://www.gpower.hhu.de))

Programa para Windows y Mac que permite hacer análisis de potencia para distintos análisis estadísticos (pruebas  $T$ , ANOVAs, ...)

Tutorials in Quantitative Methods for Psychology  
2007, Vol. 3(2), p. 51-59.



## A short tutorial of *GPower*

**Susanne Mayr**

*Heinrich-Heine-Universität, Düsseldorf, Germany*

**Edgar Erdfelder**

*Universität Mannheim, Mannheim, Germany*

**Axel Buchner**

*Heinrich-Heine-Universität, Düsseldorf, Germany*

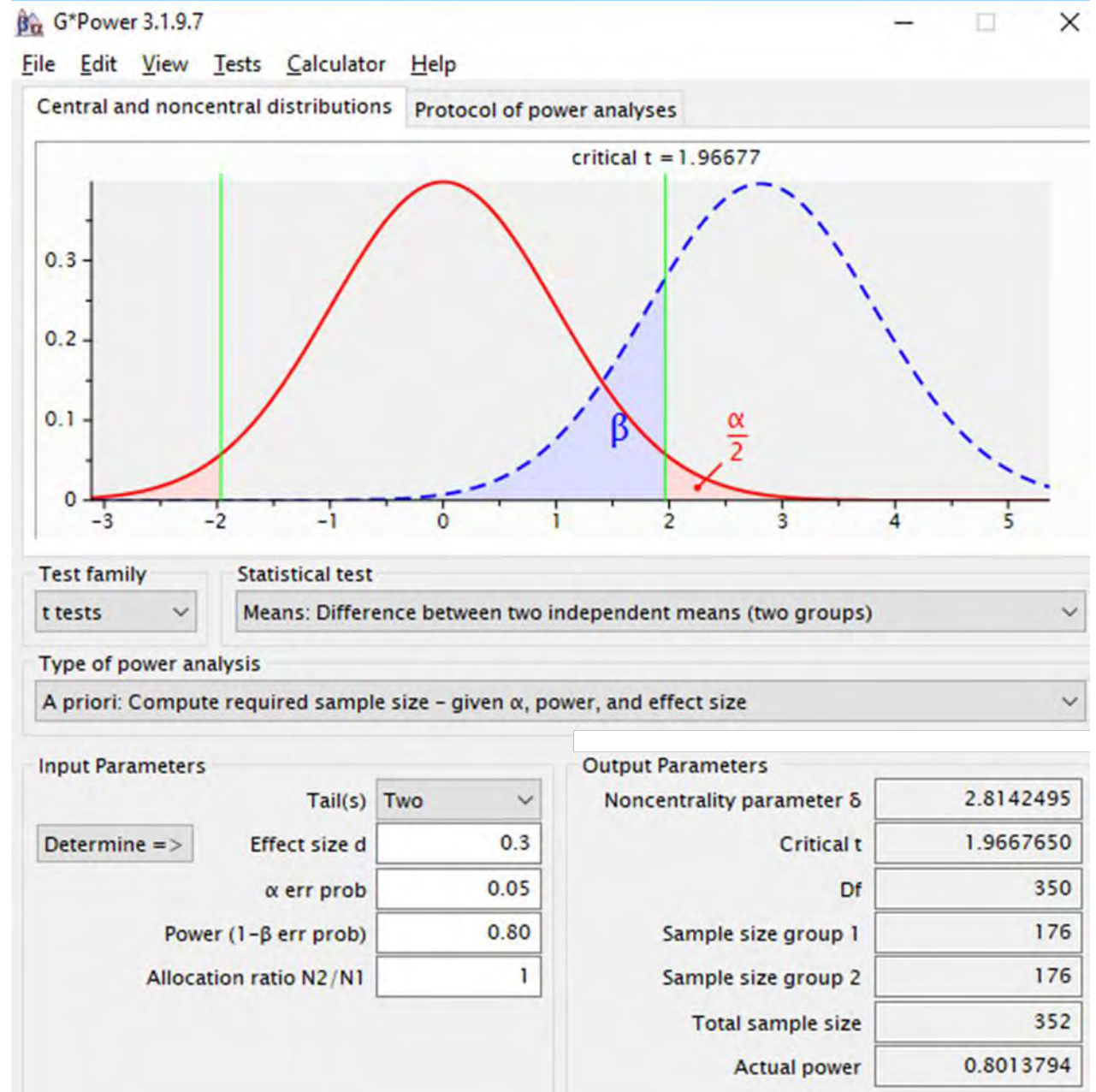
**Franz Faul**

*Christian-Albrechts-Universität, Kiel, Germany*

# G\*Power

## Comparación de medias para grupos independientes (bilateral)

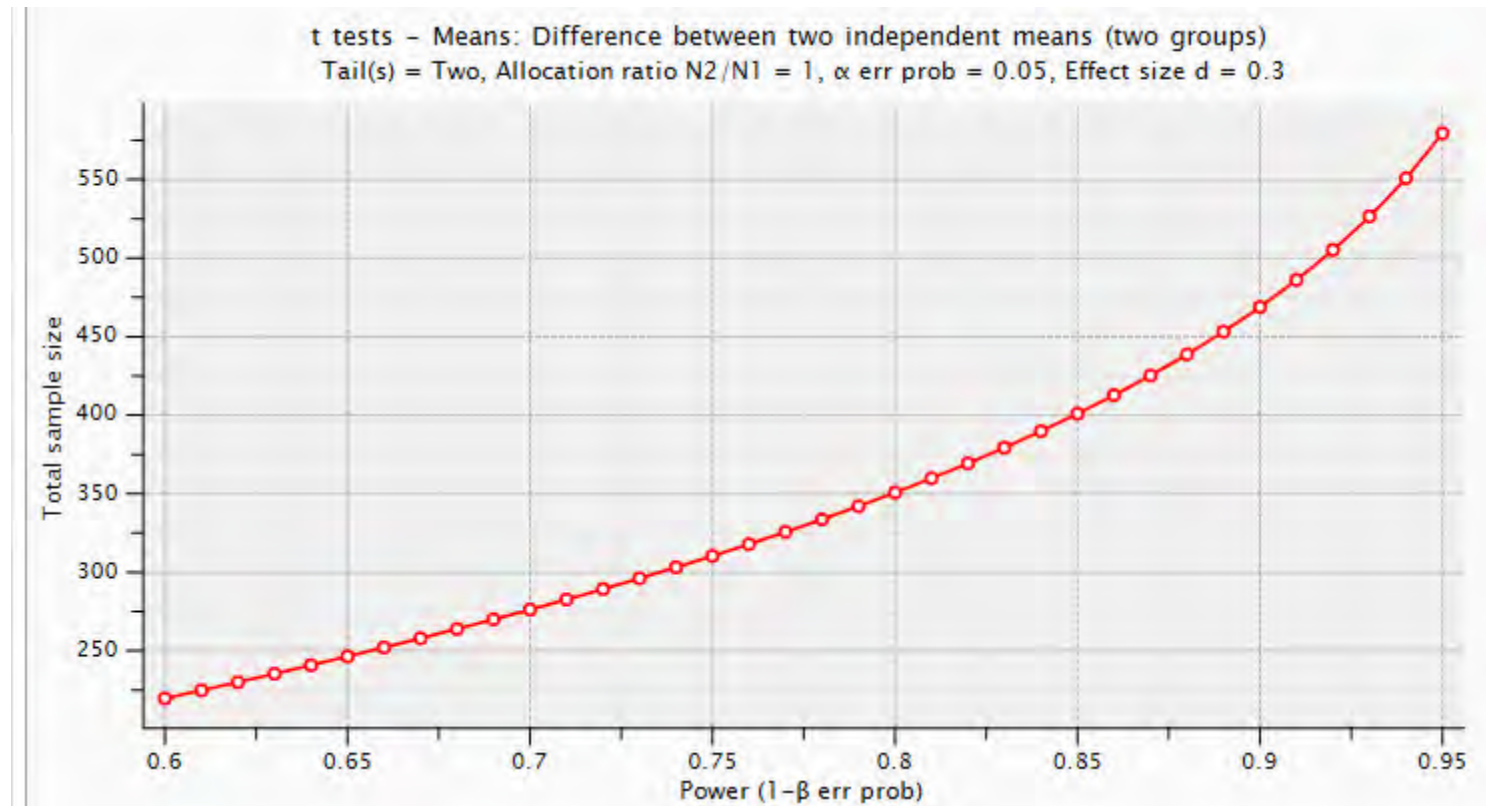
- La bibliografía previa me indica que el tamaño del efecto es pequeño ( $d = 0.30$ )
- Nivel de confianza 95% ( $\alpha = 0.05$ )
- Potencia deseada 0.80 ( $\beta = 0.20$ )



# G\*Power

- Click: [X-Y plot for a range of values](#)

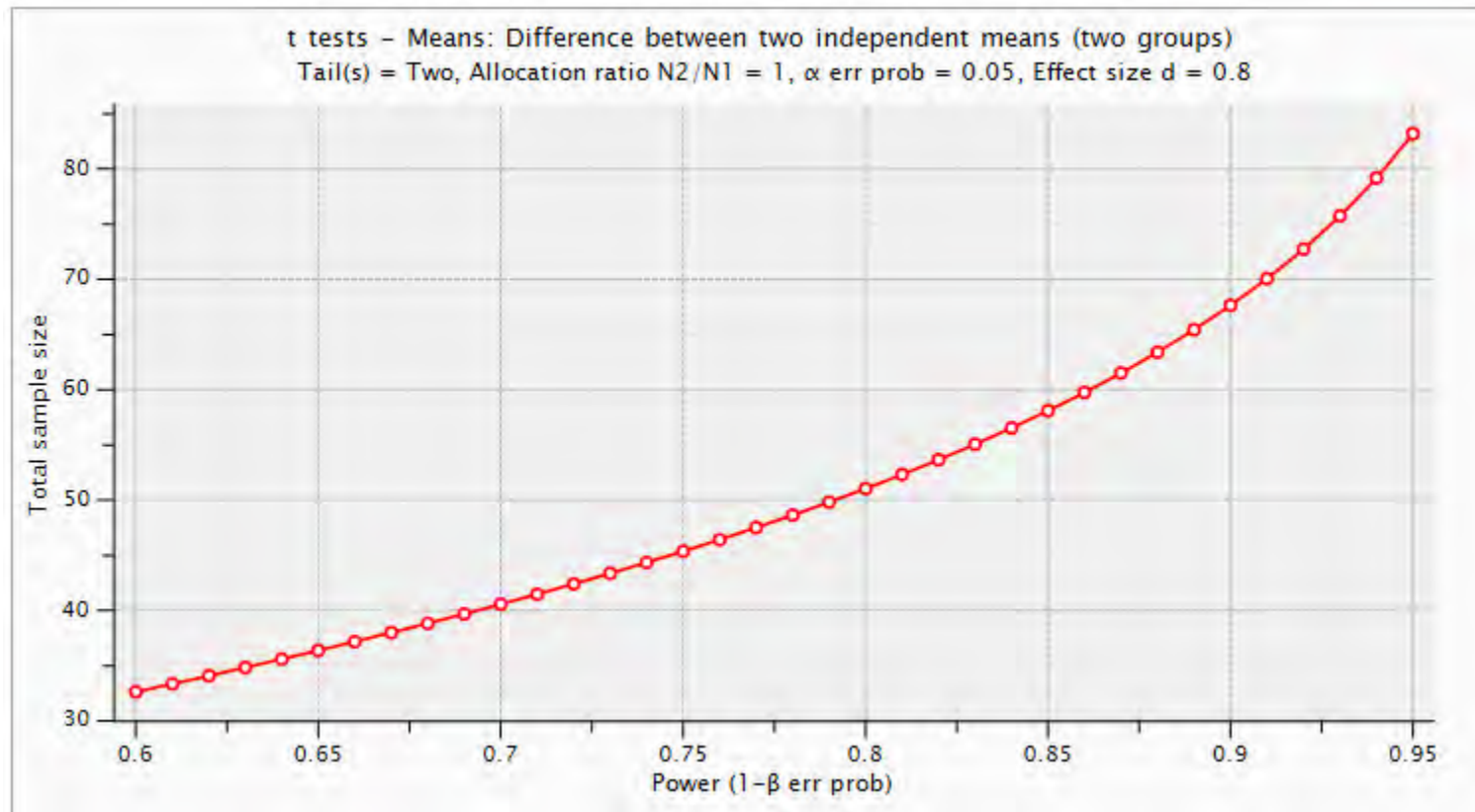
$d = 0.30$   
(small)



# G\*Power

- Click: [X-Y plot for a range of values](#)

$d = 0.80$   
(large)



# Conclusiones finales sobre el tamaño muestral

- **Depende** del análisis que se desee llevar a cabo, el nivel de confianza, el nivel de potencia deseado y el tamaño del efecto que se desea detectar
- Revisa la **literatura previa** para ver qué tamaños del efecto cabe esperar. Metanálisis y revisiones son especialmente útiles:
  - Richard, F. D., Bond Jr, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7(4), 331-363.
- Lleva a cabo estudios de potencia a priori usando software como **G\*Power**

¿Qué tamaño muestral necesito si deseo llevar a cabo un  $t$  de student 95% de confianza y deseo una potencia de al menos 0.80, si preveo que el tamaño del efecto es pequeño?

Al menos 352 sujetos





# Propuesta de un estudio para estudiar los beneficios del programa de termalismo

## (2) Muestreo

Para una hipótesis que implique comparación de medias entre dos grupos,

- Especificad el tipo de muestreo
- Define el número de participantes por grupo atendiendo a la potencia estadística (potencia = 0.80, alfa = 0.05)

Tabla sugerida: | Análisis | Tipo de muestreo | Número de participantes por grupo | Potencia resultante

## Instrumentos de medición

- La medición permite cuantificar fenómenos complejos.
- Medición: *“Asignar números o categorías a características de personas u objetos siguiendo reglas sistemáticas”*.
- Permite:
  - Comparar individuos o grupos
  - Evaluar cambios a lo largo del tiempo
  - Medir efectos de intervenciones
- Fiabilidad, validez y eficiencia son propiedades fundamentales en contextos aplicados.

**Ejemplo:** La **Satisfaction With Life Scale (SWLS, Diener et al., 1985)** mide satisfacción con la vida.

### *Escala de Satisfacción con la Vida (SWLS; Diener et al., 1985)*

*Instrucciones.* A continuación hay cinco afirmaciones con las cuales usted puede estar de acuerdo o en desacuerdo. Lea cada una de ellas y después seleccione la respuesta que mejor describa en qué grado está de acuerdo o en desacuerdo (1 = Fuertemente en desacuerdo; 2 = En desacuerdo; 3 = Ligeramente en desacuerdo; 4 = Ni de acuerdo ni en desacuerdo; 5 = Ligeramente de acuerdo; 6 = De acuerdo; 7 = Fuertemente de acuerdo).

1.	En la mayoría de los aspectos, mi vida se acerca a mi ideal	1	2	3	4	5	6	7
2.	Las condiciones de mi vida son excelentes	1	2	3	4	5	6	7
3.	Estoy completamente satisfecho/a con mi vida	1	2	3	4	5	6	7
4.	Hasta ahora, he conseguido las cosas más importantes que quiero en la vida.	1	2	3	4	5	6	7
5.	Si pudiera vivir mi vida de nuevo, no cambiaría nada	1	2	3	4	5	6	7

---

## Instrumentos de medición: Fiabilidad

- La fiabilidad indica si una medida es estable y consistente.
- Fuentes de evidencia principales:
  - Consistencia interna: todos los ítems miden lo mismo.
  - Test-retest: estabilidad de los puntajes en el tiempo.
  - Equivalencia entre formas: versiones diferentes de la misma prueba.

**Ejemplo:** Se ha encontrado una alta consistencia interna en el caso de la **SWLS** [ $\alpha = 0.88$  en Vázquez et al. (2013)]

## Instrumentos de medición: Validez

- La validez se refiere a la evidencia y los argumentos que respaldan la interpretación de los resultados de un test o medida para un propósito concreto.
- Fuentes de evidencia principales:
  - Validez de contenido: ítems representativos del constructo
  - Validez convergente/discriminante: correlación (o ausencia) con medidas relacionadas teóricamente
  - Validez de criterio: correlación con medidas externas relevantes
  - Validez de estructura interna: estructura factorial
  - Validez referida a las consecuencias de aplicación: ausencia de sesgo

## Instrumentos de medición: Validez

- **Validez de contenido:** ítems representativos del constructo

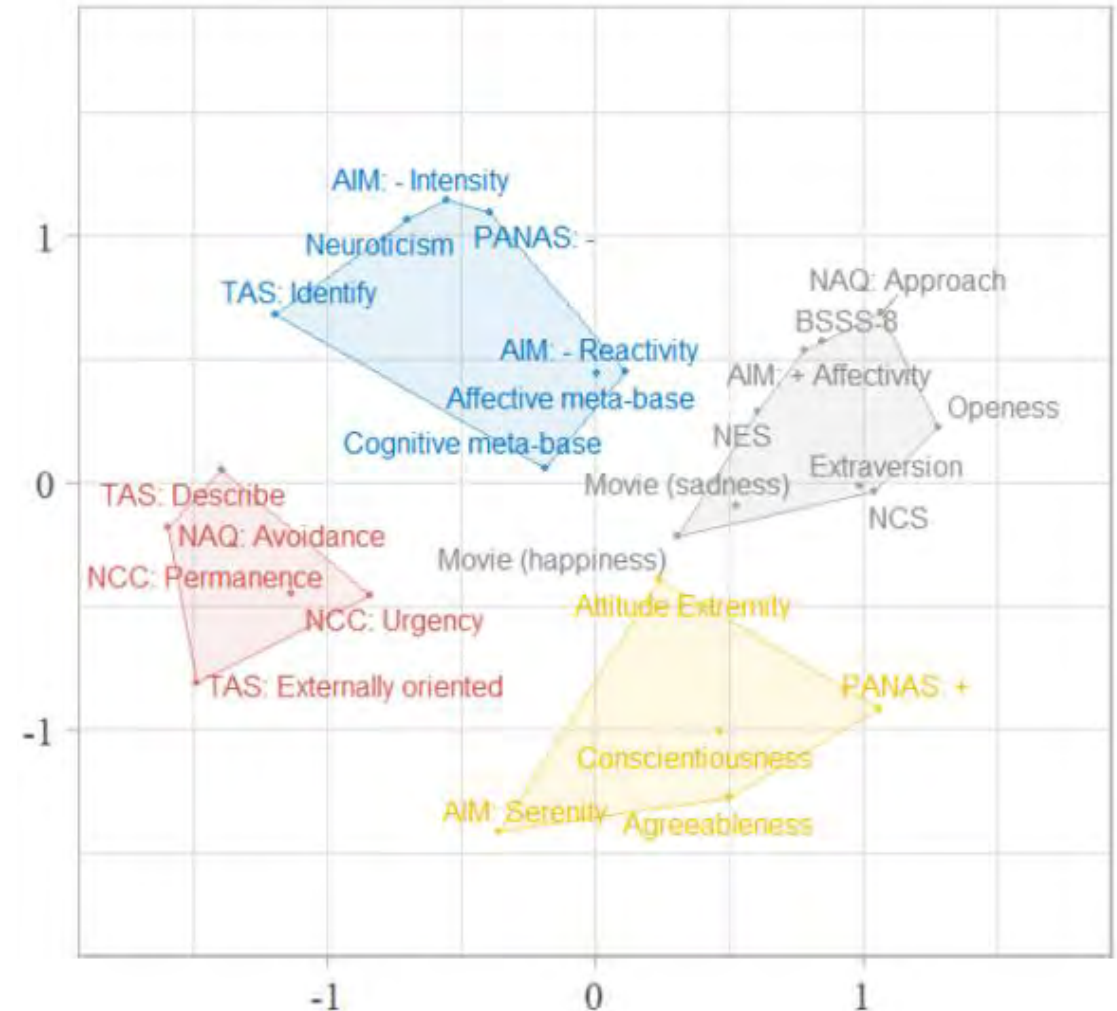
<https://doi.org/10.7334/psicothema2023.208> / <https://journals.copmadrid.org/psed/art/psed2022a15>

their thesis. The items were developed based on Bandura's theory of learning and on these competencies. To test the content validity of the scale, these items were independently evaluated by a group of four educational psychology experts. They were asked to report using a five-point scale (from 1 = none to 5 = very high) whether they thought the items measured self-efficacy, as described by Bandura (1997), as well as whether they thought the items were comprehensive. The items were also assessed by 10 Master students who had recently submitted their thesis and gone through the oral examination. Participants were asked whether they thought the items reflected the competencies they needed to acquire to complete their Master's thesis and whether the items were clear. There was an agreement that the items fully covered the basic competencies needed to succeed in the completion of a Master's thesis. Specifically, the average of the evaluations made for the 8 items for relevance and clarity was 4.97/4.80 and 4.81/4.79 for experts/students, respectively. As can be seen, the scores were always very close to the

# Instrumentos de medición: Validez

- **Validez convergente/discriminante:** correlación (o ausencia) con medidas relacionadas teóricamente

<https://doi.org/10.7334/psicothema2022.328>



## Instrumentos de medición: Validez

- **Validez de criterio:** correlación con medidas externas relevantes

### **Ejemplo: La SWLS predice:**

Bienestar emocional: afecto positivo / negativo

Salud mental: depresión y ansiedad

Salud física percibida

Éxito académico o laboral

Calidad de relaciones y apoyo social

# Instrumentos de medición: Validez

- **Validez de estructura interna:** estructura factorial (Vázquez et al., 2013)

<https://doi.org/10.3389/fpsyg.2021.614470>

**Table 7.** *Confirmatory Factor Analysis. Model Fit*

Standardized coefficients  
(Confidence intervals)

Item 1	.82 (.80–.84)					
Item 2	.82 (.80–.84)					
Item 3	.88 (.86–.90)					
Item 4	.74 (.71–.76)					
Item 5	.64 (.61–.67)					
Goodness of Fit indexes	$\chi^2$	<i>df</i>	<i>p</i>	CFI	TLI	RMSEA
Model	38.46	5	.001	.994	.988	.047

*Note:* *df* = degrees of freedom; *p* = significance level; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation.

# Instrumentos de medición: Validez

- **Validez referida a las consecuencias de aplicación:** ausencia de sesgo

<https://doi.org/10.7334/psicothema2016.394>



## Measurement invariance of the Satisfaction With Life Scale (SWLS) by country, gender and age

Igor Esnaola<sup>1</sup>, Manuel Benito<sup>1</sup>, Iratxe Antonio-Agirre<sup>1</sup>, John Freeman<sup>2</sup> and Marta Sarasa<sup>1</sup>

<sup>1</sup> Universidad del País Vasco / Euskal Herriko Unibertsitatea and <sup>2</sup> Queen's University (Canada)

## Instrumentos de medición: Eficiencia

- Philadelphia Geriatric Center Morale Scale (PGCMS) – 17 ítems
- Mental Health Continuum – Short Form (MHC-SF) – 14 ítems
- Oxford Happiness Questionnaire (OHQ) – 29 ítems
- PROMIS (test adaptativo informatizado):



### LIFE SATISFACTION

A brief guide to the PROMIS® Life Satisfaction instruments:

ADULT	PEDIATRIC	PARENT PROXY
PROMIS Item Bank v1.0 General Life Satisfaction	PROMIS Pediatric Item Bank v1.0 Life Satisfaction	PROMIS Proxy Item Bank v1.0 Life Satisfaction
PROMIS Short Form v1.0 General Life Satisfaction 5a	PROMIS Pediatric Short Form v1.0 Life Satisfaction 4a	PROMIS Proxy Short Form v1.0 Life Satisfaction 4a
	PROMIS Pediatric Short Form v1.0 Life Satisfaction 8a	PROMIS Proxy Short Form v1.0 Life Satisfaction 8a
	PROMIS Pediatric Short Form v1.0 Life Satisfaction 8b	PROMIS Proxy Short Form v1.0 Life Satisfaction 8b



# Propuesta de un estudio para estudiar los beneficios del programa de termalismo

## (3) Medición

- Para cada VI y VD, especificad el instrumento de medida
- Justifiquen la elección del instrumento (evidencia de fiabilidad, validez, ...)

Tabla sugerida:

| Variable | Instrumento | Evidencia de validez/fiabilidad (referencias) |

1

Introducción

2

Métodos

3

Análisis

4

Presentación



## Preguntas de interés en el análisis de programas

✨ *“No solo importa saber si el programa funcionó, sino también entender para quién y por qué. Eso convierte los resultados en decisiones útiles para mejorar las intervenciones.”*



## Preguntas de interés en el análisis de programas

### 1 ¿Funcionó el programa?

**Propósito:** evaluar la eficacia global.

**Indicadores típicos:**

Cambios pre–post en salud, bienestar o calidad de vida.

Comparaciones entre grupo intervención y grupo control.

**Ejemplos de análisis:**

T-test pareado (pre vs post).

ANOVA/ANCOVA.

Regresión con variable dependiente de resultado.

### 2 ¿Para quién funcionó mejor?

**Propósito:** identificar subgrupos beneficiados diferencialmente.

**Variables moderadoras:**

Sexo, edad, nivel socioeconómico, diagnóstico de base.

**Ejemplos de análisis:**

Comparaciones de medias por subgrupo.

Interacciones en modelos de regresión (*grupo × sexo*, *grupo × edad*).

Análisis estratificados.

### 3 ¿Qué factores influyen en los resultados?

**Propósito:** entender los determinantes del éxito.

**Posibles predictores:**

Características individuales (edad, salud previa).

Variables contextuales (apoyo familiar, recursos sanitarios).

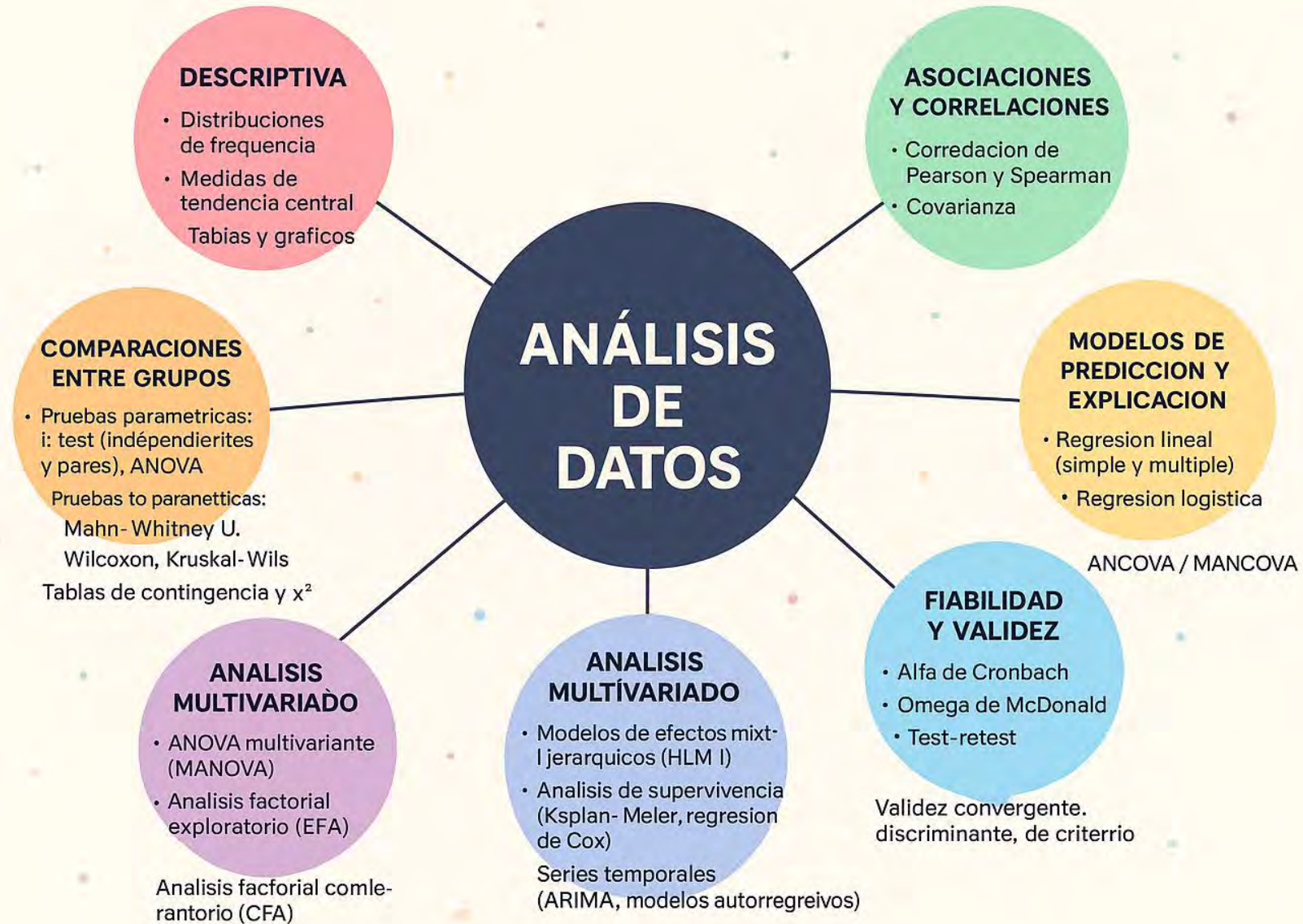
Intensidad/adhesión al programa.

**Ejemplos de análisis:**

Regresión múltiple / logística.

Modelos jerárquicos (si hay datos de pacientes dentro de centros).

Análisis multivariado (ej. modelos de ecuaciones estructurales para relaciones complejas).



+ ANCOVA, clúster, ROC, índice de cambio fiable, machine learning, bayesiana...

## Software disponible para análisis de datos

### ◆ Lenguajes y entornos estadísticos

**R** – Lenguaje de programación estadística, gran capacidad de análisis y visualización.

**Python** – Flexible, ideal para análisis de datos, machine learning y visualizaciones

### ◆ Software de estadística tradicional

**SPSS** – Popular en ciencias sociales y psicología; interfaz gráfica fácil de usar.

**Stata** – Muy usado en economía, sociología y epidemiología; potente en análisis de panel y regresión.

### ◆ Software libre y moderno

**Jamovi** – Alternativa gratuita y amigable a SPSS; análisis estadístico y gráficos.

**JASP** – Estadística bayesiana y clásica; muy intuitivo para investigación académica.

### ◆ Plataformas de análisis avanzado y Big Data

**Tableau / Power BI** – Visualización y BI interactivo; integración con múltiples fuentes de datos.

```

> str(datos)
'data.frame': 300 obs. of 30 variables:
 $ id          : int  1 2 3 4 5 6 7 8 9 10 ...
 $ center_id   : int  4 10 1 7 10 1 4 7 5 7 ...
 $ grupo       : Factor w/ 2 levels "Control", "Intervencion": 1 2 2 1 2 1 1 1 2
 1 ...
 $ sexo        : Factor w/ 2 levels "Hombre", "Mujer": 2 1 2 1 1 2 1 2 2 2 ...
 $ edad        : num  69 69 67 68 79 92 75 77 93 69 ...
 $ socioeco    : chr  "Medio" "Bajo" "Medio" "Medio" ...
 $ diagnostico : chr  "Neurodeg" "Cancer" "Neurodeg" "Otra" ...
 $ adherencia  : num  0.734 0.604 0.648 0.768 0.689 ...
 $ apoyo_familiar : num  5.9 4.4 7.1 1.2 8.8 4.6 6.6 4.1 5.8 5.5 ...
 $ salud1_pre  : num  4 3 4 3 3 3 3 3 4 4 ...
 ...
 $ salud5_pre  : num  3 2 4 3 2 3 4 3 3 4 ...
 $ salud1_post : num  4 3 4 3 4 4 3 4 3 5 ...
 ...
 $ salud5_post : num  4 1 4 3 2 3 5 3 4 5 ...
 $ salud_total_pre : num  19 13 18 14 11 16 17 16 19 19 ...
 $ salud_total_post: num  19 14 19 14 13 17 19 18 18 24 ...
 $ bienestar_pre  : num  7.4 8.3 7.5 7.3 3.6 6.8 6.4 7.8 9 7.4 ...
 $ bienestar_post : num  9.9 9.6 9 9.6 4.8 8.2 9.6 8.5 10 8.7 ...
 $ dolor_pre      : num  3.2 10 6.2 0.7 5 7.3 2.5 2.3 3.4 1.9 ...
 $ dolor_post     : num  1.2 10 5 1.1 2.2 7 4.3 0 1 2.2 ...
 $ mejora_bienestar: int  1 1 1 1 1 1 1 0 1 1 ...
 $ time           : num  7.1 11.5 11.1 8.6 3 2.7 0.4 21.4 24.4 27.2 ...
 $ event          : int  1 0 1 1 1 1 1 1 1 1 ...

```

## 1. Descriptiva (la base de todo)

- Distribuciones de frecuencia
- Medidas de tendencia central (media, mediana, moda)
- Medidas de dispersión (rango, varianza, desviación típica)
- Tablas y gráficos (barras, histogramas, boxplots)

👉 Objetivo: *describir la muestra.*

	Hombre	Mujer
Control	231	250
Intervencion	220	299

	Alto	Bajo	Medio
Control	0.067	0.161	0.253
Intervencion	0.070	0.185	0.264

grupo	n	edad_mean	edad_sd	bienestar_pre_mean	bienestar_post_mean
1 Control	481	71.7	9.90	6.42	6.82
2 Intervencion	519	72.2	9.99	6.38	7.73

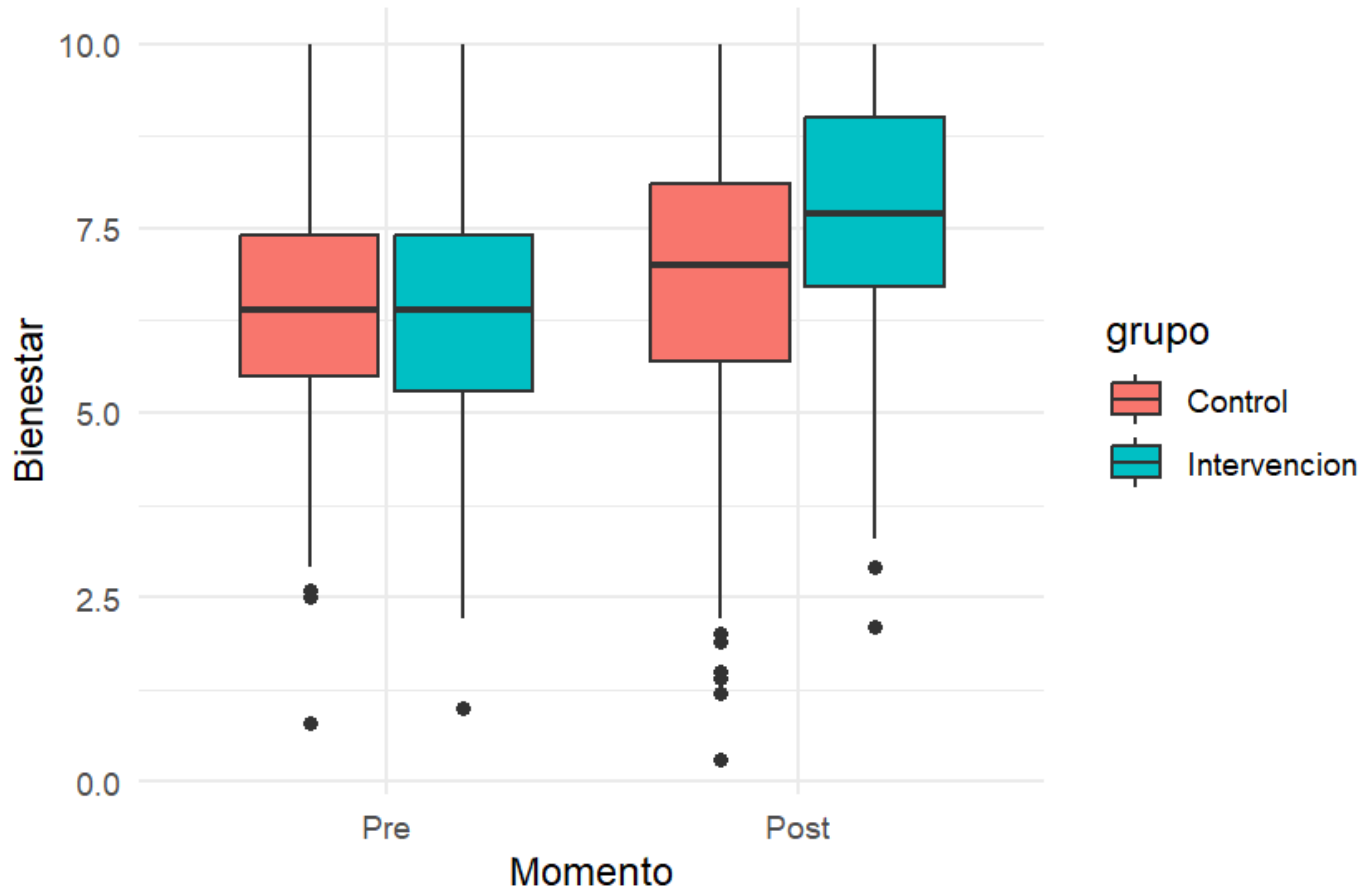
```
table(datos$grupo,
datos$sexo)prop.table(table(datos$grupo,
datos$socioeco))
```

```
datos %>%
  group_by(grupo) %>%
  summarise(
n = n(),
edad_mean = mean(edad),
edad_sd = sd(edad), bienestar_pre_mean =
mean(bienestar_pre, na.rm=TRUE),
bienestar_post_mean = mean(bienestar_post,
na.rm=TRUE))
```

## 1. Descriptiva (la base de todo)

- Distribuciones de frecuencia
- Medidas de tendencia central (media, mediana, moda)
- Medidas de dispersión (rango, varianza, desviación típica)
- Tablas y gráficos (barras, histogramas, boxplots)

👉 Objetivo: *describir la muestra.*



```
datos_largos <- datos %>%  
pivot_longer(cols = c(bienestar_pre,  
bienestar_post), names_to = "momento",  
values_to = "bienestar") %>%  
mutate(momento = factor(momento, levels =  
c("bienestar_pre", "bienestar_post"),  
labels = c("Pre", "Post")))
```

```
ggplot(datos_largos, aes(x = momento, y =  
bienestar, fill = grupo)) + geom_boxplot() +  
labs(x = "Momento", y = "Bienestar") +  
theme_minimal()
```

## 2. Comparaciones entre grupos

- Pruebas paramétricas: t-test (independientes y pareados), ANOVA
- Pruebas no paramétricas: Mann–Whitney U, Wilcoxon, Kruskal–Wallis
- Tablas de contingencia y  $\chi^2$

👉 Objetivo: *saber si los grupos difieren.*

```
t.test(bienestar_post ~ grupo, data = datos)
```

### Welch Two Sample t-test

data: bienestar\_post by grupo

t = -7.7052, df = 885.87, p-value = 3.497e-14

alternative hypothesis: true difference in means between group Control and group Intervencion is not equal to 0

95 percent confidence interval:

-1.1409580 -0.6777117

sample estimates:

mean in group Control mean in group Intervencion

6.824612

7.733947

## 2. Comparaciones entre grupos

- Pruebas paramétricas: t-test (independientes y pareados), ANOVA
- Pruebas no paramétricas: Mann–Whitney U, Wilcoxon, Kruskal–Wallis
- Tablas de contingencia y  $\chi^2$

👉 Objetivo: *saber si los grupos difieren.*

```
datos_largos <-  
datos %>% pivot_longer(cols = c(bienestar_pre,  
bienestar_post), names_to = "tiempo", values_to  
= "bienestar") %>%  
mutate(tiempo = factor(tiempo, levels =  
c("bienestar_pre", "bienestar_post"), labels =  
c("Pre", "Post")))
```

```
modelo <- lmer(bienestar ~ grupo * tiempo + (1 |  
id), data = datos_largos)
```

```
anova(modelo, type = 3)
```

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
grupo	11.96	11.96	1	999.03	19.702	1.006e-05 ***
tiempo	372.40	372.40	1	951.26	613.373	< 2.2e-16 ***
grupo:tiempo	105.61	105.61	1	951.26	173.941	< 2.2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## 2. Comparaciones entre grupos

- Pruebas paramétricas: t-test (independientes y pareados), ANOVA
- Pruebas no paramétricas: Mann–Whitney U, Wilcoxon, Kruskal–Wallis
- Tablas de contingencia y  $\chi^2$

👉 Objetivo: *saber si los grupos difieren.*

```
chisq.test(table(datos$grupo,  
datos$mejora_bienestar))
```

Pearson's Chi-squared test with Yates' continuity correction

data: table\_mejora

X-squared = 88.671, df = 1, p-value < 2.2e-16

### 3. Asociaciones y correlaciones

- Correlación de Pearson y Spearman
- Covarianza
- Correlaciones parciales y semiparciales

👉 Objetivo: *examinar relaciones lineales entre variables.*

```
cor.test(
  datos$salud_total_post,
  datos$bienestar_post, use="complete.obs")

partial.r(datos[, c("salud_total_post",
  "bienestar_post", "edad")])
```

Pearson's product-moment correlation

data: datos\$salud\_total\_post and datos\$bienestar\_post

t = 4.3629, df = 938, p-value = 1.426e-05

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.07779063 0.20314100

sample estimates:

cor

0.141031

	salud_total_post	bienestar_post	edad
salud_total_post	1.00000000	0.14086699	0.01364019
bienestar_post	0.14086699	1.00000000	0.01049373
edad	0.01364019	0.01049373	1.00000000

#### 4. Modelos de predicción y explicación

- **Regresión lineal** (simple y múltiple)
  - **Regresión logística** (cuando la DV es dicotómica)
  - ANCOVA / MANCOVA (efectos ajustados por covariables)
  - Modelos lineales generalizados (GLM)
- 👉 Objetivo: *explicar o predecir un resultado.*

```
lm1 <- lm(bienestar_post ~ grupo +  
salud_total_post + edad + sexo + apoyo_familiar,  
data = datos)summary(lm1)
```

Call:

```
lm(formula = bienestar_post ~ grupo + salud_total_post + edad + sexo + apoyo_familiar, data = datos)
```

Residuals:

Min	1Q	Median	3Q	Max
-7.4947	-1.2760	0.0458	1.4032	3.4817

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	6.315195	0.513944	12.288	< 2e-16 ***
grupoIntervencion	1.002992	0.120597	8.317	3.06e-16 ***
salud_total_post	0.073634	0.015843	4.648	3.82e-06 ***
edad	-0.010874	0.005766	-1.886	0.0596 .
sexoMujer	-0.242705	0.117182	-2.071	0.0386 *
apoyo_familiar	0.030170	0.028799	1.048	0.2951

Residual standard error: 1.801 on 964 degrees of freedom  
(30 observations deleted due to missingness)  
Multiple R-squared: 0.1029, Adjusted R-squared: 0.09829  
F-statistic: 22.12 on 5 and 964 DF, p-value: < 2.2e-16

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### 4. Modelos de predicción y explicación

- **Regresión lineal** (simple y múltiple)
  - **Regresión logística** (cuando la DV es dicotómica)
  - ANCOVA / MANCOVA (efectos ajustados por covariables)
  - Modelos lineales generalizados (GLM)
- 👉 Objetivo: *explicar o predecir un resultado.*

```
glm1 <- glm(mejora_bienestar ~ grupo +
salud_total_pre + edad + sexo + adherencia, data
= datos, family = binomial)
summary(glm1)exp(coef(glm1))
```

Call:  
glm(formula = mejora\_bienestar ~ grupo + salud\_total\_pre + edad +  
sexo + adherencia, family = binomial, data = datos)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )			
(Intercept)	0.221381	0.618829	0.358	0.7205	(Intercept)	grupointervencion	salud_total_pre
grupointervencion	1.135480	0.138004	8.228	<2e-16 ***	1.2477984	3.1126666	0.9875823
salud_total_pre	-0.012495	0.018642	-0.670	0.5027	edad	sexoMujer	adherencia
edad	-0.012044	0.006685	-1.802	0.0716 .	0.9880283	0.9594900	0.9986214
sexoMujer	-0.041353	0.134323	-0.308	0.7582			
adherencia	-0.001380	0.361471	-0.004	0.9970			

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## 5. Análisis multivariado

- Análisis factorial exploratorio (EFA)
- Análisis factorial confirmatorio (CFA)
- Análisis de componentes principales (PCA)

👉 Objetivo: *reducir dimensiones y entender estructuras latentes.*

```
salud_post <- paste0("salud", 1:5, "_post")  
fa.parallel(datos[, salud_post], fa="pc")  
efa <- fa(datos[, salud_post], nfactors=1,  
rotate="oblimin", fm="uls")  
efa$loadings  
EGA(datos[, salud_post])
```

Parallel analysis suggests that the number of factors = NA and the number of components = 1

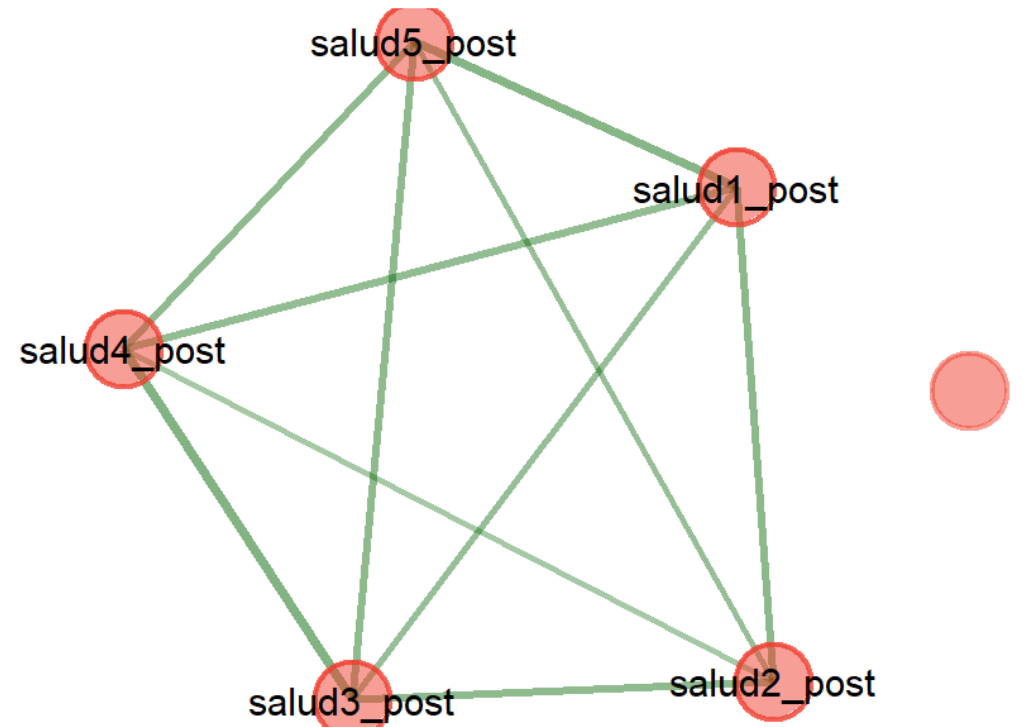
Loadings:

ULS1

salud1\_post 0.619  
salud2\_post 0.566  
salud3\_post 0.630  
salud4\_post 0.604  
salud5\_post 0.626

ULS1

SS loadings 1.859  
Proportion Var 0.372



## 6. Fiabilidad y validez

- Alfa de Cronbach
  - Omega de McDonald
  - Test-retest
  - Validez convergente, discriminante, de criterio
- 👉 Objetivo: *asegurar calidad de los instrumentos.*

```
> psych::alpha(datos[, salud_pre])$total$raw_alpha
```

```
[1] 0.8702241
```

```
> psych::alpha(datos[, salud_post])$total$raw_alpha
```

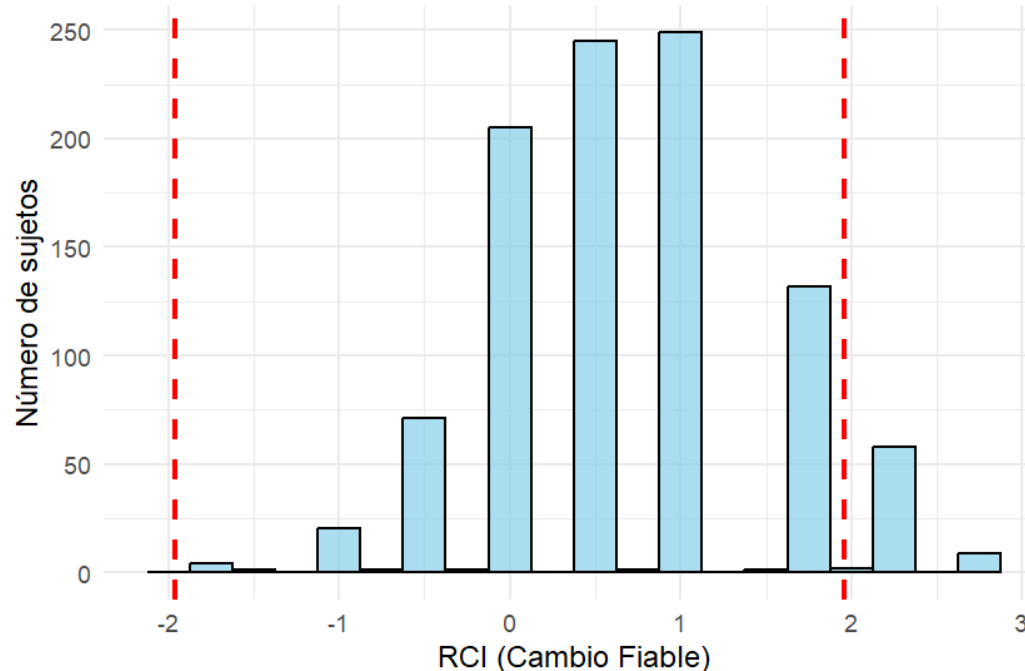
```
[1] 0.7464545
```

```
> r_observada
```

```
[1] 0.1440987
```

```
> r_corregida
```

```
[1] 0.1667855
```



```
salud_pre <- paste0("salud", 1:5, "_pre")  
psych::alpha(datos[, salud_pre])$total$raw_alpha  
psych::alpha(datos[, salud_post])$total$raw_alpha
```

```
salud_post_mean <- rowMeans(datos[, salud_post], na.rm = TRUE)  
r_observada <- cor(salud_post_mean, datos$bienestar_post, use =  
"pairwise.complete.obs")  
r_corregida <- r_observada /sqrt(0.7464545)
```

```
S_diff <- sd_pre * sqrt(2 * (1 -  
alpha_salud_pre))  
RCI <- (salud_post_mean - salud_pre_mean) /  
S_diff
```

## 7. Análisis avanzados y longitudinales

- Modelos de efectos mixtos / jerárquicos (HLM, LMM)
- Análisis de supervivencia (Kaplan–Meier, regresión de Cox)
- Series temporales (ARIMA, modelos autorregresivos)
- Redes sociales y análisis de grafos
- Machine learning aplicado a ciencias sociales y salud (árboles de decisión, random forest, clustering, SVM).
- Modelos de ecuaciones estructurales (SEM)

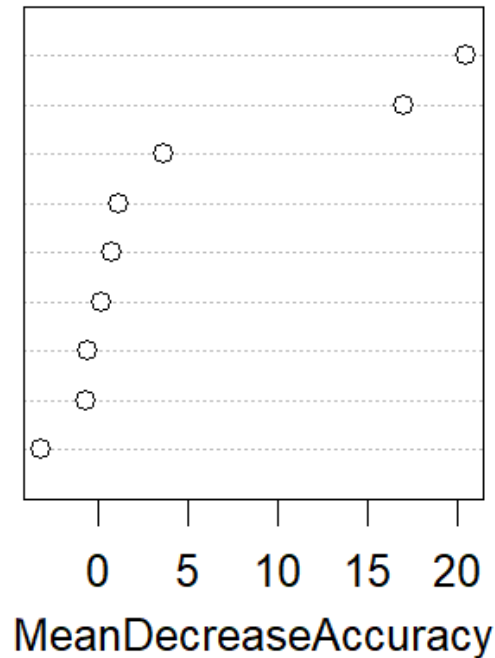
👉 Objetivo: *capturar dinámica, complejidad y predicción avanzada.*

OOB estimate of error rate: 37.7%

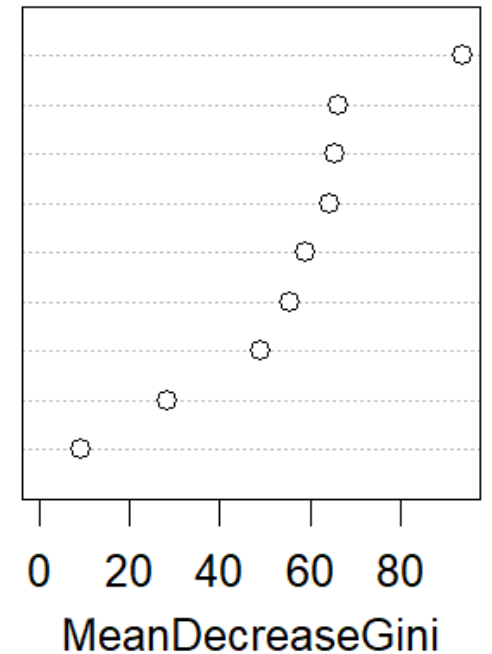
Confusion matrix:

	0	1	class.error
0	417	150	0.2645503
1	227	206	0.5242494

grupo  
qol\_post  
salud\_total\_pre  
dolor\_pre  
adherencia  
sexo  
edad  
apoyo\_familiar  
imcpases



qol\_post  
adherencia  
apoyo\_familiar  
dolor\_pre  
edad  
imcpases  
salud\_total\_pre  
grupo  
sexo



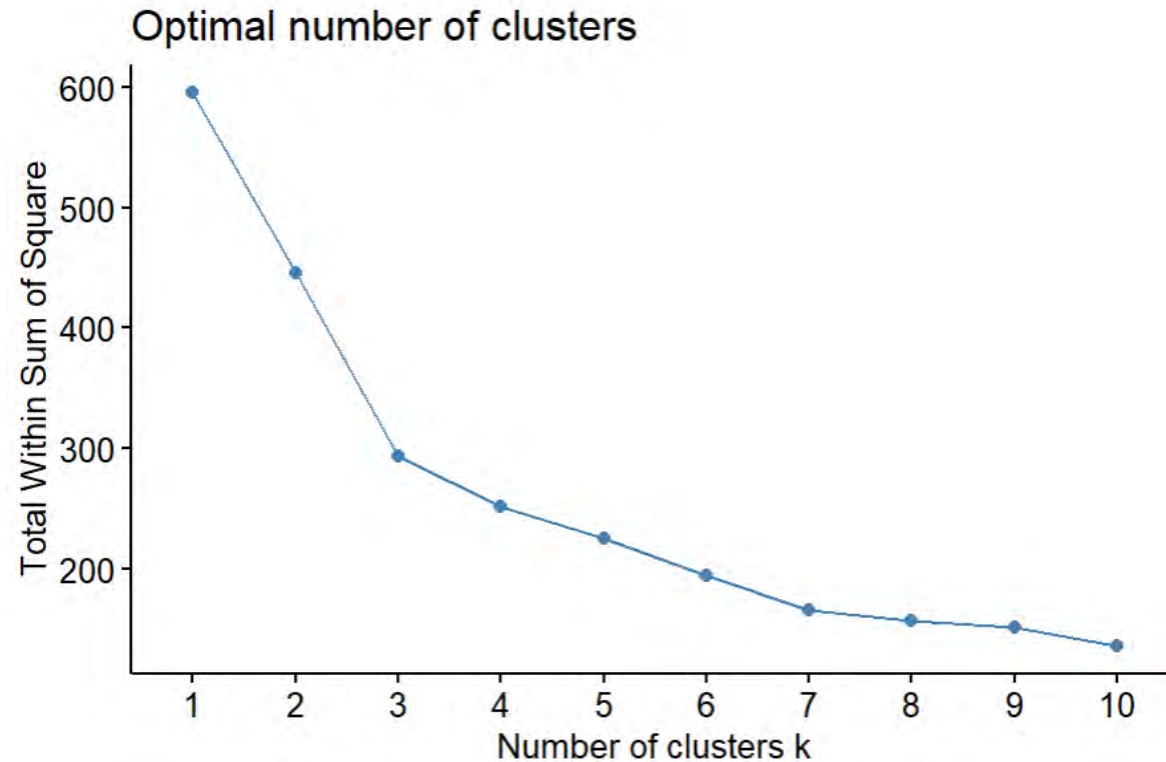
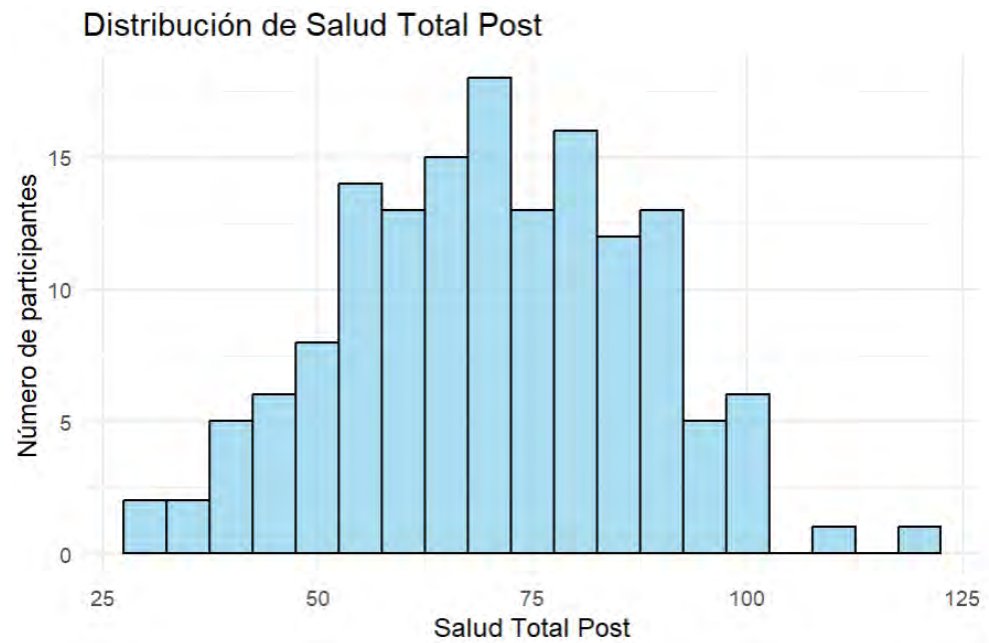
```
vars <- c("mejora_bienestar", "edad", "sexo",
"grupo", "adherencia", "apoyo_familiar",
"salud_total_pre", "dolor_pre", "imcpases",
"qol_post")
rf_data <- na.omit(datos[ , vars])
rf_data$sexo <- as.factor(rf_data$sexo)
rf_data$grupo <- as.factor(rf_data$grupo)
set.seed(123)
rf_model <-
randomForest(as.factor(mejora_bienestar) ~ .,
data = rf_data, ntree = 500, importance = TRUE)
print(rf_model)
varImpPlot(rf_model)
```

## 7. Análisis avanzados y longitudinales

- Modelos de efectos mixtos / jerárquicos (HLM, LMM)
- Análisis de supervivencia (Kaplan–Meier, regresión de Cox)
- Series temporales (ARIMA, modelos autorregresivos)
- Redes sociales y análisis de grafos
- Machine learning aplicado a ciencias sociales y salud (árboles de decisión, random forest, clustering, SVM).
- Modelos de ecuaciones estructurales (SEM)

👉 Objetivo: *capturar dinámica, complejidad y predicción avanzada.*

```
clust_vars <- datos_sim[, c("nivel_socio",  
"adherencia", "salud_total_pre")]  
clust_scaled <- scale(clust_vars)  
fviz_nbclust(clust_scaled, kmeans, method =  
"wss")  
set.seed(123)  
km_res <- kmeans(clust_scaled, centers = 3,  
nstart = 25)  
datos_sim$cluster <- factor(km_res$cluster)  
table(datos_sim$cluster)
```

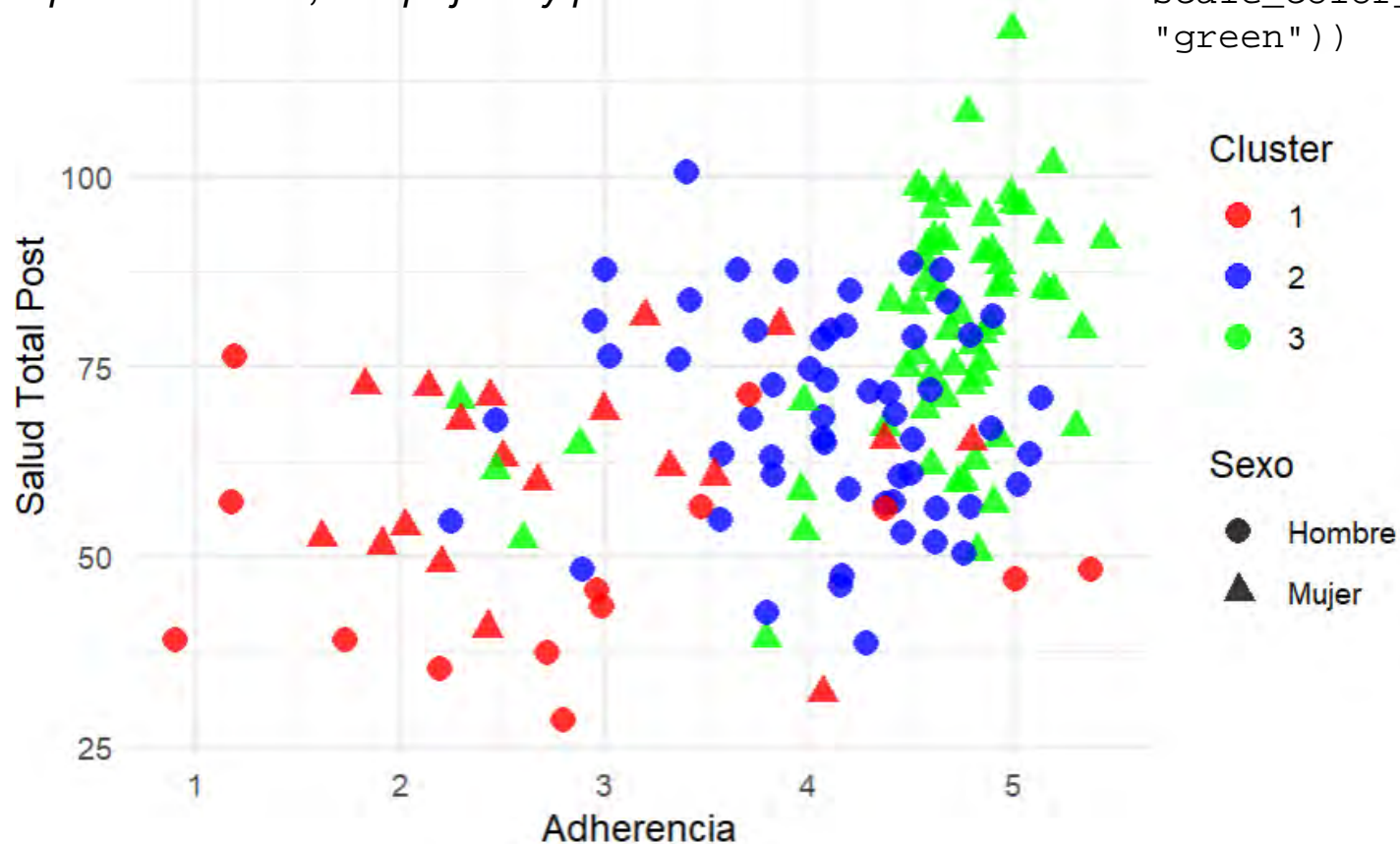


1 2 3  
0.4000000 0.2066667 0.3933333

## 7. Análisis avanzados y longitudinales

- Modelos de efectos mixtos / jerárquicos (HLM, LMM)
- Análisis de supervivencia (Kaplan–Meier, regresión de Cox)
- Series temporales (ARIMA, modelos autorregresivos)
- Redes sociales y análisis de grafos
- Machine learning aplicado a ciencias sociales y salud (árboles de decisión, random forest, clustering, SVM).
- Modelos de ecuaciones estructurales (SEM)

👉 Objetivo: *capturar dinámica, complejidad y predicción avanzada.*



```
ggplot(datos_sim, aes(x = adherencia, y = salud_total_post, color = cluster, shape = as.factor(sexo))) +  
  geom_point(size = 3, alpha = 0.8) +  
  labs(title = "Clustering de participantes según variables previas", x = "Adherencia", y = "Salud Total Post", color = "Cluster", shape = "Sexo") +  
  theme_minimal() +  
  scale_color_manual(values = c("red", "blue", "green"))
```

## 7. Análisis avanzados y longitudinales

- Modelos de efectos mixtos / jerárquicos (HLM, LMM)
- Análisis de supervivencia (Kaplan–Meier, regresión de Cox)
- Series temporales (ARIMA, modelos autorregresivos)
- Redes sociales y análisis de grafos
- Machine learning aplicado a ciencias sociales y salud (árboles de decisión, random forest, clustering, SVM).
- Modelos de ecuaciones estructurales (SEM)

👉 Objetivo: *capturar dinámica, complejidad y predicción avanzada.*

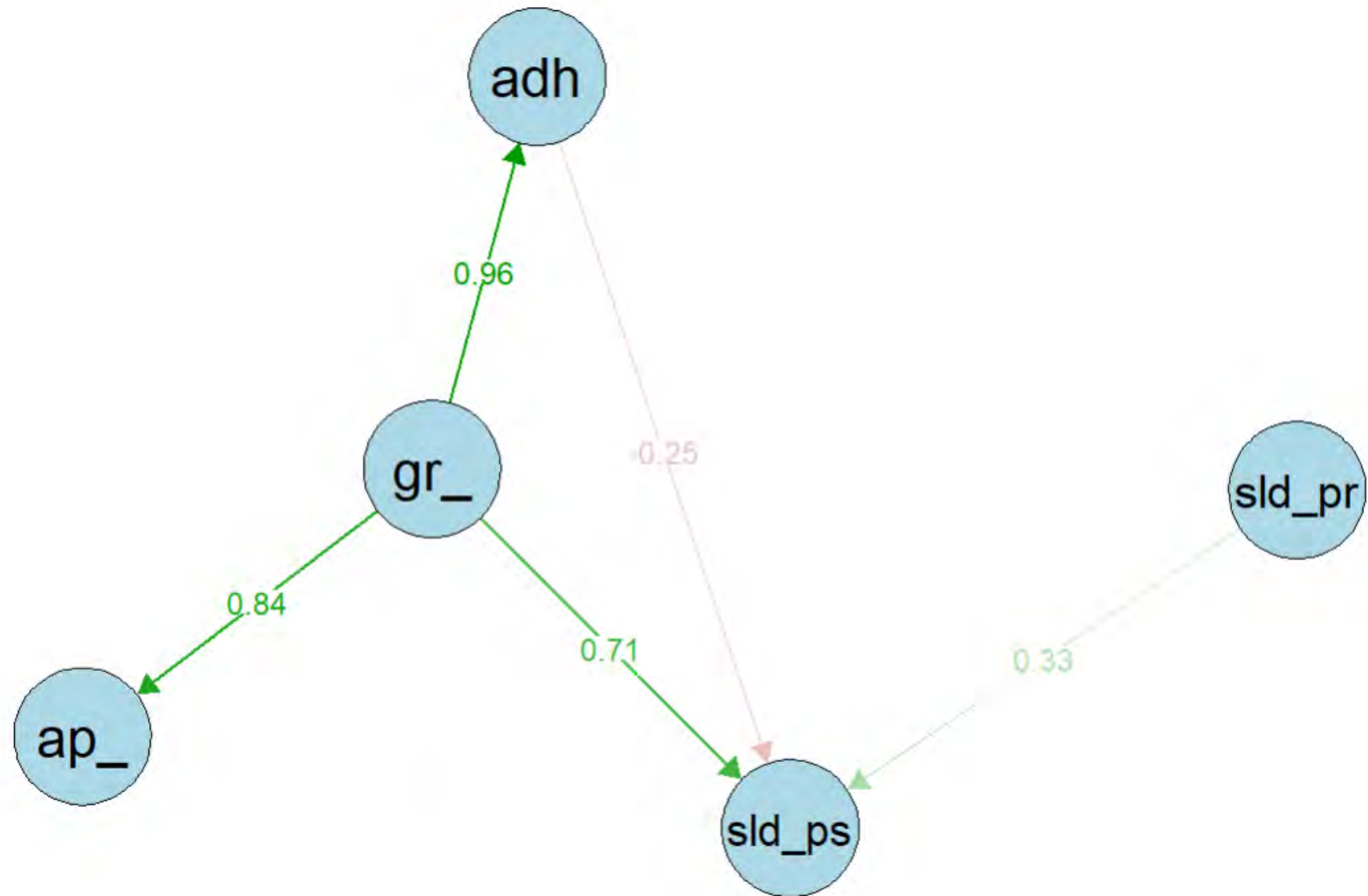
```
sem_model <- `
# Latentes
alud_pre  =~ salud1_pre + salud2_pre +
salud3_pre + salud4_pre + salud5_pre
alud_post =~ salud1_post + salud2_post +
salud3_post + salud4_post + salud5_post

# Regresión directa
salud_post ~ c*grupo_num + b1*adherencia +
b2*apoyo_familiar + b3*salud_pre

# Mediadores
adherencia ~ a1*grupo_num
apoyo_familiar ~
a2*grupo_num

# Efectos indirectos
indirect_adherencia := a1*b1
indirect_apoyo := a2*b2
total_effect := c + indirect_adherencia +
indirect_apoyo`
```

npar	chisq	df	pvalue	cfi	rmsea
40.0000	78.0405	62.0000	0.0821	0.9989	0.0227





# Propuesta de un estudio para estudiar los beneficios del programa de termalismo

## (4) Análisis de datos

- Enumera las preguntas de investigación derivadas de tus hipótesis
- Indica qué análisis estadístico permitiría responder a cada pregunta

Tabla sugerida:

| Pregunta | Análisis | Resultado esperado | Interpretación |

1

Introducción

2

Métodos

3

Análisis

4

Presentación

## ¿Por qué importa cómo comunicamos?

- Los resultados en evaluación de programas influyen en decisiones clave
- No basta con tener datos → hay que traducirlos en información útil
- Una mala comunicación = pérdida de credibilidad e impacto
- Formato: Lenguaje simple y preciso, evitar jerga técnica innecesaria, frases cortas, voz activa, explicar gráficos y tablas en el texto, destacar mensajes clave (negritas, recuadros, citas)
- Informes para distintos públicos:
  - **Técnicos:** detalle metodológico, datos completos
  - **Decisores:** síntesis ejecutiva, recomendaciones claras
  - **Comunidad:** lenguaje sencillo, resúmenes visuales

## Formato

- Formato: Lenguaje simple y preciso, evitar jerga técnica innecesaria, frases cortas, voz activa, explicar gráficos y tablas en el texto, destacar mensajes clave (negritas, recuadros, citas)
- Informes para distintos públicos:
  - **Técnicos:** detalle metodológico, datos completos
  - **Decisores:** síntesis ejecutiva, recomendaciones claras
  - **Comunidad:** lenguaje sencillo, resúmenes visuales

## Estructura de un informe técnico

- Resumen ejecutivo
- Contexto y objetivos (¿qué queremos responder?)
- Metodología aplicada (fuente, tamaño de muestra, variables)
- Resultados y análisis (cifras, tendencias, significación, tamaño del efecto, qué significa en la práctica)
- Plan de acción y recomendaciones
- Conclusiones y limitaciones (transparencia sobre posibles sesgos)
- ANEXOS (instrumentos, datos detallados)

...simplifica sin perder rigor, sé transparente, destaca lo accionable

## Ejemplo de redacción

- **ANOVA [tipo]:**  $F(gl_1, gl_2) = XX.XX$ ,  $p = .XXX$ ,  $\eta^2$  parcial = .XX, potencia = .XX.  
**Comparaciones post hoc:** diferencia media = XX, IC95% [LL, UL],  $p = .XXX$ .
- Se realizó un ANOVA de un factor para comparar los niveles de dolor entre tres grupos de intervención. Los resultados mostraron un efecto significativo del tipo de tratamiento,  $F(2, 87) = 5.42$ ,  $p = 0.006$ ,  $\eta^2$  parcial = 0.11, lo que indica un tamaño del efecto moderado.
- El análisis post hoc con corrección de Bonferroni reveló que el grupo de termalismo reportó significativamente menor dolor que el grupo control (diferencia media =  $-1.8$ , IC95% [ $-3.1, -0.5$ ],  $p = 0.004$ ). No se encontraron diferencias significativas entre los grupos de fisioterapia y control.
- El análisis de potencia post hoc mostró un poder estadístico de 0.82 para detectar efectos de tamaño moderado, lo que sugiere que la muestra fue adecuada para la hipótesis planteada.



## Prácticas cuestionables y ciencia abierta

2005

Essay

### Why Most Published Research Findings Are False

John P. A. Ioannidis

“...a research finding is **less likely to be true** when the **studies** conducted in a field are smaller; when **effect sizes** are smaller; when there is a greater number and lesser preselection of **tested relationships**; where there is greater **flexibility** in designs, definitions, outcomes, and analytical modes; when there is greater **financial and other interest** and prejudice; and when more teams are involved in a scientific field **in chase of statistical significance**”

## Prácticas cuestionables y ciencia abierta

2015

RESEARCH ARTICLE

### Estimating the reproducibility of psychological science

Open Science Collaboration<sup>\*,†</sup>

+ See all authors and affiliations

“We evaluated reproducibility using significance and **p-values, effect sizes, subjective assessments of replication teams, and meta-analysis** of effect sizes. The mean effect size ( $r$ ) of the replication effects ( $M_r = 0.197$ ,  $SD = 0.257$ ) was **half the magnitude of the mean effect size of the original effects** ( $M_r = 0.403$ ,  $SD = 0.188$ ), representing a substantial decline. **Ninety-seven** percent of original studies had significant results ( $P < .05$ ). **Thirty-six percent** of replications had significant results...”

“The claim that “we already know this” belies the uncertainty of scientific evidence. **Innovation points out paths that are possible; replication points out paths that are likely; progress relies on both**...”

## Prácticas cuestionables y ciencia abierta

### Causas de la crisis de confianza:

- Sesgo de publicación: muchas veces se persigue lo **novedoso** y **significativo**\*
- Esto favorece la aparición de prácticas cuestionables de investigación a través de los **grados de libertad del investigador**\*\*

\*Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General*, 143(2), 534.

\*\*Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359-1366.

\*\*\*Schwab, A., & Starbuck, W. H. (2017). A call for openness in research reporting: How to turn covert practices into helpful tools. *Academy of Management Learning & Education*, 16(1), 125-141.

## Prácticas cuestionables y ciencia abierta

Tres pequeñas mentiras\*:

- **Selective reporting**

Ejemplo: en el informe se muestran solo mejoras en calidad de vida y depresión. Se ocultan resultados sin cambios (dolor, movilidad)

- **HARKing (Hypothesizing After Results Are Known)**

Ejemplo: hipótesis inicial → “reducir dolor”. Tras ver datos → se cambia a “mejorar calidad de vida subjetiva”

- **p-Hacking**

Ejemplo: múltiples análisis → casi todos no significativos. Pero en un subgrupo (mujeres >70) aparece  $p = 0.04$ . Solo se reporta este resultado como “prueba” de eficacia

\*Schwab, A., & Starbuck, W. H. (2017). A call for openness in research reporting: How to turn covert practices into helpful tools. *Academy of Management Learning & Education*, 16(1), 125-141.

## Grados de libertad del investigador

Code	Related	Type of degrees of freedom
Hypothesizing		
T1	R6	Conducting explorative research without any hypothesis
T2		Studying a vague hypothesis that fails to specify the direction of the effect
Design		
D1	A8	Creating multiple manipulated independent variables and conditions
D2	A10	Measuring additional variables that can later be selected as covariates, independent variables, mediators, or moderators
D3	A5	Measuring the same dependent variable in several alternative ways
D4	A7	Measuring additional constructs that could potentially act as primary outcomes
D5	A12	Measuring additional variables that enable later exclusion of participants from the analyses (e.g., awareness or manipulation checks)
D6		Failing to conduct a well-founded power analysis
D7	C4	Failing to specify the sampling plan and allowing for running (multiple) small studies
Collection		
C1		Failing to randomly assign participants to conditions
C2		Insufficient blinding of participants and/or experimenters
C3		Correcting, coding, or discarding data during data collection in a non-blinded manner
C4	D7	Determining the data collection stopping rule on the basis of desired results or intermediate significance testing

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in psychology*, 7, 1832.

## Grados de libertad del investigador

Analyses		
A1		Choosing between different options of dealing with incomplete or missing data on <i>ad hoc</i> grounds
A2		Specifying pre-processing of data (e.g., cleaning, normalization, smoothing, motion correction) in an <i>ad hoc</i> manner
A3		Deciding how to deal with violations of statistical assumptions in an <i>ad hoc</i> manner
A4		Deciding on how to deal with outliers in an <i>ad hoc</i> manner
A5	D3	Selecting the dependent variable out of several alternative measures of the same construct
A6		Trying out different ways to score the chosen primary dependent variable
A7	D4	Selecting another construct as the primary outcome
A8	D1	Selecting independent variables out of a set of manipulated independent variables
A9	D1	Operationalizing manipulated independent variables in different ways (e.g., by discarding or combining levels of factors)
A10	D2	Choosing to include different measured variables as covariates, independent variables, mediators, or moderators
A11		Operationalizing non-manipulated independent variables in different ways
A12	D5	Using alternative inclusion and exclusion criteria for selecting participants in analyses
A13		Choosing between different statistical models
A14		Choosing the estimation method, software package, and computation of SEs
A15		Choosing inference criteria (e.g., Bayes factors, alpha level, sidedness of the test, corrections for multiple testing)

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in psychology*, 7, 1832.

## Grados de libertad del investigador

### Reporting

- R1 Failing to assure reproducibility (verifying the data collection and data analysis)
- R2 Failing to enable replication (re-running of the study)
- R3 Failing to mention, misrepresenting, or misidentifying the study preregistration
- R4 Failing to report so-called “failed studies” that were originally deemed relevant to the research question
- R5 Misreporting results and  $p$ -values
- R6 T1 Presenting exploratory analyses as confirmatory (HARKing)

Wicherts, J. M., Veldkamp, C. L., Augusteyn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in psychology*, 7, 1832.

## Prácticas cuestionables y ciencia abierta

### Soluciones a la crisis de confianza:

- Pre-registro de estudios científicos ([OSF](#))
- Ciencia abierta: artículos, datos y procedimientos (códigos)

13. Sample size (required)
  - 13.1. Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, interactions, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, how many units are you collecting at each level of the analysis?
  - 13.2. **Example:** Our target sample size is 280 participants. We will attempt to recruit up to 320, assuming that not all will complete the total task.
  - 13.3. **More information:** For some studies, this will simply be the number of samples or the number of clusters. For others, this could be an expected range, minimum, or maximum number.



\*Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General*, 143(2), 534.

\*\*Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359-1366.

\*\*\*Schwab, A., & Starbuck, W. H. (2017). A call for openness in research reporting: How to turn covert practices into helpful tools. *Academy of Management Learning & Education*, 16(1), 125-141.

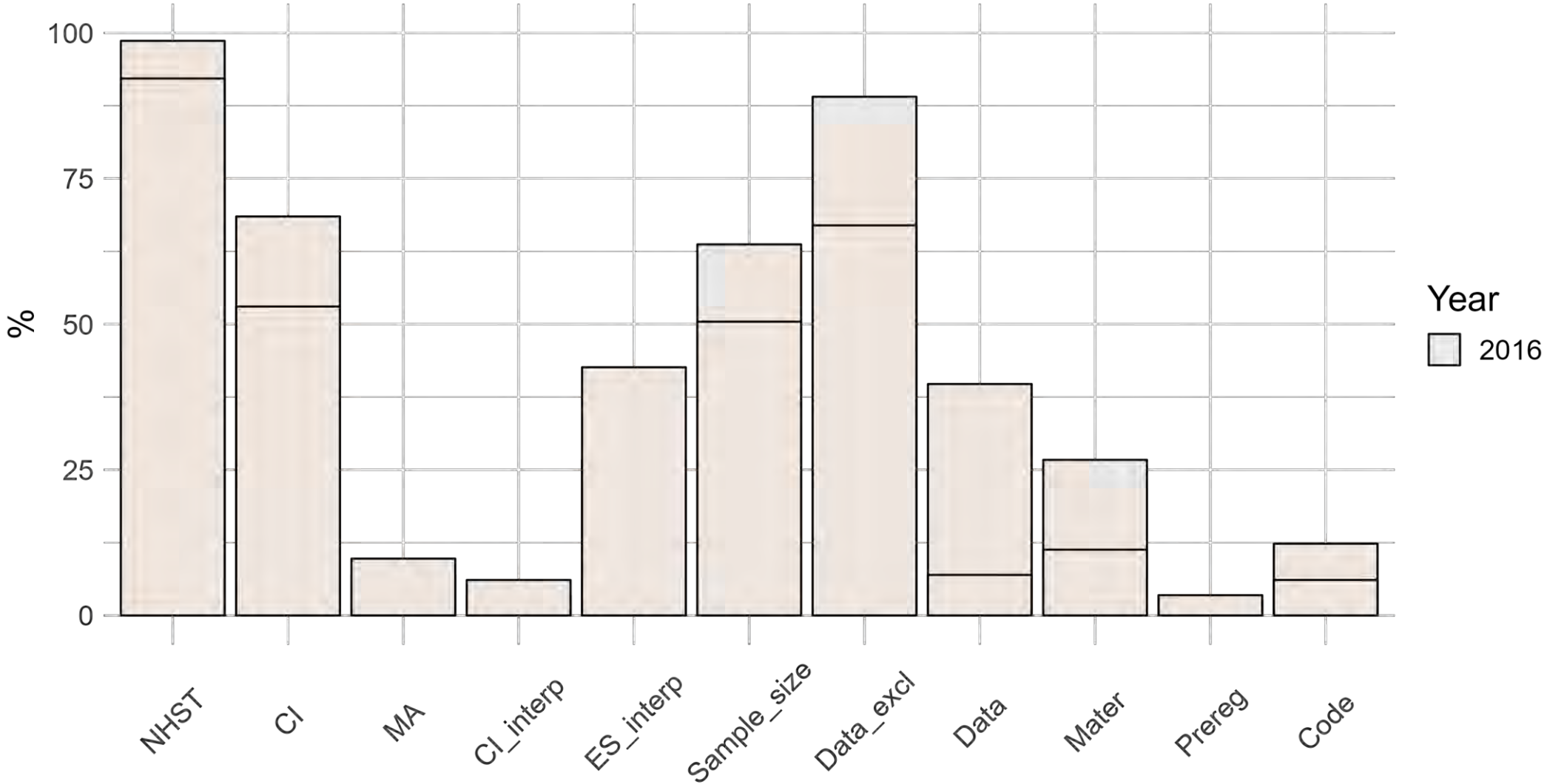
## Prácticas cuestionables y ciencia abierta

- Ciencia abierta: Guías sobre cómo reportar estadísticos\*

1. **NHST.** Reportan un  $p$ -value
2. **CI.** Reportan un intervalo de confianza
3. **MA.** Incluye un metanálisis (en el caso de más de un experimento)
4. **CI\_interp.** El intervalo de confianza es interpretado
5. **ES\_interp.** Se reporta e interpreta un tamaño del efecto
6. **Sample\_size.** Se justifica el tamaño muestral (estudio de potencia)
7. **Data\_excl.** Se indica si hubo criterio de exclusión
8. **Data.** “data are available upon request”
9. **Mater.** “materials are available upon request”
10. **Preregistered.** ¿Hubo preregistro?
11. **Code.** “the computer code or syntax (e.g., R code) needed to reproduce the analyses is available upon request”

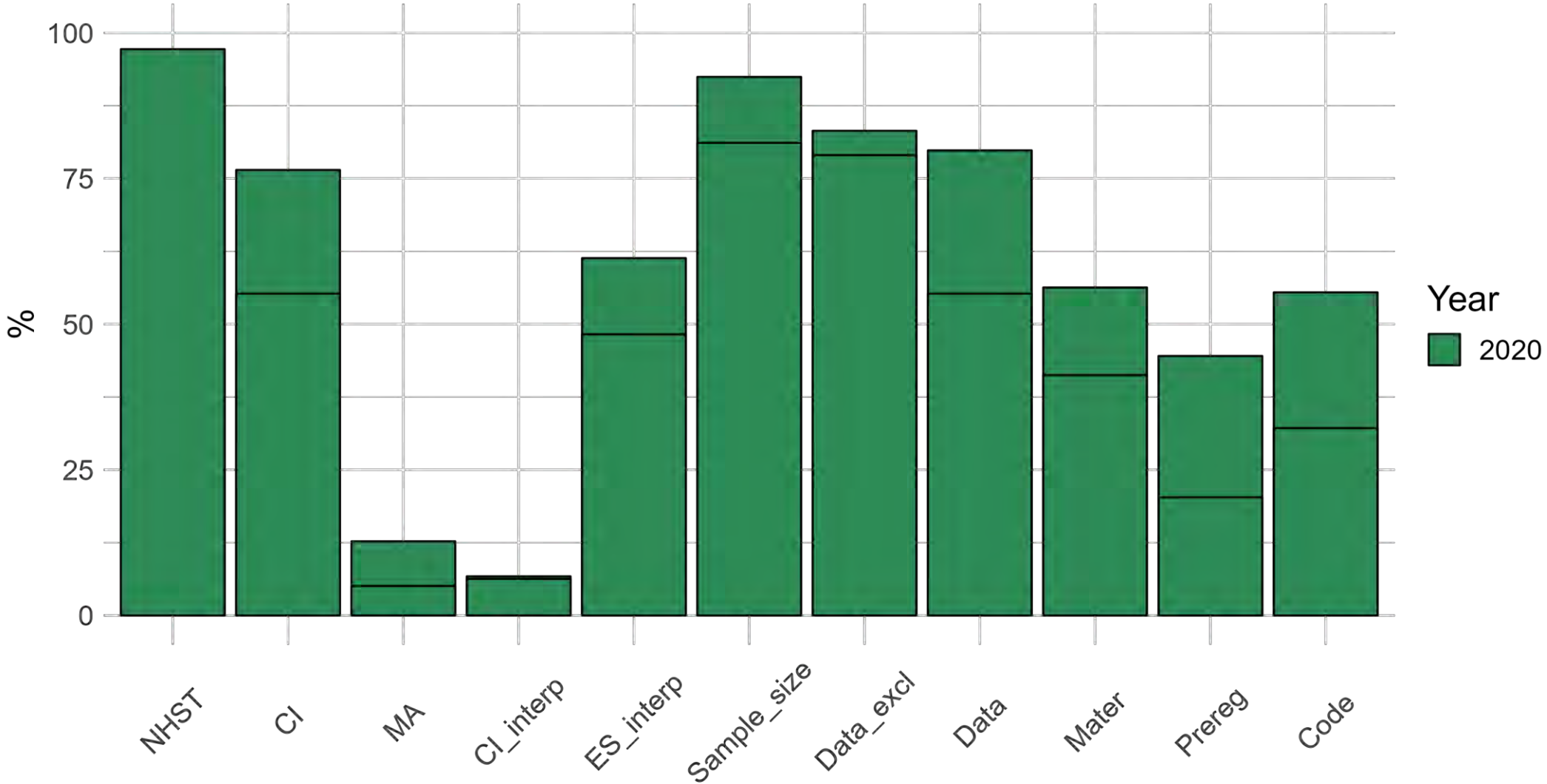
Giofrè, D., Boedker, I., Cumming, G., Rivella, C., & Tressoldi, P. (2022). The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices: Five years later. *Behavior Research Methods*

### Both journals



Giofrè, D., Boedker, I., Cumming, G., Rivella, C., & Tressoldi, P. (2022). The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices: Five years later. *Behavior Research Methods*

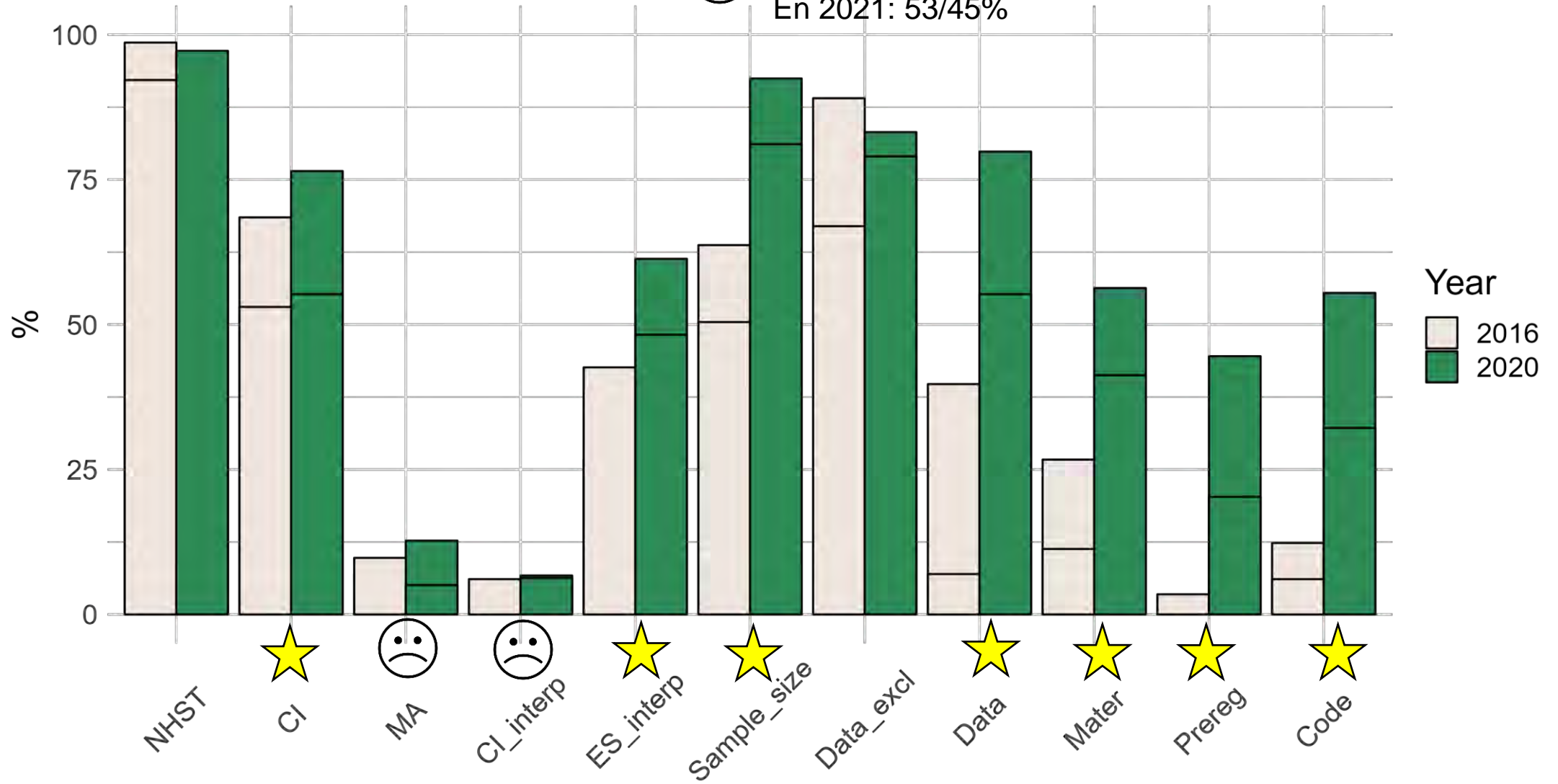
### Both journals



Giofrè, D., Boedker, I., Cumming, G., Rivella, C., & Tressoldi, P. (2022). The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices: Five years later. *Behavior Research Methods*

# Both journals

☹️ Análisis a priori de potencia  
 En 2016: 26/37%  
 En 2021: 53/45%



Giofrè, D., Boedker, I., Cumming, G., Rivella, C., & Tressoldi, P. (2022). The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices: Five years later. *Behavior Research Methods*

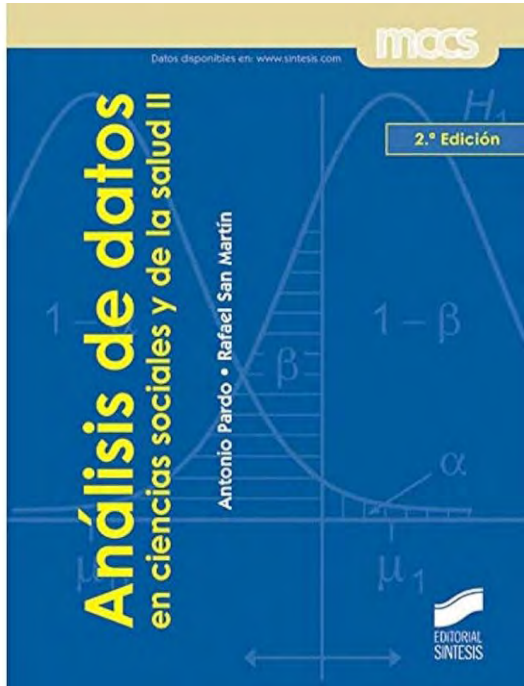
“**Data**/code/materials is available upon request to the corresponding author”

“Entre 1792 manuscritos en los que el DAS [Data availability statement] indicaba que los autores estaban dispuestos a compartir sus datos, 1669 **(93%) autores no respondieron o se negaron** a compartir sus datos con nosotros” (Gabelica et al., 2022)

Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: mixed-methods study. *Journal of Clinical Epidemiology*.

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., et al. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Sci. Data*

## Bibliografía



Pardo y San Martín



Rey y Estrada



Ximénez y Revuelta



Abad, Olea, Ponsoda y García

## Contactos

Departamentos de Metodología en las Facultades de Psicología



## **Institucionalizar la evaluación, clave para el impacto social**

La evaluación rigurosa y transparente es fundamental para mejorar programas y políticas sociales

Fomentar la colaboración entre organismos públicos y universidades permite avanzar en metodologías y resultados

Apostar por la cultura de la evaluación es apostar por la mejora continua y la rendición de cuentas

## Resumen de la jornada

- **Importancia de la evaluación del impacto social**

Medir para mejorar: más allá de los indicadores económicos, centrarse en el cambio real en la vida de las personas.

- **Metodologías y diseños de evaluación**

Diferencias entre evaluación experimental, cuasi-experimental y ex post facto. Relevancia de la validez interna y externa, y el control de variables.

- **Instrumentos y análisis de datos**

Selección y justificación de instrumentos fiables y válidos unida a un análisis de datos dirigido a responder preguntas relevantes

- **Buenas prácticas y retos actuales**

Evitar prácticas cuestionables, fomentar la colaboración entre organismos



*¡Gracias!*  
Miguel A. Sorrel  
(miguel.sorrel@uam.es)

