

# Proceso y Etapas del Diseño de la Investigación: Preguntas y Credibilidad

Fill In Your Name

16-08-2021

# El Proceso y las Etapas del Diseño de la Investigación

## El Proceso de Investigación

## El Formulario para Diseño para el Diseño la Investigación de EGAP

## DeclareDesign

## Prerregistro del plan de análisis

## Resumen

## Referencias

# El Proceso y las Etapas del Diseño de la Investigación

# ¿Qué hace a una pregunta de investigación buena?

- ▶ La respuesta a una buena pregunta de investigación debe generar conocimiento que le interese a la gente.
- ▶ Abordar la pregunta debería (ayudar a) resolver un problema, tomar una decisión o aclarar/desafiar nuestro entendimiento del mundo.
- ▶ Pero una pregunta interesante no es suficiente

# También necesitamos un buen diseño de investigación

- ▶ Un buen diseño de investigación es un plan práctico de investigación que aprovecha los recursos disponibles y produce una respuesta creíble.
- ▶ La calidad de un diseño de investigación se puede evaluar por los resultados que produce, los cuales pueden utilizarse para orientar las políticas y contribuir a la ciencia:
  - ▶ Un gran diseño de investigación produce resultados que apuntan claramente en ciertas direcciones que son importantes.
  - ▶ Un diseño de investigación deficiente produce resultados que nos dejan en la oscuridad, resultados con una interpretación confusa.

# La importancia de la teoría I

Todo diseño de investigación involucra una teoría, ya sea implícita o explícita.

- ▶ ¿Por qué investigar? Tenemos teorías y valores implícitos que guían las preguntas que hacemos. Nuestras preguntas reflejan nuestros valores; por ejemplo, los científicos sociales estudiaron el consumo de marihuana en la década de 1950 como una forma de “desviación,” las preguntas se centraron en “¿por qué la gente está tomando decisiones tan malas?” o “¿cómo pueden los legisladores prevenir el consumo de marihuana?” (ver **Becker (1998)**).
- ▶ ¿Por qué investigar? Puede que queramos cambiar la forma en la que la ciencia explica el mundo y/o cambiar las decisiones de políticas para (a) un lugar y un momento en el tiempo y/o (b) para otros lugares y momentos en el tiempo.

## La importancia de la teoría II

- ▶ La investigación que está centrada en conocer el efecto causal de  $X$  en  $Y$  requiere un modo del mundo: *cómo* la intervención  $X$  podría tener un efecto en alguna variable de interés  $Y$ , y *por qué* y *qué tan grande* podría ser el efecto. Esto nos ayuda a pensar en cómo una intervención diferente o que se dirija a diferentes destinatarios podría llevar a resultados diferentes.
- ▶ Nuestras teorías y modelos no son solo importantes para generar hipótesis, sino también para informar *el diseño* y *las estrategias de inferencia*.
- ▶ El diseño de la investigación a menudo aclarará dónde tenemos menos certeza acerca de nuestras teorías. Nuestras teorías nos indicarán dónde hay problemas con nuestro diseño. Y las preguntas que surjan del proceso de diseño pueden dejarnos ver la necesidad de trabajar más en la explicación y el mecanismo.

# Diseñando o escogiendo el tratamiento

- ▶ Desde este punto en adelante, usaremos  $T$  para el tratamiento o el efecto en el que estamos interesados. Usaremos  $X$  para referirnos a las variables de fondo.
- ▶ El tratamiento ( $T$ ) y control (la ausencia de  $T$ ) deben conectarse claramente a la pregunta de investigación. (Ver el módulo de Medición).
- ▶ Un tratamiento de interés puede ser un conjunto de varios componentes. Si su pregunta de investigación es sobre un componente específico, entonces el control debe ser diferente del tratamiento solo en ese componente. Todo lo demás debería ser igual.



## Un ejemplo

Una campaña en la que alguien visita un hogar para hablar con una familia durante 15 minutos para dar información sobre temas de salud.

- ▶ Si se tiene interés en el efecto de la información en específico, entonces el control debería tener todos los demás componentes (visita domiciliaria con una duración de 15 minutos, visitante similar, etc.) pero variar en la información que se comparte. Este diseño no enseñará sobre el efecto de las visitas, sino solo sobre el efecto de la información en sí.
- ▶ Si su pregunta se centra en el efecto de las visitas, entonces necesita un grupo de control que no recibiera visitas. Pero este diseño no servirá para responder preguntas específicas sobre la información (las visitas y la información están combinadas en un sólo tratamiento).

# Interpretación

- ▶ A veces no es posible aislar un componente específico del tratamiento.
- ▶ Por ejemplo, puede que la organización de salud comunitaria con la que usted esté asociado para visitar hogares no esté interesada en visitar hogares y compartir otra información. En ese caso el control podría ser no visitar.
- ▶ Debe tener cuidado de interpretar sus efectos como el efecto de la información impartida de esta manera en particular.
- ▶ No se podrá concluir que se ha estimado el efecto de la información únicamente .
  - ▶ Puede que esto esté bien para ciertas políticas: quizá la pregunta de políticas pueda referirse a las visitas como un paquete implícito de tratamientos.
  - ▶ Pero es difícil interpretar los resultados de este diseño como si nos dijera algo claro acerca de la información únicamente.

# El Proceso de Investigación

# Una descripción general del proceso de investigación

- ▶ Articular y afinar la pregunta (interrogando por qué se está haciendo esta pregunta y qué sucedería si se dan diferentes tipos de respuestas).
- ▶ Desarrollar el diseño de la investigación.
- ▶ Planificar el análisis, presentar y justificar hipótesis específicas y estampar la fecha en el plan con un sello impersonal creíble.
- ▶ Implementar la intervención y recopilar datos.
- ▶ Analizar los datos y escribir los resultados

# El Formulario para Diseño para el Diseño la Investigación de EGAP

# El Formulario para Diseño para el Investigación de EGAP

- ▶ Desarrollamos un Formulario para el Diseño de la Investigación que ayuda a crear una estructura para un buen diseño de investigación.
  - ▶ <https://egap.github.io/learningdays-gui/Exercises/design-form.Rmd>
- ▶ Le puede ayudar a:
  - ▶ escribir una propuesta de investigación para solicitar financiación para un proyecto, o
  - ▶ a desarrollar un plan de pre-análisis.

# Secciones del formulario de diseño de investigación EGAP

- ▶ Pregunta de investigación
- ▶ Muestra
- ▶ Tratamiento
- ▶ Variable de interés
- ▶ Estrategia de aleatorización
- ▶ Implementación
- ▶ Poder estadístico
- ▶ Análisis e interpretación

# Pregunta de investigación y motivación

- ▶ ¿Cuál es la motivación sustantiva de esta investigación? ¿Que problema está tratando de abordar? ¿Qué decisión está tratando de tomar?
- ▶ ¿A quiénes está tratando de cambiar de parecer y qué creen esas personas actualmente?
- ▶ ¿Qué preguntas teóricas se pueden abordar con esta investigación?
- ▶ Concretar pregunta de investigación en una sola oración.
- ▶ ¿Cuál es la hipótesis principal?



# Muestra

- ▶ ¿Dónde y cuándo tendrá lugar el estudio?
- ▶ ¿Quiénes o cuáles son las unidades que hay en el estudio?
- ▶ ¿Cómo se selecciona la muestra?
- ▶ ¿Es necesario dejar fuera del estudio a algunas unidades porque deben recibir el tratamiento, o por razones logísticas o de otro tipo, algunas unidades deben quedar por fuera?
- ▶ ¿Se espera que el tratamiento funcione de manera diferente para ciertos subgrupos?

# Tratamiento

- ▶ ¿Cuál es el tratamiento? ¿Habrá múltiples tratamientos?
- ▶ ¿Cuál será el control? ¿Control puro o placebo?
- ▶ ¿Existe alguna preocupación ética con el tratamiento?
- ▶ ¿A qué nivel se aleatorizará el tratamiento?

# Variable de Interés

- ▶ ¿Cuál es es la variable de interés principal?
- ▶ ¿Cómo se medirá?
- ▶ ¿Qué datos se necesitan? ¿A qué nivel se pueden medir los datos?
- ▶ ¿Qué ideas tiene usted sobre la variable de interés previo a la recolección de datos (priors)? Estas pueden venir de estudios previos o suposiciones bien fundamentadas.
- ▶ ¿Cuántas rondas de recolección de datos habrá?
- ▶ ¿Cómo minimizará la deserción?
- ▶ ¿Cómo minimizará la mala medición y las respuestas falsas?

# Estrategia de aleatorización

- ▶ ¿Qué tipo de estrategia de aleatorización utilizará? Ejemplos: simple, completo, bloques, conglomerados, factorial, de dos niveles, por etapas, etc.
- ▶ Asegúrese de que esta estrategia sea coherente con el nivel al que se está aleatorizando (posiblemente conglomerados) y la heterogeneidad que se espera en los efectos del tratamiento (posiblemente bloques).
- ▶ Defina sus bloques y conglomerados (si los hay). ¿Cuántos va a tener? ¿Qué tan grandes serán?
- ▶ ¿Es posible que haya interferencia? Si es así, ¿cómo le ayudará su muestra y la estrategia de aleatorización a minimizar la interferencia?

# Implementación I

- ▶ ¿Cómo hará la aleatorización en la práctica? ¿En público, sacando una ficha de una urna? ¿En una computadora?
- ▶ ¿Quién estará a cargo de implementar el tratamiento?
- ▶ Si es un socio el que implementará el tratamiento, ¿qué acuerdos tienen?
- ▶ ¿Cuáles son los desafíos logísticos? ¿Algún desafío especial para las unidades de control?

# Implementación II

- ▶ ¿Cómo le hará seguimiento a la calidad de la implementación?
- ▶ ¿Cómo hará seguimiento del cumplimiento del tratamiento?
- ▶ ¿Cómo minimizará el incumplimiento del tratamiento (si aplica)?
- ▶ ¿Cómo comprobará la calidad de sus datos?
- ▶ ¿Cómo se anonimizarán y almacenarán los datos de forma segura (si aplica)?

- ▶ ¿Cuál es el tamaño esperado del efecto?
  - ▶ Esta información puede venir de un estudio anterior o un valor objetivo por debajo del cual uno no estaría interesado en futuras intervenciones.
- ▶ Cálculo del poder.
  - ▶ Si tiene conglomerados, se debe tener en cuenta la correlación dentro de los conglomerados.

# Análisis e interpretación

- ▶ ¿Cuál es el estimado? (p. ej., efecto promedio del tratamiento, efecto causal promedio del cumplidor, etc.)
- ▶ ¿Cuál es el estimador? (por ejemplo, diferencia de medias, MCO con pesos de bloque, algún conglomerado). Tenga en cuenta que esto debería estar estrechamente relacionado con su diseño de aleatorización.
- ▶ Si encuentra que sus resultados son consistentes con su hipótesis, ¿qué explicaciones alternativas podría haber? ¿Qué datos le ayudarían a distinguir entre su explicación y otras alternativas?
- ▶ Si encuentra que sus resultados no son consistentes con su hipótesis, ¿qué datos le ayudarán a averiguar qué pudo haber sucedido?



# DeclareDesign

# Introducción a DeclareDesign

- ▶ Declare Design es un paquete de software en R.
- ▶ Nos ayuda a concretar las etapas del diseño de la investigación al permitirnos representarlas en código, lo que luego nos permitirá simular las etapas del diseño de la investigación para comprender las propiedades de los estimadores estadísticos y las pruebas que utilizamos.
- ▶ Para obtener más información, consulte (<https://declaredesign.org/getting-started/>)
- ▶ También puede consultar el módulo sobre estimaciones y estimadores que utiliza DeclareDesign para ayudar a determinar los estimadores apropiados.

# Introducción a DeclareDesign

- ▶ Ver <https://declaredesign.org/>
- ▶ Independientemente del método, los diseños de investigación tienen cuatro componentes
- ▶ MIDR:
  - ▶ M: Modelo (de cómo funciona el mundo)
  - ▶ I: Indagación
  - ▶ D: Datos ¿Cuál es la estrategia?
  - ▶ R: Respuesta ¿Cuál es la estrategia?
- ▶ Perspectiva crítica: Simular el diseño de investigación nos puede enseñar qué respuestas puede dar un diseño de investigación.
- ▶ Trabajar con datos simulados *antes de la recopilación de datos* ayuda a prevenir errores y descuidos.

# Modelo

- ▶ Un modelo de cómo pensamos que el mundo funciona, que incluye:
  - ▶  $T$ s y  $X$ s (tratamientos o variables causales focales como intervenciones de políticas y otras variables de fondo)
  - ▶  $Y$ s (variables dependientes)
  - ▶ Relaciones entre variables (salidas potenciales, formas funcionales, variables auxiliares y contextos)
  - ▶ Distribución de probabilidad sobre  $X$ s y quizá también sobre  $Y$ s.
- ▶ ¡Esta es la teoría!
  - ▶ Codificada numéricamente.
- ▶ Los modelos siempre serán incorrectos por definición. Si fueran correctos, no necesitaríamos realizar un estudio.
- ▶ Pero sin un modelo no tenemos cómo comenzar a evaluar lo que se puede aprender.

# Indagación

- ▶ Una pregunta con respuesta.
- ▶ ¿Cuál es el efecto de un tratamiento  $T$  sobre una variable  $Y$ ?
- ▶ Suele ser una cantidad de interés que resume los datos:
  - ▶ Descriptiva: ¿Cuál es la media de  $Y$  en el tratamiento formalmente?.
  - ▶ Causal: ¿Cuál sería la diferencia promedio de  $Y$  si cambiáramos el tratamiento por el control? Si afirmamos que  $T$  no tiene ningún efecto en  $Y$ , ¿qué tanta evidencia tendríamos para hacer esta afirmación?
  - ▶ La cantidad es el estimando o la hipótesis.
- ▶ No todas las preguntas que queremos hacer tienen respuesta.
  - ▶ Y la gama de indagaciones que podemos hacer es limitada: ¿cuánto podemos aprender de alguna cantidad que resume los datos, como el efecto de tratamiento promedio (ATE)?

# Datos

- ▶ Recolectar (generar) datos sobre el conjunto de variables (todas:  $X$ s,  $T$ s y  $Y$ s)
- ▶ Una función del modelo
- ▶ Incluye ambos:
  - ▶ Muestreo: ¿cómo llegan las unidades a su muestra?
  - ▶ Asignación de tratamiento: ¿qué valores de las variables endógenas se revelan?

# Respuesta

- ▶ Generar una respuesta dada la realización de los datos, es decir, una estimación de la cantidad de interés (indagación)
- ▶ Este es su estimador o prueba:
  - ▶ Diferencia de medias
  - ▶ prueba  $t$
  - ▶ Métodos de regresión
  - ▶ etc.
- ▶ La respuesta es una estimación de la cantidad de interés o valor de  $p$  (indagación/estimación /prueba)

## Prerregistro del plan de análisis



# Sesgo en la investigación publicada contra resultados nulos

- ▶ Anticipándose a las dificultades que hay al publicar, o para no enfrentarse ellas, manuscritos con resultados nulos no se envían nunca para revisión o se guardan en un “cajón de archivos” después de varios rechazos.
- ▶ Todos enfrentamos incentivos para cambiar nuestras especificaciones, medidas o incluso hipótesis para obtener un resultado estadísticamente significativo ( $p$ -hacking) para mejorar posibilidades de publicación.
- ▶ Incluso cuando no se enfrentan estos incentivos se toman muchas decisiones al analizar datos: manejo de valores perdidos y observaciones duplicadas, creación de escalas, etc. Y estas elecciones pueden tener consecuencias.
- ▶ Resultado general: menor credibilidad en los trabajos de investigación individuales y (con razón) menor confianza en si realmente sabemos lo que decimos saber.

# Hacia la revisión del diseño en lugar de los resultados

- ▶ Una parte de la solución de este problema es centrarse en el diseño, más que en los resultados.
- ▶ El sesgo contra los resultados nulos se puede combatir revisando el diseño antes de conocer los resultados.
- ▶ Un buen diseño bien ejecutado producirá una investigación creíble, incluso con resultados nulos. Queremos que los resultados nulos sean creíbles y procesables.
- ▶ La revisión del diseño también es una oportunidad para mejorar la investigación antes de que sea implementada.

# Prerregistro de planes de análisis y diseños de investigación

## I

- ▶ El prerregistro es la presentación de su diseño de investigación e hipótesis en un repositorio de acceso público. EGAP aloja uno que puede usar de forma gratuita (actualmente en OSF.io utilizando el formulario de registro EGAP).
- ▶ El prerregistro no implica excluir análisis exploratorios posteriores que no fueron registrados con anticipación. Simplemente se debe distinguir claramente entre los dos.

# Prerregistro de planes de análisis y diseños de investigación II

- ▶ Incluso si va a enviar un artículo con resultados en lugar de un diseño para una revista académica o está interesado principalmente en un informe final con hallazgos para una audiencia en el campo de políticas públicas, existen ventajas importantes para usted y para otros investigadores del prerregistro de su investigación.
  - ▶ Puede obtener información sobre otras investigaciones, completadas y en curso; otros pueden aprender sobre su trabajo. Podemos aprender de estudios que produjeron resultados nulos.
  - ▶ Le obliga a plantear sus hipótesis y plan de análisis antes de ver los resultados, lo que limita el  $p$ -hacking.

# Resumen

# El proceso de investigación: preguntas, teoría y credibilidad

- ▶ La investigación parte de nuestros valores y teorías sobre cómo funciona el mundo.
- ▶ Continúa articulando preguntas que se pueden abordar de forma clara mediante la observación (en este curso, utilizando experimentación aleatoria).
- ▶ Las buenas preguntas tienen respuestas consecuentes: cambian las explicaciones científicas y/o cambian las decisiones de políticas.

# El proceso de investigación: preguntas, teoría y credibilidad

- ▶ Los buenos diseños satisfacen todos los requisitos y dan a los lectores razones para creer en los resultados.
- ▶ Las listas de verificación (como el formulario de diseño de investigación o los formularios de preinscripción) ayudan a evitar errores por descuido.
- ▶ El prerregistro aumenta aún más la credibilidad y, por lo tanto, las probabilidades de que sus resultados tengan un impacto en la ciencia y en las políticas.

## Referencias



# Referencias

Becker, Howard S. 1998. *Tricks of the Trade : How to Think about Your Research While You're Doing It (Chicago Guides to Writing, Editing, and Publishing)*. University Of Chicago Press.