

# Proceso y etapas del diseño de investigación: Preguntas y credibilidad

Fill In Your Name

02-03-2022

El Proceso y las etapas del diseño de investigación

El Proceso de investigación

El Formulario para el diseño de investigación de EGAP

DeclareDesign

Pre-registro del plan de análisis

Resumen

Referencias

# El Proceso y las etapas del diseño de investigación

# ¿Qué hace que una pregunta de investigación sea 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 excelente diseño de investigación produce resultados que nos guían claramente en ciertas direcciones que son relevantes.
  - ▶ Un diseño de investigación deficiente produce resultados que nos dejan en la oscuridad, o sea, resultados cuya interpretación es confusa.

# La importancia de la teoría I

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

- ▶ ¿Por qué investigar? Todos tenemos teorías y valores implícitos que guían las preguntas que hacemos. Nuestras preguntas reflejan nuestros valores. Por ejemplo, en la década de los 1950 científicos sociales estudiaron el consumo de marihuana 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 pensar en un modelo del mundo: *cómo* la intervención  $X$  podría afectar a una variable de resultado  $Y$ , y *por qué* y *qué tan grande* podría ser el efecto. Esto nos ayuda a pensar en cómo una intervención diferente o que esté dirigida a diferentes destinatarios podría producir 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 investigación a menudo mostrará 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 contexto.
- ▶ El tratamiento ( $T$ ) y control (la ausencia de  $T$ ) deben estar claramente relacionados a la pregunta de investigación. (Ver el módulo de Medición).
- ▶ Un tratamiento de interés puede ser un conjunto con 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 debe ser igual.



## Un ejemplo

Una campaña de salubridad en la que un voluntario 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 debe incluir todos los demás componentes (visita domiciliaria con una duración de 15 minutos, voluntario con características similares, etc.) pero variar la información que se comparte. Este diseño no nos enseñará el efecto de las visitas, sino solo el efecto de la información en sí.
- ▶ Si su pregunta se centra en el efecto de las visitas, entonces se necesita un grupo de control que no reciba visitas. Sin embargo este diseño no servirá para responder preguntas específicas sobre el contenido de 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 ser que la organización de salud comunitaria con la que están colaborando para visitar hogares no esté interesada en visitar hogares y compartir otra información. En ese caso el control podría ser no visitar.
- ▶ Se debe tener cuidado en interpretar los resultados como el efecto de la información comunicada de esta manera en particular.
- ▶ No se puede concluir que se ha estimado solamente el efecto de la información comunicada.
  - ▶ Puede que esto sea adecuado para ciertas políticas: quizá la pregunta de la política pública sea del efecto de las visitas como parte de una combinación implícita de tratamientos.
  - ▶ Pero es difícil interpretar los resultados de este diseño si queremos saber algo específico sobre el efecto de la información

# El Proceso de investigación

# Una descripción general del proceso de investigación

- ▶ Articular y afinar la pregunta (preguntándonos la razón de hacer esta pregunta y qué sucedería si obtenemos diferentes tipos de respuestas).
- ▶ Desarrollar el diseño de investigación.
- ▶ Planificar el análisis, presentar y justificar hipótesis específicas, así como registrar el plan con un sello impersonal creíble señalando la fecha del registro.
- ▶ Implementar la intervención y recopilar los datos.
- ▶ Analizar los datos y escribir los resultados

# El Formulario para el diseño de investigación de EGAP

# El Formulario para el diseño de investigación de EGAP

- ▶ Creamos un **Formulario para el diseño de investigación** que provee de una estructura para desarrollar un buen diseño de investigación.
  - ▶ <https://egap.github.io/learningdays-resources/Exercises/design-form.Rmd>
- ▶ Les puede ayudar a:
  - ▶ escribir una propuesta de investigación para solicitar financiación para un proyecto, y/o
  - ▶ desarrollar un plan de pre-análisis.

# Secciones del formulario para el diseño de investigación EGAP

- ▶ Pregunta de investigación
- ▶ Muestra
- ▶ Tratamiento
- ▶ variable de resultado
- ▶ Estrategia de asignación aleatoria
- ▶ Implementación
- ▶ Poder estadístico
- ▶ Análisis e interpretación

# Pregunta de investigación y motivación

- ▶ ¿Cuál es la motivación sustantiva de esta investigación? ¿Qué problema se está tratando de abordar? ¿Qué decisión se está tratando de tomar?
- ▶ ¿A quiénes están tratando de cambiar de parecer? y ¿qué creen esas personas actualmente?
- ▶ ¿Qué preguntas teóricas se pueden abordar con esta investigación?
- ▶ Concretar la 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 de investigación 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 es el control? ¿Control puro o placebo?
- ▶ ¿Existe alguna preocupación ética relacionada al tratamiento?
- ▶ ¿A qué nivel se realizará la asignación aleatoria del tratamiento?

# Variable de resultado

- ▶ ¿Cuál es la variable de resultado principal?
- ▶ ¿Cómo se medirá?
- ▶ ¿Qué datos se necesitan? ¿A qué nivel se pueden medir los datos?
- ▶ ¿Qué ideas tienen sobre la variable de resultado previo a la recolección de datos (priors)? Éstas pueden basarse en estudios previos o suposiciones bien fundamentadas.
- ▶ ¿Cuántas rondas de recolección de datos habrá?
- ▶ ¿Cómo se minimizará la deserción?
- ▶ ¿Cómo se minimizará la mala medición y las respuestas deshonestas?

# Estrategia de aleatorización

- ▶ ¿Qué tipo de estrategia de asignación aleatoria utilizarán?  
Ejemplos: simple, completo, bloques, conglomerados, factorial, de dos niveles, por etapas, etc.
- ▶ Asegúrense 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).
- ▶ Definan sus bloques y conglomerados (si los hay). ¿Cuántos van a tener? ¿Qué tan grandes serán?
- ▶ ¿Es posible que haya interferencia? Si es así, ¿cómo les ayudará su muestra y la estrategia de aleatorización a minimizar la interferencia?

# Implementación I

- ▶ ¿Cómo llevará a cabo la asignación aleatoria 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 dará seguimiento a la calidad de la implementación?
- ▶ ¿Cómo dará seguimiento al 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)?

# Poder estadístico

- ▶ ¿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). Tengan 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 investigación al permitirnos representarlas en código, lo que luego nos permitirá simular las etapas del diseño de 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 pueden consultar el módulo sobre estimaciones y estimadores que utiliza DeclareDesign para ayudarles a determinar los estimadores apropiados.

# Introducción a DeclareDesign

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

# Modelo

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

# Indagación (Inquiry)

- ▶ 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 una cantidad que resume los datos, como lo es 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 expresan?

# Respuesta (Answer Strategy)

- ▶ Generar una respuesta dada la recolección 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)

## Pre-registro del plan de análisis



# Sesgo en la investigación publicada contra resultados nulos

- ▶ Anticipándose a las dificultades que hay al publicar, o para no enfrentarse a 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 las posibilidades de publicación.
- ▶ Incluso cuando no se enfrentan estos incentivos se toman muchas decisiones al analizar los 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.

# Pre-registro de planes de análisis y diseños de investigación I

- ▶ El pre-registro es incluir el diseño de investigación e hipótesis en un repositorio de acceso público. EGAP aloja uno que se puede usar de forma gratuita (actualmente en [OSF.io](https://osf.io) utilizando el formulario de registro EGAP).
- ▶ El pre-registro no implica excluir análisis exploratorios posteriores que no fueron registrados con anticipación. Simplemente se debe distinguir claramente entre ambos.

# Pre-registro de planes de análisis y diseños de investigación II

- ▶ Incluso si se va a enviar un artículo con resultados en lugar de un diseño para una revista académica o se está interesado principalmente en el informe final con hallazgos para una audiencia en el campo de las políticas públicas, existen ventajas importantes para ustedes y para otros investigadores del pre-registro de su investigación.
  - ▶ Por ejemplo, se puede obtener información sobre otras investigaciones, completadas y en curso; otros pueden aprender sobre su trabajo. Podemos aprender de estudios que produjeron resultados nulos.
  - ▶ Les 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 de investigación 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 pre-registro 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.