

# Transparencia y Reproducibilidad en Investigación Social

Julio Iturra y Martín Venegas

2021-07-20



# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introducción</b>  | <b>5</b>  |
| <b>2</b> | <b>Transparencia</b>   | <b>9</b>  |
| 2.1      | ¿Qué es la transparencia? Un concepto multidimensional . . . . . | 9         |
| 2.2      | ¿Crisis en las ciencias? . . . . .                               | 12        |
| 2.3      | ¿Que podemos hacer? . . . . .                                    | 18        |
| <b>3</b> | <b>Reproducibilidad</b>  | <b>25</b> |
| 3.1      | Cómo se entiende . . . . .                                       | 25        |
| <b>4</b> | <b>Herramientas para la transparencia y irreproducibilidad</b>   | <b>29</b> |
| 4.1      | Preregistros . . . . .   | 29        |
| 4.2      | Flujos de trabajos reproducibles. . . . .                        | 32        |
| <b>5</b> | <b>Palabras finales</b>  | <b>33</b> |



# Chapter 1

## Introducción

Imagine que usted es un chef, y como tal, disfruta enormemente de las artes culinarias. Como amante de la cocina, cada vez que alguna celebración especial se avecina, es tradición preparar sus mejores recetas para sus invitados. Sin embargo, en esta ocasión, la celebración se llevará a cabo en un restaurante, lo cual implica que la preparación de la comida no depende de usted. Con bastantes dudas, usted accede, teniendo en mente la siguiente pregunta: ¿Cómo puedo tener certeza de que se seguirán los procedimientos adecuados para que la preparación sea de calidad? Pues, estimado lector, es esta la pregunta que subyace a todo el proceso que engloba al conocimiento científico.

La ciencia, al igual que la cocina, no se trata solamente de los productos. Una preparación no aparece por generación espontánea, sino que requiere el seguimiento riguroso de una receta, donde usamos los ingredientes, las cantidades y tiempos de cocción adecuados. Esta receta no solo contribuye a la preparación de la comida, sino que hace posible que dicho plato pueda ser **reproducido** cada vez que se lo desee, garantizando el mismo resultado. Entonces, como buen chef, la respuesta a la pregunta no es tan compleja: necesitamos ser capaces de evaluar el proceso de elaboración de la comida para asegurarnos de que es el adecuado. Dicho de otro modo, el proceso de preparación debe ser **transparente** y estar abierto al escrutinio de otros cocineros expertos, como también para quienes estén adentrándose en la cocina. En el caso de la comida, que el restaurante cuente con una cocina abierta o construida en torno a ventanales bastaría para lograr este objetivo. Sin embargo, dentro del campo de la ciencia ¿cómo logramos que el proceso de producción del conocimiento científico esté abierto al escrutinio público?

Esta pregunta puede ser algo engañosa ¿acaso los procesos de investigación en la ciencia no están ya abiertos al escrutinio público? La narrativa actual pareciese sugerir que no. En el último tiempo ha primado el diagnóstico de que la ciencia está viviendo una crisis, donde polémicas situaciones de malas

prácticas académicas han salido a la luz. Un ejemplo de estas malas prácticas es el falseamiento de datos. Uno de los casos más emblemáticos es el de Diderik Stapel, una figura académica con alto prestigio en el campo de la psicología social a quién se le acusó y confirmó de falseamiento de datos. Más de 10 años de investigación y 150 artículos -algunos de ellos en las revistas más prestigiosas- fueron puestos en duda a raíz de las malas prácticas de Stapel. Así también el trabajo de muchos colegas y estudiantes fue desacreditado. El caso del ex doctor Stapel acabó en la revocación de su doctorado, la retracción de 58 artículos de investigación y, prácticamente, el fin de su carrera académica.

Casos como el de Diderik Stapel existen muchos (ver Abril Ruiz, 2019), sin embargo, no son las prácticas más recurrentes. Dentro de la investigación científica existen una serie de prácticas que caen en un terreno gris cuando se trata de su evaluación ética, estas son las llamadas *prácticas cuestionables de investigación*. La preocupación dentro de la comunidad científica es que la acumulación y la poca fiscalización de estas prácticas lleven a una ciencia poco transparente, con dificultad en torno a la reproducibilidad de los análisis y de sus resultados, y que a la larga se pierda la confianza en el quehacer científico.

La pérdida de la confianza en el quehacer científico afecta el objetivo de las ciencias sociales. Generalmente se tiene la concepción de que es parte de las tareas de los científicos sociales el contribuir al bienestar de la sociedad por la vía de las herramientas de investigación. Esta es la noción de las ciencias sociales como un bien público (Thibodeaux, 2016). Bajo esa idea, los efectos de una crisis de credibilidad tienen dos posibles efectos concretos en las ciencias sociales. Primero, la falta de credibilidad podría afectar la confianza en los hallazgos y en las disciplinas que componen las ciencias sociales. Ya existe evidencia sobre un incremento de desconfianza hacia los científicos en países como Estados Unidos (Motta, 2018) y sumarle una crisis de la ciencia a esa situación solo generaría un ambiente mucho más complejo. Segundo, y estrechamente relacionado a lo anterior, una pérdida de credibilidad en la ciencia podría impactar en la elaboración de políticas sociales, mal orientando las prioridades y los recursos del país. Si se quieren evitar estas situaciones, es necesario tomar cartas en el asunto y orientar los esfuerzos a devolverle la credibilidad a las ciencias sociales.

Este manual busca contribuir a una ciencia social más creíble, basándose en el marco de la ciencia abierta. La ciencia abierta se entiende como:

“...la práctica de la ciencia de manera que otros puedan colaborar y contribuir, donde los datos de la investigación, las notas de laboratorio y otros procesos de investigación estén disponibles libremente, en términos que permitan la reutilización, redistribución y reproducción de la investigación y sus datos y métodos subyacentes”. (FORSTER, Open Science Teaching Resource)

Dentro de la ciencia abierta, nos centraremos en dos conceptos: transparencia y reproducibilidad. Si seguimos la metáfora de la cocina que planteamos en un principio, la transparencia y la reproducibilidad son dos conceptos similares,

pero no idénticos. La transparencia implicaría la posibilidad de evaluar y poner en discusión la receta y la ejecución de la misma. En cambio, la reproducibilidad apuntaría a que la receta sea lo suficientemente clara y precisa para que el mismo plato, con el mismo sabor, pueda ser preparado por cualquier persona que contase con los ingredientes y recursos necesarios. En el campo de las ciencias, esta distinción la hacemos entre: 1) la transparencia de los procesos de producción científica (e.g. transparentar el plan de análisis) y 2) la reproducibilidad de los análisis en los artículos (e.g. código de procesamiento y análisis de datos permite reproducir el artículo). Este manual estará estructurado en torno a estas dos conceptos.

Relacionado al primer concepto (transparencia), haremos un barrido un tanto más detallado sobre la crisis de la reproducibilidad que aquí hemos mencionado brevemente. Ahondaremos en los factores que contribuyen a su reproducción, a las razones éticas por el cuales es necesario hacer un cambio y en las recomendaciones que se pueden seguir para adoptar ciertos principios de transparencia. En el caso del segundo concepto (reproducibilidad), nos centraremos en los análisis reproducibles, teniendo un carácter mucho más práctico. Presentaremos las distintas consideraciones que hay que tener para que un análisis sea fácilmente reproducible, desde tipos de flujos de trabajo hasta herramientas específicas que faciliten la reproducibilidad de los análisis.

Estimado investigador o investigadora de las ciencias sociales, este documento va dirigido a usted. Independiente de su disciplina, de su trayectoria académica o de su conocimiento previo con respecto a estas temáticas, en este manual tenemos dos simples objetivos. El primero es que usted pueda convencerse de que, efectivamente, es necesario dar un giro en la forma que hacemos ciencia actualmente y que la adopción de prácticas relacionadas a la ciencia abierta son el primer paso en ese giro. El segundo es poder instruirlo en esa adopción de prácticas, particularmente en lo que respecta a la transparencia y reproducibilidad. Al final de este manual, usted será capaz tanto de argumentar por qué la transparencia y la reproducibilidad son un paso importante en el avance de las ciencias sociales, así como también contará con una serie de herramientas para llevar esto al quehacer académico del día a día.





## Chapter 2

# Transparencia

Esta sección tratará sobre la transparencia en la investigación científica, haciendo énfasis en las ciencias sociales. Nuestro objetivo es poder comunicar de forma clara y concisa tres puntos: a) qué es la transparencia, b) por qué la necesitamos y c) cómo podemos adoptarla. En esta sección profundizaremos en los dos primeros puntos. Respecto al tercer punto, entregaremos algunos consejos y recomendaciones que se han dado en la literatura, para después, en la siguiente sección, profundizar en torno a las herramientas que nos permitirían adoptar la transparencia. La tónica de este escrito es la práctica, es decir; todo lo que presentemos acá tiene la finalidad de servir de camino para poder aprender y aprehender herramientas que promueven la transparencia. Dicho esto, comencemos con los dos primeros puntos.

### 2.1 ¿Qué es la transparencia? Un concepto multidimensional

Nuestro punto de partida es que la transparencia es un concepto amplio y multidimensional. Por eso, antes de adentrarnos en sus dimensiones, comencemos con una definición de diccionario. Según la Real Academia Española la transparencia es la cualidad de un cuerpo que permite ver a través de él. Un ejemplo para llevar esta definición a la práctica es el vidrio de una ventana. La transparencia del vidrio nos permite ver con claridad lo que está el otro lado, como por ejemplo un paisaje. No obstante, ¿qué ocurre cuándo la transparencia del vidrio se va perdiendo? La respuesta es simple, pero potente: la claridad con la que veíamos el paisaje se va difuminando. Esta pérdida de claridad puede dar como resultado que nuestra observación del paisaje se torne ambigua y errónea, o dicho de otro modo, cada vez será más difícil analizar el paisaje. Desde este ejemplo se podría plantear un símil con la ciencia; el paisaje equivale al proceso de producción científica, y el vidrio representa la claridad con la que podemos

analizar este proceso. De esta manera, la base de la idea de transparencia es que permite analizar con claridad un fenómeno, una situación, o en este caso, un proceso.

¿Qué implica un proceso científico transparente? Ya existen algunas respuestas a esta pregunta. Por ejemplo, Breznau (2021) entiende la transparencia como una forma en que los investigadores pueden revelar el proceso, ideas y materiales que sustentan un argumento o una teoría, con tal de contribuir a una comunidad científica más ética. Otra perspectiva es la de Aczel et al. (2020), quienes proponen a la transparencia como un principio que permite evaluar y reproducir los hallazgos científicos, así como también a sintetizar investigaciones y contribuir a la ejecución de metanálisis. Klein et al. (2018) proponen que la credibilidad de los productos científicos dependen de la transparencia en la que se basan, y que, en consecuencia, avanzar hacia la transparencia simplemente significa adoptar ciertas prácticas de gestión de la investigación que la hagan menos propensa a errores y más reproducibles. Estas perspectivas son una primera aproximación a las implicancias de un proceso científico transparente, no obstante, no representan una presentación exhaustiva del concepto.

Volvamos a nuestro planteamiento inicial: la transparencia es un concepto multidimensional. Para abarcar las múltiples formas que puede adoptar, nos basamos en la taxonomía de Elliott (2020). La taxonomía se estructura en torno a cuatro preguntas: ¿por qué?, ¿quienes?, ¿qué? y ¿cómo? Cada una de estas preguntas tiene al menos una dimensión asociada. La primera pregunta (*¿por qué?*) se refiere a los propósitos por los cuales es necesario adoptar la transparencia; la segunda pregunta (*¿quienes?*) apunta a la audiencia que está recibiendo la información; la tercera pregunta (*¿qué?*) hace alusión al contenido que es transparentado y la cuarta pregunta (*¿cómo?*) consiste en cuatro dimensiones distintas sobre cómo adoptar la transparencia: actores, mecanismos, tiempo y lugar. También, la taxonomía propone una dimensión sobre los peligros que podrían afectar a las iniciativas que busquen promover la transparencia.. Una representación gráfica puede verse en la Figura N° 2.1.

## 2.1. ¿QUÉ ES LA TRANSPARENCIA? UN CONCEPTO MULTIDIMENSIONAL<sup>11</sup>

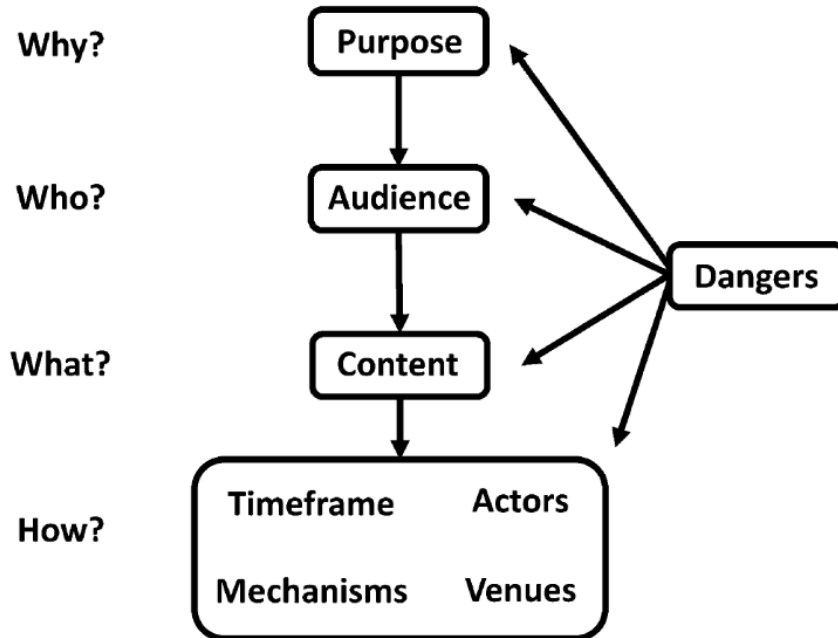


Figure 2.1: Taxonomía de Transparencia

Para comprender mejor la taxonomía la Tabla N° 2.1 presenta una versión detallada de cada dimensión en conjunto a una lista no exhaustiva de variaciones. Por ejemplo, la dimensión del propósito sintetiza varios de los puntos que ya se han señalado en la literatura sobre ciencia abierta: la transparencia contribuye a formar una ciencia más replicable, facilita la interacción crítica, facilita el reanálisis de resultados, entre otros. La mayoría de estos propósitos están estrechamente relacionados con el horizonte de una ciencia con mayor credibilidad, cuestión que profundizaremos en breve. Dentro de las dimensiones restantes, existe una en particular que queremos recalcar: los contenidos. Aquí, los distintos contenidos que pueden ser transparentados van desde cosas complejas como los juicios de valor o factores que podrían influenciar las interpretaciones de resultados, hasta lo más concreto como datos, métodos y materiales, o en otras palabras, el diseño de investigación. Como podemos ver, la transparencia abarca varias dimensiones, sin embargo, a fin de mantener un documento simple y práctico, esta sección apuntará específicamente a la promoción de la transparencia del diseño de investigación.

En síntesis, un proceso de investigación transparente es uno que se puede evaluar con claridad y facilidad. La taxonomía de Elliott (2020) nos ayuda a comprender la diversidad de dimensiones en las que se puede pensar la idea de transparencia.

Ahora conocemos un poco más sobre la transparencia y lo que han dicho algunas voces de la literatura, no obstante, probablemente esté pensando algo como “Sí, tiene sentido que la transparencia contribuya a una ciencia mejor, pero ¿de verdad vale el esfuerzo y el tiempo de adoptar todas estas prácticas? La ciencia parece estar bien como está” Es un pensamiento razonable, sin embargo, la comunidad de adherentes a la ciencia abierta y la literatura asociada han adoptado el discurso de que efectivamente es necesario cambiar la forma que estamos haciendo hoy en día. Ha emergido la narrativa de que la ciencia está en crisis.

## 2.2 ¿Crisis en las ciencias?

En los últimos años, ha venido tomando fuerza la idea de que existe una crisis en la ciencia. Esta idea se basa en el diagnóstico de que gran parte de los artículos científicos en distintas disciplinas no son posibles de reproducir ni replicar (e.g. Wilson et al., 1973; Camerer et al., 2018). Tanto la reproducibilidad (emplear el mismo diseño y datos para reproducir los hallazgos de un artículo) o la replicabilidad (emplear el mismo diseño y distintos datos para obtener los mismos resultados) son componentes centrales de la ciencia. Es más, gran parte del avance del conocimiento científico recae en la verificabilidad de sus hallazgos, por lo que un fallo en poner a prueba los hallazgos es un golpe directo a la credibilidad de la ciencia. La pregunta es ¿existe realmente una crisis en las ciencias?

Unas de las fuentes más comúnmente citadas para introducir la idea de crisis es una encuesta realizada en la revista *Nature*. En esta encuesta, Baker (2016) logró obtener las opiniones de poco más de 1,500 investigadores de disciplinas como la química, ingeniería y la medicina, sobre tópicos relacionados a la reproducibilidad en las ciencias. El resultado principal muestra que un **90% de los encuestados está de acuerdo en la existencia una crisis**, donde un 52% piensa que es una gran crisis y un 38% la percibe como una ligera crisis. En este mismo estudio, se les pregunta por los factores que contribuyen a esta crisis, donde la cultura del *pública o perece* y el *reporte selectivo de resultados* aparecen como los protagonistas. Si bien la encuesta no es una muestra representativa de toda la comunidad científica, presenta una panorámica que lleva a, por lo menos, considerar la crisis de la ciencia como tema que merece atención.

Actualmente, existe un cuerpo de literatura que se ha dedicado diagnosticar y proponer alternativas de solución ante la idea de una ciencia en crisis. Dentro del diagnóstico, variados estudios han orientado sus esfuerzos a esclarecer cuáles son los factores que podrían estar influenciando esta crisis. En línea con los resultados de Baker (2016), es posible plantear dos áreas: una relacionada al modelo de producción científica actual, donde los incentivos por publicar[], los criterios de ascenso en la jerarquía [] académica y el formato de la revisión por pares contribuyen [] a generar una cultura del *pública o perece* que fuerza a los investigadores a aumentar su productividad académica; y otra relacionada a la

flexibilidad (también conocida como grados de libertad) que tienen los investigadores a la hora de realizar investigaciones, generando la oportunidad para prácticas académicas que pueden afectar la credibilidad de los resultados. En esta sección, nos centraremos en la dimensión de las prácticas de investigación.

Comencemos con las prácticas de investigación. Para esquematizar de mejor manera qué es lo problemático de ciertas prácticas, es que utilizamos el esquema conceptual de Steneck (2006). El esquema parte de una distinción básica entre la *ética en investigación* y la *integridad en investigación*, englobando ambas bajo el gran concepto de *Conducta Responsable de Investigación (RCR)* (ver Figura N° 2.2. A grandes rasgos, la RCR se puede entender como el “llevar a cabo la investigación de forma que se cumplan las responsabilidades profesionales de los investigadores, tal y como las definen sus organizaciones profesionales, las instituciones para las que trabajan y, en su caso, el gobierno y el público” (Steneck, 2006). En palabras simples, una conducta integra es atenerse a un conjunto de reglas sobre conducta científica.

\begin{figure}[H]



{

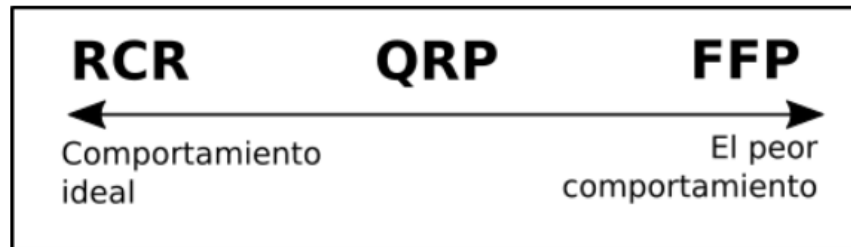
}

\caption{Conducta Responsable de Investigación. Imagen de Abril Ruiz (2019) basada en Steneck (2006)} \end{figure}

Dentro de la RCR, la ética de investigación está relacionada al comportamiento académico visto desde la óptica de los principios morales (Steneck, 2006), lo que se expresa en tópicos como el uso de datos, los consentimientos informados o el trato con pacientes -en el caso de las ciencias biomedicas-, por dar algunos ejemplos. La definición que ofrece Steneck (2006) señala que la ética de investigación se define como “el estudio crítico de los problemas morales asociados o que surgen en el curso de la investigación” (p.56) En cambio, la integridad en investigación se entiende como “poseer y adherirse firmemente a las normas profesionales, tal y como las señalan las organizaciones profesionales, las instituciones de investigación y, en su caso, el gobierno y el público” (Steneck, 2006, p.56). A diferencia de la ética de investigación, el concepto de integridad está regido por los estándares profesionales más que por los principios morales, su función es plantear una

guía clara para la conducta investigativa, de ahí que sea el concepto utilizado por distintos códigos de conducta (ver Abril Ruiz, 2019).

`\begin{figure}[H]`



`{`

`}`

`\caption{Gradación del comportamiento integro en investigación. Imagen de Abril Ruiz (2019) basada en Steneck (2006)}` `\end{figure}`

Habiéndonos situado dentro del concepto de integridad en la investigación, podemos pasar a delinear las principales prácticas que atentan contra él y que se han propuesto como factores que contribuyen a la crisis en la ciencia. Tanto Steneck (2006)

como distintos códigos de conducta de universidades e instituciones de financiamiento (ver Abril Ruiz, 2019) evalúan las prácticas de investigación en un continuo, que representa cuánto adhieren los investigadores a los principios de integridad científica. La Figura N° 2.2 esquematiza esta idea mostrando dos extremos, donde a la izquierda está el mejor comportamiento (RCR), y a la

derecha el peor comportamiento (FFP). Las FFP son una abreviación en lengua inglesa para referirse a *Fabrication, Falsification, Plagiarism* (Invención, Falsificación y Plagio), también conocidas como *mala conducta académica*. En el medio del continuo están las *prácticas cuestionables de investigación* (QRP, por sus siglas en inglés) las cuáles refieren a “acciones que violan los valores tradicionales de la empresa de investigación y que pueden ser perjudiciales para el proceso de investigación” (*National Academies of Science* 1992 en Steneck, 2006, p.58). Las QRP se caracterizan porque tienen el potencial de dañar la ciencia, en tanto las FFP la dañan directamente.

Comprendidos ambos conceptos, veamos las situaciones que podrían categorizarse como FFP y las que podrían concebirse como QRP.

### 2.2.1 Mala conducta académica (FFP)

La mala conducta académica suelen ser situaciones polémicas y que muchas veces alcanzan gran cobertura mediática. El libro de Abril Ruiz (2019) hace una revisión de una serie de situaciones, en distintas disciplinas y años, en las que investigadores han sido descubiertos cometiendo prácticas que atentan directamente a la ciencia. Las situaciones son variadas, existen casos de manipulación de imágenes, exageración de los registros de laboratorio, o de

plano la invención de conjuntos de datos enteros. En esta sección veremos el caso de Diderik Stapel como ejemplo de malas prácticas de investigación y plantearemos la relación que tiene con la transparencia en la investigación.

### 2.2.1.1 Diderik Stapel

Probablemente, el caso de Diderik Stapel sea uno de los más emblemáticos y representativos de este problema. Diderik Stapel era un investigador de la *Tilburg University* que se dedicaba al campo de la psicología social. Su carrera se caracterizó por una trayectoria ejemplar, obtuvo su M.A en psicología y comunicaciones el año 1991, se doctoró en psicología social el año 1997 y trabajó como profesor asociado primero en la *University of Gronigen* (2000-2006) y desde el 2006 en la *Tilburg University*. Fue fundador del *Tilburg Institute for Behavioral Economics Research*, del cual el 2010 se convirtió en decano. Así también, fue galardonado con el premio a la trayectoria académica por la *Society of Experimental Social Psychology*. En breve, Stapel era una figura de alto estatus en el mundo académico, entre 1995 y 2015 publicó aproximadamente 150 artículos en revistas científicas, algunas de las más prestigiosas (e.g. *Science*). Sin embargo, el año 2011 se confirmó que gran parte de su trayectoria académica era una farsa.

Después de la verificación de su culpa ante acusaciones de malas prácticas, la carrera de Diderik Stapel acabó. Fue desvinculado de la *Tilburg University*, se le rebocó su título de doctorado y toda su trayectoria académica fue investigada acusativamente. El informe final de la investigación encontró que más de 50 de sus artículos eran fraudulentos. De hecho, Stapel ocupa el tercer lugar en con 58 artículos listados en *Retraction Watch* (<https://retractionwatch.com/>), una plataforma que se dedica a sistematizar y mantener una lista actualizada sobre retracciones de artículos. A efectos de este escrito, lo que más destacable es que Stapel cometió conductas de mala conducta académica durante más de 15 años. La pregunta es ¿cómo fue esto posible? ¿cómo ninguno de sus colegas, alumnos o co-autores se dio cuenta de sus malas prácticas? La respuesta breve es por la falta de transparencia durante el proceso de investigación.

Los artículos periodísticos que han profundizado en el caso han relatado parte del proceso investigativo de Stapel (e.g. Carey, 2011). Dentro de los procesos de producción y análisis de datos, Stapel se caracterizaba por hacer todo el trabajo solo y “a puertas cerradas”. Es decir, nadie más que él tenía acceso a los datos brutos, ni tampoco a la ejecución de las pruebas estadísticas. Generalmente, Stapel compartía con sus colegas y alumnos de doctorado la base de datos lista, con las pruebas estadísticas ya hechas y, claro está, con resultados estadísticamente significativos. Estas prácticas no causaron sospechas durante muchos años, es más, a muchos de sus estudiantes les parecía una práctica normal y eficiente. Además, con el estatus de Stapel ¿qué podría estar mal? Sin embargo, era a raíz de esta práctica que Stapel tenía la oportunidad de inventar y falsificar datos a su conveniencia. Esto explica en

gran parte de su trayectoria académica llena de grandes hallazgos y una producción académica increíble.

El caso de Stapel deja un punto importante sobre la mesa: la falta de transparencia en el proceso investigativo dio cabida a la mala conducta académica. Cómo nadie más colaboraba con el procesamiento de datos, ni tampoco parecía extraño que así fuera, las oportunidades para la falsificación de los datos estaba abierta. Ahora, esta no es necesariamente una relación de causalidad, la falta de transparencia no tiene porque terminar en conductas como fabricación o falsificación de datos. Sin embargo, tal y como lo argumentan O’Boyle et al. (2017) si son una oportunidad para violaciones a la integridad científica más sutiles, tales como las QRP.

### 2.2.2 Prácticas cuestionables de investigación (QRP)

Recordemos, las QRP son prácticas que en si mismas no dañan directamente la empresa científica, pero si tienen el potencial de hacerlo. Son prácticas que alteran el correcto funcionamiento del método científico. En la literatura, existen una variedad de términos que se utilizan para describir las prácticas cuestionables, así como también distintas listas de prácticas que han emitido instituciones. Abril Ruiz (2019) hace una recopilación y traducción de distintos códigos de conducta de distintas universidades y organismos, el cuál nosotros sistematizamos en la Tabla N° 2.2. La tabla muestra un conjunto de prácticas divididos en cuatro categorías, de acuerdo a si las prácticas tienen que ver con: el diseño y el procesamiento de los datos, la redacción y el reporte de los resultados, temas de citación y uso de ideas ajenas y, por último, sobre relaciones con otros actores en el campo de la ciencia.

Existe una serie de estudios que han intentado medir directamente la existencia de estas prácticas a través de encuestas. Fanelli (2009) hizo un metanálisis que tenía por objetivo sistematizar los resultados de estudios que hasta esa fecha habían abordado las prácticas de investigación desde encuestas cuantitativas.

Los resultados mostraron que un 1.97% de investigadores había inventado datos al menos una vez (FFP) y que un 33.7% había realizado alguna vez una QRP como “borrar puntos de los datos basados en un sentimiento visceral”.

Un estudio más reciente, también basado en encuestas cuantitativas sobre prácticas, es el de John et al. (2012). En este estudio, los resultados mostraron que un 36.6% de quienes participaron alguna vez habían practicado alguna QRP. En detalle, analizando los porcentajes práctica a práctica se halló que el 50% de los psicólogos encuestados alguna vez reportaron selectivamente estudios que apoyaran su hipótesis; un 35% alguna vez reportaron resultados inesperados como esperados; y un 2% alguna vez reportó datos falsos. Estos estudios son una primera aproximación a la existencia de las QRP en la ciencia.

Existen ciertas prácticas que han sido tratadas con más énfasis en la literatura, y que son las que queremos destacar acá. Por un lado, está el *sesgo de publicación*, el cual significa, a grandes rasgos, que un artículo es publicado en



base a sus resultados. Específicamente, el sesgo de publicación ocurre cuando el criterio determinante para que un artículo sea publicado es que sus resultados sean significativos, en desmedro de la publicación de resultados no significativos. El estudio de Franco et al. (2014) logra cuantificar esta situación bastante bien y esclarecer el *file drawer problem* (en español: problema del cajón de archivos), que hace alusión a los resultados perdidos dentro de un cuerpo de evidencia (Christensen et al., 2019, p.39). Franco et al. (2014) encontraron que, efectivamente, los resultados nulos tienen un 40% menos de probabilidades de ser publicados en revistas científicas, en comparación a estudios con resultados significativos. Es más, muchas veces los resultados nulos simplemente llegan a ser escritos: más de un 60% de los experimentos que componen la muestra del estudio de Franco et al. (2014) nunca llegaron a ser escritos, en contraste al menos del 10% de resultados significativos.

El principal problema del sesgo de publicación es que puede impactar en la credibilidad de cuerpos enteros de literatura. Existen ejemplos de cuerpos de literatura que se han hallado sesgados, como el *Value of Statistical Life* (Valor de la Vida Estadística) y el de salario mínimo y desempleo, ambos en economía (para detalle ver p.42 y p.46 en Christensen et al., 2019). También, a partir del desarrollo de distintos métodos de detección, se ha podido diagnosticar el sesgo de publicación en importantes revistas en economía Brodeur et al. (2016), sociología y ciencias políticas (Gerber and Malhotra, 2008b,a).

Otra práctica bastante cuestionada es el *p-hacking*, que de forma literal sería un “hackeo” de los valores  $p$ . El *p-hacking* se da cuando el procesamiento de los datos tiene por objetivo obtener resultados significativos. Si el sesgo de publicación afecta la credibilidad de un cuerpo de literatura, el *p-hacking* afecta a la credibilidad de los artículos mismos, ya que al forzar la significancia estadística la probabilidad de que en realidad estemos frente a un falso positivo aumenta. Un trabajo que da sustento a esta idea es el de Simmons et al. (2011), quienes calculan la posibilidad de obtener un falso positivo (error Tipo I) de acuerdo a al nivel de manipulación intencionada de los datos. El resultado principal es que a medida que aumenta la cantidad de manipulación en los datos, la posibilidad de obtener un falso positivo aumenta progresivamente.

El *p-hacking* también contribuye a sesgar cuerpos enteros de literatura. Para diagnosticar esto se ha utilizado una herramienta denominada *p-curve*, la cual “describe la densidad de los *p-values* reportados en una literatura, aprovechando el hecho de que si la hipótesis nula no se rechaza (es decir, sin efecto), los *p-values* deben distribuirse uniformemente entre 0 y 1” (Christensen et al., 2019, p.67.). De esta manera, en cuerpos de literatura que no sufran de *p-hacking*, la distribución de *p-values* debería estar cargada a la izquierda (siendo precisos, asimétrica a la derecha), en cambio, si existe sesgo por *p-hacking* la distribución de *p-values* estaría cargada a la derecha (asimetría a la izquierda). Simonsohn et al. (2014) proponen esta herramienta y la prueban en dos muestras de artículos de la *Journal of Personality and*

*Social Psychology (JPSP)*. Las pruebas estadísticas consistieron en confirmar que la primera muestra de artículos (que presentaban signos de p-hacking) estaba sesgada, en cambio la segunda muestra (sin indicios de p-hacking), no lo estaba. Los resultados corroboraron las hipótesis, en detalle, los artículos que presentaban solamente resultados con covariables, resultaron tener una p-curve cargada a la derecha (asimétrica a la izquierda).

Por último, pero no menos importante existe la práctica del *HARKing*. El nombre es una nomenclatura en lengua inglesa: *Hypothesizing After the Results are Known*, que literalmente significa establecer las hipótesis del estudio una vez que se conocen los resultados. El principal problema de esta práctica es que confunde los dos tipos de razonamiento que están a la base de la ciencia: el exploratorio y el confirmatorio. El objetivo principal del razonamiento exploratorio es plantear hipótesis a partir de un análisis de datos. En cambio, el razonamiento confirmatorio busca plantear hipótesis basado en teoría y contrastar esas hipótesis con datos empíricos. Como señala Nosek et al. (2018), cuando se confunden ambos tipos de análisis y se hace pasar un razonamiento exploratorio como confirmatorio se está cometiendo un sesgo inherente, ya que se está generando un razonamiento circular: se plantean hipótesis a partir del comportamiento de los datos y se confirman las hipótesis con esos mismos datos.

## 2.3 ¿Que podemos hacer?

Tanto en las ciencias sociales, como en otras disciplinas han ido emergiendo una variedad de recomendaciones que contribuyen a la adopción de la transparencia. Por ejemplo, Crüwell et al. (2018) propone formar investigadores y estudiantes a partir de la promoción de siete principales tópicos: entender la ciencia abierta; acceso abierto; la importancia de los datos, material y código abierto, los análisis reproducibles; los pre-registros; la replicación y, por último, la enseñanza de la ciencia abierta. Cada uno de estos tópicos son un paso para lograr avanzar hacia una ciencia más abierta y transparente. De forma similar, Miguel et al. (2014) enfatiza en tres ideas: el *disclosure* (divulgación), los pre-registros y los datos y materiales abiertos. La divulgación consta de que los investigadores deben declarar todo tipo de procesamiento realizada a los datos y detallar como se llegó a las muestras finales. Los preregistros son una forma de ser más transparente con los procedimientos y dividir los tipos de formulación de hipótesis. Los datos y materiales abiertos permite que otros investigadores puedan reproducir el trabajo, hacer otras pruebas de hipótesis, identificar malas prácticas etc. También Lindsay (2020) discute los efectos que tienen el sesgo de publicación y otras prácticas similares en la credibilidad de la ciencia, proponiendo siete pasos para mejorar la transparencia: decir la verdad, por ejemplo, si la idea de investigación surgió analizando datos plantearlo de esa manera; evaluar la comprensión de la estadística inferencial; estandarizar la aproximación para

probar las hipótesis, por ejemplo, realizar un plan de análisis; hacer un manual de laboratorio (en el caso de diseños experimentales); hacer abiertos los materiales, datos y análisis; abordar las limitaciones de la generabilidad de las conclusiones; y por último, considerar enfoques colaborativos para conducir investigaciones. O'Boyle et al. (2017) proponen tres simples consejos para hacer la ciencia más transparente: incluir una cláusula ética en el envío de manuscritos que declaren no haber cometido QRP, que todo artículo original (e.g. tesis doctoral) debe estar disponible para descarga y que las revistas cuenten con un espacio dedicado a replicación.

Una propuesta sistemática para adoptar la transparencia, que incluye varias de las recomendaciones mencionadas son las *Transparency and Openness Promotion (TOP) Guidelines* (Guías para la Promoción de la Transparencia y la Accesibilidad). Estos son principios que buscan alcanzar un formato de investigación reproducible a través del aumento de la transparencia en el proceso y los productos de investigación (Nosek et al., 2014). Estos principios sirven de guía tanto para la adopción de buenas prácticas en los investigadores, como para que las revistas académicas puedan adherir progresivamente al ideal de transparencia en la ciencia. Son ocho principios:

1. Citación
2. Transparencia de datos
3. Transparencia de métodos analíticos (código)
4. Transparencia de los materiales
5. Transparencia del diseño y el análisis
6. Preregistro de estudios
7. Pre registro de planes de análisis
8. Replicación

A grandes rasgos, el principio de citación propone que las normas de citado deben ampliarse también a los datos y códigos, permitiendo reconocer su autoría intelectual (Nosek et al., 2015). Los principios de transparencia de datos, métodos analíticos, materiales y diseño y análisis (2 a 5) refieren a la transparencia en su forma más concreta: la apertura del proceso de investigación para su evaluación. El detalle puesto a los principios responde a la generalizabilidad que se le busca dar a los principios. Por ejemplo, un estudio observacional cuantitativo no tiene material que transparentar, pero si datos y métodos analíticos. Así también, un estudio cualitativo quizás no tenga código que transparentar, pero si un diseño y una bitacora detalla del proceso de análisis. En el caso de los principios relacionados al pre-registro, Nosek et al. (2015) argumenta que registrar los estudios los hace más descubribles, incluso si no son publicados. Así también, los preregistros del plan de análisis contribuyen a distinguir entre los análisis confirmatorios y exploratorios (ver Nosek et al. (2018) para un manejo detallado del tema). Por último, el principio de replicación fomenta las oportunidades para la corrección de artículos y redirecciona la investigación en vías más prometedoras (Nosek et al., 2015).

Cada uno de estos principios cuenta con tres niveles, que sirven para medir el grado de inclusión de la transparencia por parte de una revista científica (ver Figura N° 2.2 para detalle). La adopción de prácticas transparentes va desde el nivel 1, siendo lo menos transparente, hasta el nivel 3 siendo lo más transparente. Se añade un nivel 0 que no cumple los estándares de transparencia con la finalidad de tener una comparación. Por ejemplo, para los estándares de transparencia del método de análisis (código), el nivel 1 dicta que las revistas deben solicitar la existencia del código de análisis, en cambio, el nivel 3 es más estricto en plantear que el código de análisis debe estar almacenado en un repositorio confiable y que el análisis será reproducido durante el proceso de revisión. El mismo método se puede aplicar para la el preregistro del plan de análisis. En el nivel 1 las revistas promueven el uso de preregistros, en cambio, en el nivel 3 los preregistros son obligatorios y también reconocidos. En suma, las TOP Guideliness son una iniciativa que contribuye a la apertura en la ciencia proponiendo nuevas prácticas para los requerimientos de las revistas.

| Summary of the eight standards and three levels of the TOP guidelines  |   |   |   |  |
|--|---|---|---|--|
| Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard. |   |   |   |  |
|  | LEVEL 0   | LEVEL 1   | LEVEL 2   | LEVEL 3  |
| <b>Citation standards</b>  | Journal encourages citation of data, code, and materials—or says nothing. | Journal describes citation of data in guidelines to authors with clear rules and examples.                  | Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.                             | Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.     |
| <b>Data transparency</b>   | Journal encourages data sharing—or says nothing.                          | Article states whether data are available and, if so, where to access them.                                 | Data must be posted to a trusted repository. Exceptions must be identified at article submission.   | Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.              |
| <b>Analytic methods (code) transparency</b>  | Journal encourages code sharing—or says nothing.                          | Article states whether code is available and, if so, where to access them.                                  | Code must be posted to a trusted repository. Exceptions must be identified at article submission.   | Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.              |
| <b>Research materials transparency</b>   | Journal encourages materials sharing—or says nothing.                     | Article states whether materials are available and, if so, where to access them.                            | Materials must be posted to a trusted repository. Exceptions must be identified at article submission.                                      | Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.         |
| <b>Design and analysis transparency</b>  | Journal encourages design and analysis transparency or says nothing.      | Journal articulates design transparency standards.  | Journal requires adherence to design transparency standards for review and publication.   | Journal requires and enforces adherence to design transparency standards for review and publication.                                 |
| <b>Preregistration of studies</b>  | Journal says nothing.   | Journal encourages preregistration of studies and provides link in article to preregistration if it exists. | Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements. | Journal requires preregistration of studies and provides link and badge in article to meeting requirements.                          |
| <b>Preregistration of analysis plans</b>   | Journal says nothing.   | Journal encourages preanalysis plans and provides link in article to registered analysis plan if it exists. | Journal encourages preanalysis plans and provides link in article and certification of meeting registered analysis plan badge requirements. | Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.      |
| <b>Replication</b>   | Journal discourages submission of replication studies—or says nothing.    | Journal encourages submission of replication studies.   | Journal encourages submission of replication studies and conducts blind review of results.  | Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes. |

Figure 2.2: Variaciones por dimensión de transparencia

Como podemos ver, en general las recomendaciones sobre transparencia giran en torno a los preregistros, los datos y materiales abiertos y la replicación. En la siguiente sección profundizaremos en los preregistros como una forma de aumentar la transparencia durante el proceso de investigación.

Table 2.1: Variaciones por cada dimensión de transparencia

| Dimensión | Variaciones  |
|-----------|--|
| Proposito | Facilitar el reanálisis de los resultados  |
|           | Hacer a la ciencia más replicable  |
|           | Promover la innovación   |
|           | Mantener la rendición de cuentas de los expertos                                   |
|           | Facilitar la interacción crítica   |
|           | Promover el desarrollo de política pública de alta calidad                         |
|           | Permitir al público tomar decisiones de acuerdo a sus valores                      |
|           | Promover la integridad   |
| Contenido | Datos, métodos, código y materiales  |
|           | Interpretaciones de los datos, métodos, y códigos para no especialistas            |
|           | Juicios de valor, factores que los influyen o sus implicancias                     |
|           | Deliberaciones que subyacen a los reportes   |
| Actores   | Científicos que desarrollaron la investigación                                     |
|           | Otros científicos de la misma disciplina u otras                                   |
|           | Periodistas  |
|           | Sociedades científicas   |
|           | Agencias gubernamentales   |
|           | Organizaciones no gubernamentales y de la sociedad civil                           |
| Espacios  | Comunicación desde los científicos (oral o escrita, incluyendo las redes sociales) |
|           | Registros o repositorios   |
|           | Divulgación científica   |
|           | Reportes de agencias gubernamentales   |
|           | Reportes de agencias no gubernamentales o grupos de la comunidad                   |
|           | Científicos haciendo la investigación  |
|           | Otros científicos  |
|           | Periodistas  |

Table 2.2: Algunas situaciones de QRP

| Dimensión                        | Práctica   |
|----------------------------------|--|
| Diseño y procesamiento           | Ignorar ciertos aspectos de los requerimientos de las personas participantes.  |
|                                  | Pasar por alto el uso de datos cuestionables o de interpretaciones cuestionables que otros hacen.                            |
|                                  | Cambiar partes de un estudio como respuesta a la presión de una fuente de financiación                                       |
|                                  | Eliminar observaciones de los análisis basados en la intuición de que eran inexactos.  |
|                                  | Redondear un valor p (por ejemplo, reportar que un p-value de 0,054 es menor a 0,05)   |
|                                  | Eliminación, adición o alteración de datos después de pruebas de hipótesis.  |
|                                  | Supresión selectiva o adición de variables.  |
| Redacción, reporte y publicación | Invertir la dirección o reformular hipótesis para respaldar los datos  |
|                                  | Ampliar de manera innecesaria la bibliografía de un estudio.   |
|                                  | Tergiversar los logros de la investigación.  |
|                                  | Exagerar la importancia y la relevancia práctica de los resultados.  |
|                                  | Retener detalles de la metodología de investigación (e.g. no reportar todas las variables dependientes de un estudio)        |
|                                  | Retener resultados de la investigación (e.g. no presentar datos que contradicen una propia investigación previa).            |
|                                  | Establecer publicaciones o brindar apoyo a publicaciones que no cumplen el proceso de control de calidad de la investigación |
|                                  | Publicar los mismos datos o resultados en dos o más publicaciones.   |
|                                  | Selectivamente reportar estudios   |





## Chapter 3

# Reproducibilidad

¿Cuántas veces nos hemos enfrentado a un trabajo publicado que no comparte sus materiales, y que por tanto, es imposible acceder a los procedimientos que dieron luces de sus resultados? En el marco de la denominada “crisis de la ciencia”, la comunidad científica se ha organizado para dar salida a este problema a través de una diversidad de iniciativas que, a modo general, buscan promover los principios de la ciencia abierta. En este sentido, dichas iniciativas han puesto sus esfuerzos en contribuir con herramientas que permitan dar salida a los problemas en torno a la transparencia de los procesos de investigación y la **reproducibilidad** en la investigación empírica. En este apartado, revisaremos cómo ha sido entendido este problema y luego se presentarán las propuestas de tres iniciativas en torno a cómo abordar la reproducibilidad en las ciencias sociales empíricas.

### 3.1 Cómo se entiende

En la discusión sobre los problemas de transparencia en torno a los procesos y procedimientos de investigación, se vuelve necesario precisar de qué manera es entendido el concepto de reproducibilidad en la ciencia. En esta línea, la laxitud en que se ha empleado el término ha llevado a definiciones poco claras, lo cual ha generado una tendencia a confundir lo que refiere a la transparencia de un proceso único que ya ha sido realizado, con un proceso nuevo y que puede realizarse de manera reiterativa, obteniendo los mismos resultados.

La discusión en torno a cómo se entiende la **reproducibilidad**, habitualmente lleva al contraste respecto al concepto de **replicabilidad**. Al respecto on Earth and on Behavioral (2019) menciona que con el incremento de las herramientas computacionales a principios de los años 90’, el término de “investigación reproducible” era concebido como las investigaciones que

proveían un compendio detallado de la documentación, código y datos que permitieran obtener los mismos resultados publicados por los autores, enfatizando que los análisis fueran transparentes y claros con el objetivo de ser verificados por sus pares. Por otro lado, los autores sostienen que en otras disciplinas, el concepto de reproducibilidad era asociado a investigaciones independientes entre sí en términos de los datos empleados, los materiales e implementación, lo cual estaría orientado a robustecer o cuestionar la evidencia previa (on Earth and on Behavioral, 2019, pp 33-34). Actualmente, a esta práctica se la entiende como “replicabilidad” de una investigación y no debe ser confundida con el concepto de “reproducibilidad” (Barba, 2018).

Barba (2018) sugiere que la confusión entre reproducibilidad y replicabilidad ha contribuido a obstaculizar las prácticas en ambas dimensiones. En una revisión reciente realizada por la autora se han identificado al menos tres escenarios o versiones de cómo se entienden ambos conceptos en una amplia gama de disciplinas que van desde las ciencias sociales hasta estudios clínicos. El primer escenario (A), y a la vez el más común, es donde el uso de ambos conceptos es indistinto, contribuyendo a la ya mencionada confusión. El segundo escenario (B1) es cuando la reproducibilidad es entendida como la situación que los datos originales y el código de análisis son empleados para **regenerar** los resultados originales, mientras que la replicabilidad es entendida cuando investigadores o equipos independientes utilizan datos nuevos para obtener los mismos resultados que la investigación previa. Finalmente, un tercer escenario (B2) es cuando la reproducibilidad es entendida cuando investigadores o equipos independientes obtienen los mismos resultados empleando sus propios datos y métodos, mientras que la replicabilidad es entendida cuando investigadores o equipos independientes llegan a los mismos resultados empleando los “artefactos digitales”<sup>1</sup> originales del autor con menores o mayores modificaciones, los cuales han sido puestos previamente a disposición de sus pares. La Figura 3.1 ilustra cómo podemos entender el escenario B1 y B2 en relación a la distinción entre reproducibilidad y replicabilidad.

---

<sup>1</sup>Barba (2018) lo define como un compendio que detallar la estrategia de medición, diseño del estudio o código de análisis originales de un autor

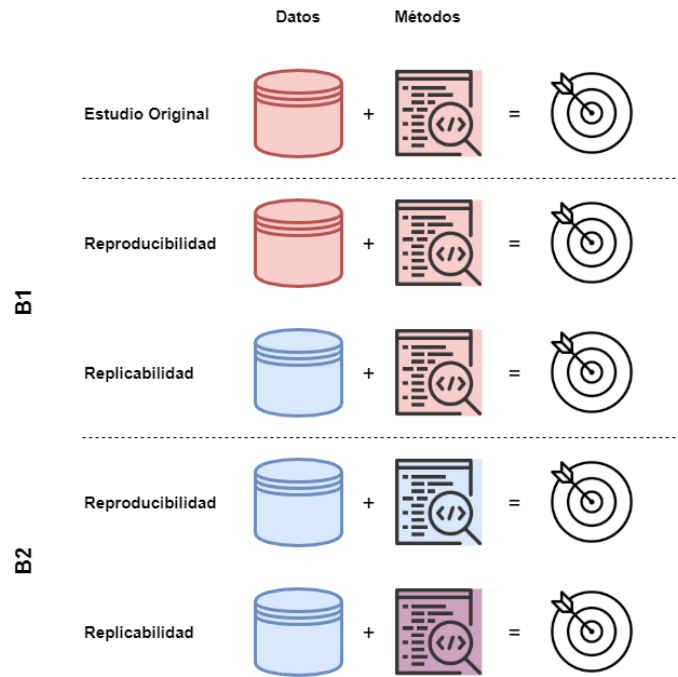


Figure 3.1: Escenarios B1 y B2 en reproducibilidad y replicabilidad.

En las ciencias sociales, el debate en torno a la investigación reproducible y la replicabilidad no ha estado ausente. Como fue reseñado en la sección anterior, existen casos icónicos en torno a prácticas cuestionables de investigación que han afectado la confianza en la investigación científica, lo cual ha contribuido a incrementar los esfuerzos por una ciencia social abierta y reproducible (Breznau, 2021; Nosek et al., 2015). En los tres escenarios descritos por Barba (2018), las ciencias sociales han experimentado de manera diversa el ajuste hacia una cultura de mayor apertura y precisión en torno a los problemas de la crisis de reproducibilidad, principalmente a través del estudio sistemático de dicha problemática, dentro de lo cual la psicología ha sido un pionera en proveer evidencia para este debate (Open Science Collaboration, 2015; Gilbert et al., 2016). Al respecto Bishop (2019) sostiene que una de las principales amenazas para el progreso de la ciencia en general ha sido a la falta de reproducibilidad de los resultados (*irreproducibility*), lo cual ha afectado principalmente la robustez y credibilidad de la evidencia reportada por las investigaciones, problema que también ha sido identificado en las ciencias sociales, principalmente por la falta de precisión en los procedimientos y las barreras de acceso a materiales clave del proceso de análisis (Freese and Peterson, 2017).

Entonces, retomando la distinción clave entre lo que entendemos por

**reproducibilidad y replicabilidad**, en su revisión, Barba (2018) sugiere que una manera de entender y distinguir ambos conceptos de manera minimalista puede descansar en el carácter de los *datos* y los *métodos*. Al respecto Nosek et al. (2015) sostiene que en lo que refiere a estas dos dimensiones, los niveles en que una publicación los incorpora es gradual y puede entenderse como un continuo o espectro (Peng, 2011), y por tanto, el nivel en que se cumplen con determinados criterios nos permite definir el carácter de una investigación en términos de su reproducibilidad. Por ejemplo, la Figura XX nos muestra cómo podemos caracterizar una investigación publicada en torno al acceso y vinculación entre código y datos. Por un lado, se observa que en el polo donde únicamente disponemos de la publicación, se entiende como la ausencia de reproducibilidad. Por otro lado, en la medida que incrementa el acceso a los materiales, y se explicita el enlace entre ellos, se puede caracterizar a una publicación como reproducible.<sup>2</sup>

Figura XX: Espectro de Reproducibilidad.

---

<sup>2</sup>En la figura original, Peng (2011) muestra el polo derecho como el mejor escenario y lo clasifica como *Full replication*, sugiriendo que el mejor estándar para poner a prueba los hallazgos de una investigación científica es la replicación, pero en la ausencia de dicha posibilidad la reproducibilidad de los resultados debiese ser un estándar mínimo

## Chapter 4

# Herramientas para la transparencia y irreproducibilidad

En la sección anterior hemos hecho un repaso sobre la narrativa

### 4.1 Preregistros

Los preregistros son una marca temporal sobre las decisiones del diseño, el método y el análisis de un artículo científico y se suelen hacer antes del levantamiento de datos (Stewart et al., 2020). Básicamente, preregistrar un artículo o un proyecto implica que un grupo de investigadores dejarán por escrito una pauta de investigación a la cuál se atenderán lo más posible cuando desarrollen la investigación, especialmente la recopilación y el análisis de los datos. Ahora ¿por qué habría de hacer algo así? Llevar a cabo una investigación ya es lo suficientemente complejo como para añadirle una tarea adicional. La respuesta es que, cómo señalamos en secciones anteriores, los preregistros son una herramienta que permite hacerle frente a las QRP y, a la larga, contribuir a la realización de una ciencia de mayor calidad.

Ciertamente, los preregistros no son la respuesta a cada una de las QRP existentes, pero sí son una herramienta eficaz para evitar las más frecuentes. En secciones anteriores hablamos de los sesgos de publicación, el p-hacking y el HARKing pues, cada uno de ellos puede ser evitado a partir de un preregistro.

Primero, vimos que el sesgo de publicación se trataba de publicar selectivamente los resultados de investigación: resultados que no hayan sido significativos, o hipótesis que “no funcionaron” simplemente se omiten. Sin embargo, cuando existe un documento como un preregistro, el cual deja

estipulado claramente las hipótesis que deben ponerse a prueba y los análisis que se emplearan para ello, se torna un poco difícil el reportar selectivamente

los resultados. Dicho de otra forma, cuando existe una pauta a la cual apegarse, la discrecionalidad en el reporte de los resultados disminuye. En el caso del p-hacking, el efecto del preregistro es parecido. Como vimos, el p-hacking consistía en abusar de las pruebas estadísticas para obtener resultados significativos. “Abusar” en el sentido de buscar toda vía posible para obtener un valor  $p$  que confirme las hipótesis planteadas. El hecho de preregistrar el plan de análisis y el procesamiento que se le efectuara a las variables permite evitar este tipo de búsqueda intencionada: como hay una guía que seguir, cualquier desviación debe ser justificada. En esta misma línea, un preregistro evita el HARKing ya que las hipótesis están previamente planteadas y no es posible cambiarlas una vez que se han visto los resultados. En suma, el plantear un registro *a priori* de la investigación, disminuye la flexibilidad que suele dar paso a las QRP.

Si bien los preregistros pueden ser una herramienta en contra de las QRP, existen resquemores de los que es preciso hacerse cargo. Una de las principales preocupaciones es que el uso de preregistros tendería a coartar la creatividad y la producción de conocimiento exploratoria []. La lógica es que, como cada parte de la investigación debe ser registrada detalladamente previo a la recopilación, no queda espacio para la espontaneidad durante el análisis el análisis de datos. Nada puede estar más lejos del sentido de un preregistro. Más que inhibir la investigación exploratoria, el objetivo de especificar una pauta *a priori* es poder separar lo que es la investigación confirmatoria (pruebas de hipótesis) y la exploratoria (generación de hipótesis) (Nosek et al., 2018). En ese sentido, es posible la investigación exploratoria bajo el modelo de preregistros, solo que hay que especificarla como tal. Una segunda creencia es que realizar un preregistro añade un nivel de escrutinio mayor del necesario, es decir, como se conoce cada detalle, la investigación se vuelve un blanco fácil de críticas. Sin embargo, la situación es todo lo contrario (Moore, 2016), por ejemplo, haber preregistrado un plan de análisis para una regresión logística binaria con datos que originalmente eran ordinales hará más creíble los resultados, ya que quienes evalúen la investigación tendrán pruebas de que el nivel de medición no se cambió solo para obtener resultados significativos. Una tercera idea en torno a los preregistros es que conllevan una gran inversión de tiempo y energía. Si bien es cierto que se añade un paso más al proceso de investigación, el avance en la temática ha logrado que existan una variedad de plantillas que hacen el proceso más rápido y eficiente. Desde una lógica racional, el tiempo que toma este paso nuevo en la investigación es un costo bajo en contraste a los beneficios que trae.

Una característica principal de los preregistros es que deben ser efectuados previo a la recolección de datos. Este requisito es lo que permite asegurar la credibilidad de los resultados, ya que, si no hay datos que alterar, entonces las probabilidades de que ocurra una QRP son básicamente nulas. Generalmente, para las ciencias médicas o la psicología experimental (disciplinas donde cada

vez se usan más los preregistros), esto no suele ser un problema ya que se utilizan diseños experimentales. Los diseños experimentales se apegan al método científico clásico: se plantean hipótesis basadas en la teoría, se diseña un experimento para probar esas hipótesis y luego se recopilan y analizan los datos para ver si dan soporte a las hipótesis planteadas. Sin embargo, ¿qué ocurre cuando trabajamos con datos pre-existentes? En muchas disciplinas de las ciencias sociales los diseños experimentales son una pequeña fracción del conjunto de la literatura (e.g. según Card et al., 2011, en 2010, un 3% de los papers en las mejores revistas de economía eran experimentales), donde lo que prima son los diseños observacionales, los que suelen trabajar con datos administrativos o generados a partir de censos o encuestas. A diferencia de muchos estudios experimentales, los cuales deben generar de primera fuente el experimento que les permita testear sus hipótesis, en los estudios observacionales se utilizan datos ya existentes, lo cual afecta al principal componente de credibilidad de los preregistros: nada puede asegurar que los datos fueron analizados antes de la escritura del preregistro y que, por ejemplo, las hipótesis se están planteando una vez conocidos los patrones significativos (HARKing). De ahí que nace la pregunta sobre la posibilidad (y la utilidad) de utilizar preregistros en estudios con datos pre-existentes.

En la literatura sobre preregistros se han discutido los desafíos que implica preregistrar estudios que utilicen datos pre-existentes [e.g. editors\_Observational\_2014]. Existen posturas que proponen que, en realidad, no existe una forma creíble para preregistrar este tipo de estudios (Christensen and Miguel, 2018). No obstante, otras posturas han profundizado en las situaciones en las que aun es posible pre-registrar estudios con datos elaborados previamente. Burlig (2018) proponen tres escenarios donde el preregistro de datos observacionales es valioso. El primero es, básicamente, cuando el grupo que planteo el diseño de investigación genera sus propios datos, tal y como sucede en investigaciones que cuentan con los recursos para poder aplicar instrumentos originales. En este caso, los investigadores sí pueden elaborar un preregistro previo a la recolección de datos. El segundo escenario se da cuando se preregistra un estudio que tenga como objeto de interés un suceso que aun no ha ocurrido, lo que se conoce como estudios prospectivos. Por ejemplo, un grupo de investigadores puede estar interesado en el efecto que tendrá la introducción de una ley en las prácticas sociales, o el efecto de un tratado en las variaciones del PIB. Para esos casos, el preregistro aun mantiene su validez original ya que, si bien los datos ya existen, no es posible hacer los análisis antes del preregistro porque el evento de interés no ha ocurrido. El tercer escenario ocurre cuando los datos existen, pero no están abiertos al público. En estos casos, es la accesibilidad lo que determina la credibilidad del preregistro. Por ejemplo, el grupo de investigadores que elaboraron los datos pueden establecer que serán accesibles con previo contacto y que se solicitará un preregistro. Por ende, en orden de analizar los datos, los investigadores interesados deberán elaborar un preregistro para utilizar los datos. Según Mertens and Krypotos (2019), también se pueden

adoptar ciertas prácticas para asegurar la credibilidad de un preregistro con datos secundarios. Dos de ellas son: que el grupo de investigadores que analiza los datos sea distinto e independiente de quien propuso el diseño de investigación y que el equipo realice sintaxis de análisis con datos simulados, con tal de demostrar que las hipótesis ya existían previo a acceder a los datos. En suma, lo que permite mantener el efecto “puro” de un preregistro, es que los datos no hayan sido observados por ningún integrante del grupo de investigación que los analizará (Nosek et al., 2018).

El principio básico de un preregistro entonces, es que los datos no deben haber sido observados previos al análisis. Sin embargo, según Nosek et al. (2018) aun pueden existir ciertos sesgos en el planteamiento de hipótesis a raíz de cosas como el reporte de resultados descriptivos de la base de datos o las recomendaciones sobre cómo aproximarse a la base de datos. Este tipo de influencias son un poco más sutiles y es difícil deshacerse completamente de ellas. Es por esto que, quizás la recomendación más transversal y a la vez simple para registrar análisis con datos secundarios, es ser sincero y detallado respecto a lo que se ha hecho y lo que no (Nosek et al., 2018). Si es que se ha leído el reporte descriptivo sobre la base de datos, estipularlo como tal. Es preciso transparentar cualquier tipo de aproximación a los datos previo haberlos analizados. Para lograr este nivel de detalle y ser eficiente con los tiempos y la comunicación hacia otros investigadores, es que existen plantillas predeterminadas para preregistrar distintos tipos de artículos. A continuación veremos las más comunes y nos centraremos en describir una de las más usadas y su variante para estudios que usen datos secundarios.

- ¿Cómo se hacen los preregistros? ¿Qué plantillas hay? ¿Qué hay que incluir?

#### 4.1.1 Otros

- Otras cosas que contribuyen a los preregistros: reporting standards y registered reports

## 4.2 Flujos de trabajos reproducibles.



## Chapter 5

# Palabras finales

Hemos aprendido sobre transparencia y reproducibilidad.



# Bibliography

- Abril Ruiz, A. (2019). *Manzanas podridas: Malas prácticas de investigación y ciencia descuidada*.
- Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š., Benjamin, D., Chambers, C. D., Fisher, A., Gelman, A., Gernsbacher, M. A., Ioannidis, J. P., Johnson, E., Jonas, K., Kousta, S., Lilienfeld, S. O., Lindsay, D. S., Morey, C. C., Munafò, M., Newell, B. R., Pashler, H., Shanks, D. R., Simons, D. J., Wicherts, J. M., Albarracin, D., Anderson, N. D., Antonakis, J., Arkes, H. R., Back, M. D., Banks, G. C., Beevers, C., Bennett, A. A., Bleidorn, W., Boyer, T. W., Cacciari, C., Carter, A. S., Cesario, J., Clifton, C., Conroy, R. M., Cortese, M., Cosci, F., Cowan, N., Crawford, J., Crone, E. A., Curtin, J., Engle, R., Farrell, S., Fearon, P., Fichman, M., Frankenhuis, W., Freund, A. M., Gaskell, M. G., Giner-Sorolla, R., Green, D. P., Greene, R. L., Harlow, L. L., de la Guardia, F. H., Isaacowitz, D., Kolodner, J., Lieberman, D., Logan, G. D., Mendes, W. B., Moersdorf, L., Nyhan, B., Pollack, J., Sullivan, C., Vazire, S., and Wagenmakers, E.-J. (2020). A consensus-based transparency checklist. *Nature Human Behaviour*, 4(1):4–6.
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604):452–454.
- Barba, L. A. (2018). Terminologies for Reproducible Research. *arXiv:1802.03311 [cs]*.
- Bishop, D. (2019). Rein in the four horsemen of irreproducibility. *Nature*, 568(7753):435–435.
- Breznau, N. (2021). Does Sociology Need Open Science? *Societies*, 11(1):9.
- Brodeur, A., Lé, M., Sangnier, M., and Zylberberg, Y. (2016). Star Wars: The Empirics Strike Back. *American Economic Journal: Applied Economics*, 8(1):1–32.
- Burlig, F. (2018). Improving transparency in observational social science research: A pre-analysis plan approach. *Economics Letters*, 168:56–60.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick,

- N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E.-J., and Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9):637–644.
- Card, D., DellaVigna, S., and Malmendier, U. (2011). The Role of Theory in Field Experiments. *Journal of Economic Perspectives*, 25(3):39–62.
- Carey, B. (2011). Fraud Case Seen as a Red Flag for Psychology Research. *The New York Times*.
- Christensen, G. and Miguel, E. (2018). Transparency, Reproducibility, and the Credibility of Economics Research. *Journal of Economic Literature*, 56(3):920–980.
- Christensen, G. S., Freese, J., and Miguel, E. (2019). *Transparent and Reproducible Social Science Research: How to Do Open Science*. University of California Press, Oakland, California.
- Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J., Orben, A., Parsons, S., and Schulte-Mecklenbeck, M. (2018). 7 Easy Steps to Open Science: An Annotated Reading List.
- Elliott, K. C. (2020). A Taxonomy of Transparency in Science. *Canadian Journal of Philosophy*, pages 1–14.
- Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *PLOS ONE*, 4(5):e5738.
- Franco, A., Malhotra, N., and Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203):1502–1505.
- Freese, J. and Peterson, D. (2017). Replication in Social Science. *Annual Review of Sociology*, 43(1):147–165.
- Gerber, A. and Malhotra, N. (2008a). Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals. *Quarterly Journal of Political Science*, 3(3):313–326.
- Gerber, A. S. and Malhotra, N. (2008b). Publication Bias in Empirical Sociological Research: Do Arbitrary Significance Levels Distort Published Results? *Sociological Methods & Research*, 37(1):3–30.
- Gilbert, D. T., King, G., Pettigrew, S., and Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science". *Science*, 351(6277):1037–1037.
- John, L. K., Loewenstein, G., and Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, 23(5):524–532.
- Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., IJzerman, H., Nilsson, G., Vanpaemel, W., and Frank, M. C. (2018). A Prac-

- tical Guide for Transparency in Psychological Science. *Collabra: Psychology*, 4(1).
- Lindsay, D. S. (2020). Seven steps toward transparency and replicability in psychological science. *Canadian Psychology/Psychologie canadienne*, 61(4):310–317.
- Mertens, G. and Krypotos, A.-M. (2019). Preregistration of Analyses of Preexisting Data. *Psychologica Belgica*, 59(1):338–352.
- Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B. A., Petersen, M., Sedlmayr, R., Simmons, J. P., Simonsohn, U., and der Laan, M. V. (2014). Promoting Transparency in Social Science Research. *Science*, 343(6166):30–31.
- Moore, D. A. (2016). Preregister if you want to. *The American Psychologist*, 71(3):238–239.
- Motta, M. (2018). The Dynamics and Political Implications of Anti-Intellectualism in the United States. *American Politics Research*, 46(3):465–498.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S., Breckler, S., Buck, S., Chambers, C., Chin, G., and Christensen, G. (2014). Transparency and Openness Promotion (TOP) Guidelines.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E. L., Simonsohn, U., Soderberg, C., Spellman, B. A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E. J., Wilson, R., and Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242):1422–1425.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., and Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11):2600–2606.
- O’Boyle, E. H., Banks, G. C., and Gonzalez-Mulé, E. (2017). The Chrysalis Effect: How Ugly Initial Results Metamorphosize Into Beautiful Articles. *Journal of Management*, 43(2):376–399.
- on Earth, D. and on Behavioral, C. B. (2019). Reproducibility and Replicability in Science. *undefined*.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716–aac4716.
- Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, 334(6060):1226–1227.

- Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11):1359–1366.
- Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2):534–547.
- Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics*, 12(1):53–74.
- Stewart, S., Rinke, E. M., McGarrigle, R., Lynott, D., Lunny, C., Lautarescu, A., Galizzi, M. M., Farran, E. K., and Crook, Z. (2020). Pre-registration and Registered Reports: A Primer from UKRN.
- Thibodeaux, J. (2016). Production as social change: Policy sociology as a public good. *Sociological Spectrum*, 36(3):183–190.
- Wilson, F. D., Smoke, G. L., and Martin, J. D. (1973). The Replication Problem in Sociology: A Report and a Suggestion\*. *Sociological Inquiry*, 43(2):141–149.