

Master en Psicología General Sanitaria

TRABAJO DE FIN DE MASTER

La Crisis de la Replicación en Psicología:

Análisis de los factores que la causan y recomendaciones

Sayoa María Górriz Eguaras

Directoras: Jone Aliri Lazcano y Arantxa Gorostiaga Manterola

Índice

Resumen	1
1. Fundamentación teórica: Origen de la crisis de la replicación.....	3
2. Objetivos.....	7
3. Procedimiento.....	8
4. Resultados.....	8
4.1. Factores que afectan a la capacidad de replicación en cada fase de la investigación.8	
4.1.1. Planteamiento del problema.....	9
4.1.2. Revisión de la literatura y desarrollo del marco teórico	10
4.1.3. Definición de las hipótesis y diseño de la investigación.....	11
4.1.4. Definición de la población y selección de la muestra.....	12
4.1.5. Recogida de los datos	13
4.1.6. Análisis de los datos	14
4.1.7. Elaboración de informes, publicación y divulgación.....	16
4.2. Recomendaciones para fomentar la transparencia en la investigación.....	19
4.2.1. Pre-registro y planes de análisis.....	19
4.2.2. Recomendaciones metodológicas	21
4.2.3. Mejoras en la formación sobre metodología de la investigación.....	23
4.2.4. Investigación colaborativa	24
4.2.5. Unificación de criterios de presentación de informes, publicación y métodos de difusión	25
4.2.6. Flujo de datos, materiales y condiciones experimentales abiertas.....	26
5. Conclusiones.....	28
6. Referencias Bibliográficas.....	33
Anexo A. Gráfico Interactivo	39
Anexo B. CONSORT	40
Anexo C. STROBE	41
Anexo D. Declaración PRISMA	43
Anexo E. Pautas TREND	44
Anexo F. Directrices TOP	47
Anexo G. Ejemplo Lista de Verificación	48

Resumen

La crisis de la replicación en psicología es un problema al que lleva enfrentándose esta disciplina, junto a otras ciencias de la salud, desde 2010, cuando se comenzaron a destapar casos de fraude y dificultades en la replicación de diversas investigaciones. El objetivo del presente trabajo consiste en identificar los factores que inciden en las dificultades en la replicación de las investigaciones, centrándonos, sobre todo en las prácticas cuestionables. Asimismo, hemos considerado muy importante plantear posibles soluciones frente a estos factores. Por ello, el segundo objetivo del trabajo consiste en proponer recomendaciones que permitan hacer frente a las posibles prácticas cuestionables. Con el fin de cumplir con estos objetivos, se realizó una búsqueda en las bases de datos Scopus y Web of Science, en las que se encontraron numerosos artículos centrados tanto en los factores como en las propuestas de mejora para esta crisis. Entre los factores se han identificado diferentes sesgos como el de confirmación o publicación y prácticas como el HARKing o *P*-Hacking, entre otras, que podrían estar en la base de esta crisis de la replicación. Entre las recomendaciones propuestas por investigadores/as del campo de la psicología que pueden ayudar a mejorar la transparencia en futuras investigaciones y, por lo tanto, fomentar su replicación, destacan la ciencia colaborativa, el pre-registro y los planes de análisis, el flujo de datos abiertos y la estandarización de los informes. El trabajo finaliza con una serie de reflexiones acerca de hacia dónde debe dirigirse la psicología como disciplina para mejorar su transparencia.

Palabras clave: Crisis de la replicación, prácticas de investigación cuestionables, psicología, transparencia, sesgos, recomendaciones.

Abstract

The replication crisis in psychology is a problem that this science, along with other health sciences, has been facing since the 2010, when cases of fraud and difficulty in replicating various investigations began to be uncovered. The objective of this work is to identify the factors that affect the difficulties in replicating research, focusing, above all, on questionable research practices. Likewise, we have considered very important to propose possible solutions to these factors. Therefore, the work's second objective is to propose recommendations that allow dealing with possible questionable research practices. In order to meet these objectives, a search was carried out in Scopus and Web of Science electronic databases, where numerous papers were found that investigated both the factors and the improvement proposals for this crisis. Among the factors, different biases have been identified, such as confirmation or publication biases and practices such as HARKing or *P*-Hacking, among others, which could be at the base of this replication crisis. Among the recommendations proposed by researchers in the field of psychology that can help improve transparency in future research and, therefore, encourage its replication, stand out collaborative science, pre-registration and analysis plans, flow of open data and standardization of reports. The work ends with a series of reflections about where psychology as a discipline should go to improve its transparency.

Keywords: Replication crisis, questionable research practices, psychology, transparency, bias, recommendations.

1. Fundamentación teórica: Origen de la crisis de la replicación.

El término “crisis de la replicación” comenzó a utilizarse en psicología como consecuencia de algunos incidentes detectados en publicaciones científicas en la década de 2010. Esta crisis, no solo se da en la disciplina de la psicología, sino que también se da en otras ciencias de la salud como la medicina, la biomedicina y la farmacia. En lo que respecta a nuestro ámbito, una de las consecuencias de esta crisis sería generar dudas sobre el nivel de evidencia de los tratamientos psicológicos, lo que en última instancia podría interferir en la adherencia a los mismos. Por ello consideramos importante analizar el origen y las posibles causas de dicha crisis, y finalmente proponer ciertas pautas de actuación.

Con el fin de clarificar la confusión terminológica existente en este ámbito, y de delimitar el alcance del presente trabajo cuando nos referimos a “crisis de la replicación”, comenzaremos con la definición previa de una serie de conceptos relacionados entre sí. En primer lugar, distinguiremos los conceptos de replicación y reproducción, y definiremos los diferentes tipos de replicación.

Aunque no exista una definición universal del concepto, se define *replicación* como la capacidad de consecución de resultados similares a través de una investigación que se realiza de manera independiente a la original y en un momento posterior, con un diseño similar, llevado a cabo por los/as mismos/as investigadores/as o por otros/as independientes. Por su parte, la *reproducción*, hace referencia a la capacidad de repetir de manera exacta los mismos resultados de un estudio, a partir de los datos originales de este (Bretz et al. 2019). En este trabajo nos centraremos en el concepto de replicación.

Podemos distinguir dos tipos de replicación: la literal y la constructiva. Aunque distintos autores han utilizado distintas nomenclaturas, como recoge Schmidt (2016), en esencia hacen referencia a los tipos que hemos señalado. La *replicación literal*, hace referencia a una replicación lo más exacta posible de la investigación original (condiciones experimentales, participantes, procedimientos...), mientras que la *replicación constructiva o conceptual*, utiliza diferentes métodos para verificar la misma teoría o hipótesis del estudio original (Blomquist, 1986). Tanto en la replicación conceptual como en la literal, el hecho de que un estudio de replicación no consiga replicar los resultados del original no implica que el estudio original se haya realizado de manera indebida. En ocasiones puede resultar difícil o imposible replicar un estudio

debido a que ciertas circunstancias pueden no ser replicables. Por lo tanto, son los/as investigadores/as quienes deben analizar el porqué de las diferencias en los resultados y deben justificar si el hecho de no haberse replicado los resultados implica que los resultados del estudio original se han de poner en duda o no (Nelson, 2018). Una de las principales razones por las que se podrían cuestionar los resultados de un estudio sería por el hecho de que los/as investigadores/as hubieran utilizado prácticas de investigación cuestionables o inadecuadas.

Continuando con la clarificación terminológica, consideramos importante distinguir entre las prácticas de investigación cuestionables (a partir de ahora QRP por el término en inglés *Questionable Research Practices*) y las prácticas indebidas de investigación (a partir de ahora RM por el término en inglés *Research Misconduct*). A pesar de no existir una definición única, se engloba bajo el término QRP a aquellas prácticas que pueden ser perjudiciales dentro del proceso de investigación, que van en contra de los valores tradicionales de la investigación y, por tanto, atentan contra la transparencia y validez de ésta. Por ejemplo, se englobarían dentro de las QRP prácticas como *P-Hacking*, *HARKing*, dragado de datos, expediciones de pesca o minería de datos, *Cherry picking* y sesgos como el de confirmación, apofenia, selección, esencias ilusorias, retrospectivo y de publicación. Tales prácticas estarían situadas entre las buenas prácticas y las RM. Estas últimas, engloban conductas tales como la fabricación y la falsificación de datos y el plagio (Xie et al., 2021). En este trabajo nos centraremos en las QRP, ya que consideramos que, en la gran mayoría de casos, son conductas que no se realizan de mala fe, y por lo tanto se pueden prevenir y evitar. Tras esta clarificación terminológica, continuamos con los acontecimientos que dieron lugar a la “crisis de la replicación”.

En relación con las prácticas indebidas, en el ámbito de la psicología uno de los incidentes más reseñables fue el de Diederik Stapel, investigador de gran prestigio, que había conseguido publicar investigaciones en revistas de gran impacto como *Science*. Tras un informe de investigación realizado por la revista *Nature*, se descubrió el fraude en los datos de numerosas de sus más brillantes y trascendentes publicaciones, que habían sido ovacionadas por gran parte de la comunidad científica. Estas publicaciones se situaban principalmente en el área de la psicología social y se centraban en constructos tales como los estereotipos o la discriminación (Callaway, 2011).

Otro de los incidentes más notables del inicio de esta crisis en el ámbito de la psicología fue el estudio de Bem (2011), que concluía que los humanos podían sentir

eventos futuros y utilizar dicha información para juzgar el presente. Dicho estudio, aunque mostraba resultados controvertidos, ya que apoyaba la existencia de fenómenos paranormales, fue publicado en la revista de prestigio *Journal of Personality and Social Psychology* y, por lo tanto, se consideró que aportaba evidencias científicas del citado fenómeno. Cuando Bem publicó su artículo en el año 2011, otros/as investigadores/as trataron de replicar el mismo, y comprobaron que no podían obtener los mismos resultados. Sin embargo, tuvieron dificultades para publicar sus estudios, ya que el editor de la revista que publicó los hallazgos de Bem, al principio no permitió que estos artículos accedieran a la fase de revisión por pares (Schimmack, 2012). A día de hoy, ya se conoce que Bem utilizó prácticas cuestionables a la hora de realizar dicho estudio (Francis, 2012; Schimmack, 2012), sin embargo, si no se hubieran realizado y publicado los estudios de replicación, es posible que no se hubieran detectado las prácticas llevadas a cabo por Bem.

Como se ha señalado previamente, el hecho de que un estudio de replicación no consiga obtener los resultados del estudio original no necesariamente se debe a que en el estudio original se hayan llevado a cabo prácticas cuestionables o indebidas. Sin embargo, también, se pueden considerar antecedentes de esta crisis, los problemas de replicación de ciertos experimentos clásicos como el de Bargh et al. (1996) sobre el *priming*, en el que se concluía que los jóvenes caminaban más despacio tras salir de un experimento que consistía en la observación de personas ancianas. El equipo de Doyen et al. (2012) realizó una replicación literal de dicho estudio, y concluyó que los resultados obtenidos no se parecían a los obtenidos en el original, ya que en su estudio la velocidad de los jóvenes no se redujo tras el experimento. La publicación de dicha replicación generó un debate sobre las posibles causas de las discrepancias en los resultados, ya que también se realizaron y publicaron numerosas replications conceptuales que sí habían obtenido resultados que estaban en línea con los del estudio original. En este caso, la posible explicación fue la falta de validez externa histórica de los resultados originales. Es decir, se concluyó que una razón plausible de que dicho estudio de replicación no obtuviera resultados similares al original se debía a que el “caminar despacio” no fuera una característica central del estereotipo de las personas mayores en la sociedad belga 20 años más tarde (Stroebe y Strack, 2014).

Tras constatar el hecho de que solo un 1% de los estudios publicados en revistas de impacto era un estudio de replicación (Makel et al. 2012), la Open Science Collaboration (2015) realizó un gran esfuerzo colaborativo en diferentes laboratorios o

grupos de investigación para replicar 100 experimentos en el área de la psicología. Así, se observó que gran parte de los hallazgos contaron con una evidencia más débil respecto a los resultados de los experimentos originales, aun habiendo usado el mismo material que los autores originales. La mayoría de los efectos estaban en consonancia con los resultados originales, pero no se encontraron los mismos tamaños del efecto y niveles de significación. Concretamente, concluyeron que únicamente 36 de los 100 estudios replicados presentaron resultados estadísticamente significativos, es decir, únicamente estos estudios habían rechazado la hipótesis nula con una p inferior a 0,05. Este hecho creó una gran controversia, ya que la prensa se hizo eco de los resultados y publicó noticias bastante alarmistas (por ejemplo, en el New York Times se publicó un artículo con el título “Many psychology findings not as strong as claimed, study says”, Carey, 27 de agosto de 2015). Sin embargo, aunque tradicionalmente las replications se han considerado fallidas cuando el efecto descrito en el estudio original no ha sido estadísticamente significativo en el estudio de replicación, dicha consideración tiene al menos dos problemas importantes. Por un lado, el estudio de replicación puede haber obtenido resultados no significativos por haberse realizado con un tamaño muestral insuficiente y, por otro lado, el estudio de replicación puede haber obtenido resultados significativos, por haberse realizado con un tamaño muestral muy grande, aunque el tamaño del efecto sea muy inferior al estudio original. En este segundo caso no parece muy lógico concluir que el estudio de replicación haya sido exitoso. Otros/as autores/as han optado por comparar los tamaños del efecto para concluir si la replicación es exitosa o no, pero esta estrategia también tiene sus problemas. Uno de los problemas que presentan ambas estrategias es que no tienen en cuenta que una replicación puede tener resultados no concluyentes, es decir, solo consideran que un estudio de replicación puede ser exitoso o fallido. Sin embargo, en ocasiones es posible que los resultados no sean concluyentes. Así, si tenemos en cuenta este hecho, la interpretación de los resultados del Open Science Collaboration cambia. De nuevo el 36% de los resultados serían replications exitosas, pero del 64% restante solo el 25% serían replications fallidas, y el 49% serían replications no concluyentes (Nelson et al., 2018).

Los citados sucesos y el hecho de que en muchos casos los/as investigadores/as no proporcionaran los datos originales para la reproducción, hicieron que se comenzasen a cuestionar los conocimientos hasta entonces adquiridos en ciencias de la salud como la psicología (Pashler y Wagenmakers, 2012). Asimismo, Fanelli (2009), publicó uno de los

primeros meta-análisis en el ámbito de la Medicina y la Farmacia sobre prácticas cuestionables e indebidas en la investigación de ambas ciencias. En dicho estudio, se analizó la prevalencia de las prácticas de investigación cuestionables e indebidas y se concluyó que casi un 2% de científicos admitía haber fabricado, falsificado o alterado de algún modo los datos o los resultados al menos una vez. Asimismo, se puso en evidencia que casi un 10% de investigadores/as de estas disciplinas admitía haber realizado alguna otra conducta que podría incluirse dentro del paraguas de prácticas de investigación cuestionables.

A raíz de estos hechos, la psicología comenzó a realizar una cierta autocrítica y a observar cuáles eran los problemas que se estaban dando en este sentido. Algunos/as expertos/as consideran que comenzó un nuevo periodo en el que se concede mayor importancia a la metodología de la investigación y se produce una tendencia hacia una mejora y transparencia de la investigación (Nelson et al. 2018).

2. Objetivos

Al tratar de profundizar en la denominada crisis de replicación en psicología, nos encontramos con distintos tipos de factores y prácticas no adecuadas por parte de los investigadores/as, que pueden estar en el origen de la crisis o la están afectando en distinta medida. Por ello, el primer objetivo del presente trabajo consiste en realizar una síntesis de los factores que pueden intervenir de manera directa o indirecta en las dificultades de replicación de las investigaciones originales. Concretamente, en este trabajo nos centraremos en las denominadas prácticas cuestionables o QRP. Asimismo, consideramos muy importante plantear posibles soluciones frente a estos factores, ya que muchas de estas prácticas se pueden producir por desconocimiento o por errores involuntarios. Por ello, el segundo objetivo del trabajo es recopilar las recomendaciones más relevantes que encontramos en la literatura, planteadas por autores/as expertos/as en el tema que fomenten una mayor transparencia en la investigación y en las publicaciones. Creemos que cumpliendo estos objetivos se pueden ir dando pasos para mejorar la transparencia y la calidad de la investigación psicológica del futuro. Asimismo, consideramos que una literatura científica más transparente y de mayor calidad metodológica redundará en una mejor identificación de los tratamientos basados en la evidencia, así como en una mayor credibilidad de los mismos.

3. Procedimiento

Con el fin de responder a los objetivos planteados, se realizó una revisión utilizando dos bases de datos electrónicas: Scopus (considerada la mayor base de datos de citas y resúmenes de bibliografía revisada por pares) y Web of Science (la plataforma multidisciplinar más grande con estudios de alta calidad). La búsqueda no se acotó temporalmente y se incluyeron artículos científicos, capítulos de libros y libros. Los términos de búsqueda utilizados fueron “replication crisis” o “replicability crisis”, de manera individual o combinada con el término “psychology”. La primera selección se realizó mediante la lectura de títulos y resúmenes, seleccionando aquellos estudios que se centraban en el tema de interés, excluyendo aquellos que no fueran en inglés o español. Tras ello, se procedió a leer el texto completo de las publicaciones seleccionadas. Además de extraer la información necesaria para la revisión, la lectura de estas publicaciones permitió identificar a muchos/as de los/as investigadores/as expertos/as en el tema, de manera que se pudieron identificar otros trabajos que permitieron lograr los resultados que se presentan a continuación.

4. Resultados

4.1. Factores que afectan a la capacidad de replicación en cada fase de la investigación.

Como hemos señalado previamente, en el presente trabajo nos centraremos en las QRP más que en las RM, sobre todo porque entendemos que las primeras ocurren sin mala fe por parte de los/as investigadores/as, por desconocimiento, por errores cometidos o por dificultades de diversa índole. Sin embargo, son prácticas que, en algunos casos, se encuentran ya normalizadas en muchas de las tareas investigadoras y pueden atentar contra la validez y transparencia de la investigación, por lo que es imprescindible conocerlas y detectarlas. Así, se presentan algunos de los múltiples factores citados en la literatura científica en los que podemos encontrar claves para entender las razones de la crisis investigadora en la que la psicología se ha visto inmersa como disciplina científica, al igual que otras ciencias de la salud.

Con respecto a la prevalencia, diversos meta-análisis han estudiado la frecuencia de estas prácticas en la investigación. Uno de los más actuales, que busca cubrir la brecha temporal entre los meta-análisis anteriores, como el antes mencionado de Fanelli (2009) y la situación actual, muestra que el 2,9% de investigadores/as señalan haber cometido al

menos una RM en sus investigaciones y el 12,5% había realizado una o más QRP. Además, este meta-análisis reporta que casi un 16% declaraba haber sido testigo de RM por parte de otros/as investigadores/as y casi el 40% había sido testigo de la realización de diferentes QRP (Xie et al., 2021).

A su vez, en una encuesta de 2022 realizada a 6.813 investigadores/as, alrededor de un 4% reportaron haber fabricado datos y otro 4% haberlos falsificado. Además, alrededor del 51% de los/a encuestados/as reconocieron participar con frecuencia en al menos una QRP. Respecto a los perfiles y características de los/las investigadores/as, se observó que ser doctorando/a o joven investigador/a, así como la presión por publicar, aumentaron las probabilidades de incurrir en al menos una QRP. Sin embargo, una adhesión a las normas científicas y la percepción de la probabilidad de detección por parte de los/as revisores/as se asoció con menos conductas cuestionables (Gopalakrishna et al., 2022).

Dentro de las QRP podríamos encontrar desde errores involuntarios que se pueden dar en todas las publicaciones y en todos los procesos como, por ejemplo, al codificar los datos manualmente, los análisis estadísticos mal ejecutados, las erratas... siendo más leves y muchas veces corregibles (Nelson et al., 2018), hasta conductas más graves, como el P-Hacking, conductas todas ellas que pueden llegar a comprometer los resultados de la investigación. Ante estas conductas plantearemos posibles soluciones en el siguiente apartado.

Con fines didácticos, estructuraremos las QRP basándonos en el proceso de la investigación, aunque algunas de ellas se pueden dar en más de una fase y no de manera aislada, sino una precedida o suscitada por otra como mencionaremos a continuación. Concretamente, las fases en las que las dividiremos serán: Planteamiento del problema, Revisión de la literatura y desarrollo del marco teórico, Diseño de la investigación e hipótesis, Definición y selección de la muestra, Recogida de los datos, Análisis de los datos, Elaboración de informes, publicación y divulgación. Al inicio de cada una de las fases haremos una breve descripción de dicha fase para contextualizar los problemas que pueden darse en la misma.

4.1.1. Planteamiento del problema

El problema, hace referencia a una pregunta que se plantea al inicio de la investigación para la cual se espera obtener respuesta al finalizar dicha investigación

(Castro, 2002). Por tanto, en esta fase se formula una oración interrogativa, acerca de la relación entre dos o más variables (Kerlinger, 2002). A priori, esta relación no ha sido estudiada anteriormente o al menos existe cierta controversia sobre la misma, y es, por tanto, la pieza angular de la que parte toda investigación.

En general, el interés que prima en los/as investigadores/as cuando se plantean el problema es el del avance del conocimiento. Sin embargo, en ocasiones pueden surgir otros tipos de interés. Cuando la confluencia de diversos intereses incompatibles interviene a la hora de realizar una investigación, nos encontramos ante un fenómeno conocido como **conflicto de intereses** (Nosek et al., 2012). Esto implica la existencia de diversos intereses económicos, políticos... que pueden entrar en conflicto con los intereses propios de los/as científicos/as. Esta situación puede generar que no siempre se priorice el avance del conocimiento, por ejemplo, no dándole importancia a investigaciones de replicación, y valorando otros aspectos con la intención de tener más oportunidades de conseguir financiación o tener más facilidades para publicar los resultados. Estos hechos pueden hacer que se prime la novedad de la temática, el que la temática este en boga o que trate sobre temas socialmente más atractivos (Fernández et al., 2017). Sin embargo, en esta crisis de la replicación se hace patente la importancia del eslabón de la replicación para afianzar el conocimiento, ya que gracias a ella se han podido poner en tela de juicio multitud de investigaciones, que finalmente no eran tan sólidas como aparentaban.

4.1.2. Revisión de la literatura y desarrollo del marco teórico

La revisión bibliográfica hace referencia a una fase del proceso investigador que consiste en una revisión exhaustiva del estado del problema, para con ello, poder realizar un marco teórico que contextualice la investigación. El marco teórico es el conjunto de principios teóricos preexistentes que guían la investigación y que ayudan a establecer las unidades de análisis relevantes para cada problema (Castro, 2002).

En esta fase destaca el **sesgo de confirmación**, un sesgo de los más polémicos y cuestionados. Se refiere a la tendencia de los/as investigadores/as a encontrar información y/o buscar evidencias que confirmen sus expectativas sobre el problema, descartando la información que sea contradictoria (Glick, 2017).

En 1998, Nickerson, publicó un estudio de revisión sobre este sesgo mostrando las posibles formas en las que se podía dar y exponiendo diferentes situaciones reales y

ámbitos en los que se encontraba. Entre las diversas maneras que señaló, destacamos aquellas que puedan afectar directamente en el porqué de esta crisis, como por ejemplo, la búsqueda sesgada de información, sobreponderando positivamente los puntos de vista y casos que apoyen la hipótesis del/de la investigador/a y desechando o desvalorizando aquellas que no vayan en línea con la hipótesis; la interpretación sesgada, tanto de la información como de los resultados, donde se tiende a ver patrones coherentes con la hipótesis en la información encontrada o en los datos (relacionado con el sesgo de apofenia, que se describe en la siguiente fase); y la memoria selectiva, que hace que se recuerden mejor los datos que apoyan la hipótesis. Al respecto de esta última, podemos destacar también cómo el efecto de primacía y persistencia en las creencias actúan de manera que se tiende a validar y apoyar aquello que se aprendió anteriormente, desvalorizando las nuevas hipótesis o conocimientos.

Este sesgo cobra gran importancia en el ámbito de la investigación, puesto que podría llegar al extremo de que todo el marco teórico en el que se fundamenta la investigación esté sesgando la realidad de la que se parte, pudiendo llegar a conclusiones que no estén basadas en la realidad y a sentirnos tentados a realizar algunas de las diversas QRP, como las que explicaremos a continuación, que nos ayudarían a adaptar nuestros resultados a las hipótesis.

4.1.3. Definición de las hipótesis y diseño de la investigación

Tras delimitar el problema y realizar la fundamentación y el marco teórico, llega el momento de avanzar una posible explicación al problema mediante el planteamiento de las hipótesis. Las hipótesis son enunciados que, tras ser comprobados empíricamente, permitirán resolver el problema y comprenderlo (Montserrat, 1983), o al ser refutados, descartar estas posibles soluciones. Es importante remarcar que no deben ser explicaciones subjetivas, sino basadas en conocimientos preexistentes, para no ser planteadas de manera arbitraria (Losada y López-Feal, 2003).

Para llegar a la comprobación de las hipótesis se decide un diseño de investigación que guía al/a la investigador/a hacia esta comprobación y, por tanto, hacia la verificación o falsación de estas (Popper, 1990). Con ese fin, habrá que decidir cuál será el diseño adecuado de la investigación: experimental, cuasiexperimental, selectivo u observacional (Castro, 2002).

Principalmente, el sesgo más importante en esta fase es el denominado **HARKing** (acrónimo de la frase inglesa *hypothesizing after the results are known*). Este fenómeno consiste en definir las hipótesis una vez analizados los resultados, presentándolas como si hubieran sido definidas a priori (Kerr, 1998). Esta práctica, llevaría al investigador o la investigadora a reformular las hipótesis tantas veces como fueran necesarias para que sean coherentes con los resultados obtenidos.

Por ejemplo, supongamos que se está estudiando si los niveles de ansiedad del alumnado en general aumentan cuando llega el periodo de exámenes y la hipótesis inicial fuera que sí aumentan dichos niveles de ansiedad. Supongamos también que los resultados mostraran un aumento de los niveles de ansiedad únicamente en los cursos de la ESO y no en los de primaria, y que al redactar el artículo se dijera que la hipótesis inicial era que la ansiedad del alumnado aumenta cuando llega el periodo de exámenes en los/as estudiantes de los cursos de la ESO, pero no en primaria. En este caso, se estaría incurriendo en esta práctica, reformulando la hipótesis y aumentando la probabilidad de publicar, como se señalará más adelante.

En esta misma fase también encontramos el **sesgo de apofenia**. Realmente, este sesgo podría situarse tanto en esta fase como en la anterior o incluso en la fase de resultados, ya que está muy relacionado con el sesgo de confirmación. Este sesgo, se define como la capacidad de percibir patrones de fenómenos en principio no relacionados, normalmente de manera inconsciente (Buetow, 2019), como ocurre en el juego con la “Falacia del Jugador” donde tras conseguir una secuencia de pérdidas en un juego de azar existe la creencia de que de un momento a otro comenzarán las victorias para compensar las pérdidas (Attorresi et al., 2006). Este sesgo afecta directa o indirectamente a los resultados, puesto que gracias a él se pueden llegar a gestar teorías o hipótesis que no tienen una base respaldada científicamente. De esta manera, se podría llegar a plantear hipótesis no justificadas y arbitrarias, llegando, como pasaba en el sesgo de confirmación, a incurrir en QRP. Este hecho produciría que la investigación fuera más viable para su publicación, a pesar de su inicial arbitrariedad en el planteamiento.

4.1.4. Definición de la población y selección de la muestra

Definir la población diana en la que se pretende contrastar las hipótesis, es realmente importante, porque afectará a la selección de la muestra y, por tanto, a las posibilidades de generalización del estudio (Castro, 2002). Una vez seleccionada la población, se determinan los criterios, tanto de inclusión como de exclusión, con los que

delimitar la muestra y, además, se selecciona la técnica de muestreo más adecuada para conseguirla.

En esta fase, nos podemos encontrar con el **sesgo de selección**, que implica seleccionar la muestra a partir de una subpoblación acorde con los resultados que se pretenden obtener (Padmanabhan, 2014), produciendo así, unos resultados predeciblemente favorables en cuanto a la verificación de la hipótesis. Siguiendo con el ejemplo de la ansiedad, si se está estudiando el aumento de la ansiedad durante los exámenes, y se pretende probar la hipótesis previamente comentada (que los niveles de ansiedad del alumnado aumentan cuando llega el periodo de exámenes) se podría seleccionar a los/as alumnos/as de bachillerato cuando van a realizar las pruebas de acceso a la universidad. En este contexto sería más probable confirmar la hipótesis en comparación con una situación en la que se seleccione una muestra representativa de estudiantes de 12 a 18 años (incluyendo estudiantes de la ESO, por ejemplo), o en una situación que no implique las pruebas de acceso a la universidad.

Asimismo, dentro de esta fase de la investigación cabe destacar un fenómeno en el que se ha puesto el foco en esta crisis, la **falta de generalizabilidad** de los estudios originales que posteriormente son replicados. Este sesgo puede darse como consecuencia directa del sesgo de selección o de la falta de representación muestral. En este caso diferentes estudios pueden mostrar efectos estadísticamente significativos, aunque la muestra utilizada no siempre represente adecuadamente la población de interés, limitando la generalizabilidad de los resultados. A veces, las muestras son poco representativas, y por tanto no son realmente útiles a la hora de generalizar los efectos a la población, aunque se pudieran publicar como tal (Ioannidis, 2005; Yarkoni, 2020). Esta falta de generalizabilidad estaría atentando contra la validez externa de la investigación, siendo esta, el grado de confianza con que una presumible relación causal entre variables se puede generalizar a otras personas, poblaciones y contextos diferentes a los/as de la investigación (Balluerka, 2011).

4.1.5. Recogida de los datos

Una decisión fundamental a la hora de realizar la investigación es la que se refiere a la estrategia más adecuada para la recogida de datos (Castro, 2002) y ponerla en práctica siguiendo fielmente el procedimiento diseñado.

Algunos de los instrumentos que se utilizarán para recoger los datos serán test estandarizados o entrevistas estructuradas o semiestructuradas, entre otros. En esta fase es de gran importancia utilizar definiciones operacionales adecuadas de los constructos.

Algunos autores apuntan en este punto a la posible falta de coherencia entre los instrumentos para medir los diferentes constructos (inteligencia, agresividad, empatía...), con los enunciados de las hipótesis que realmente se quieren contrastar. Como afirman Brick et al. (2022), nos encontramos ante un sesgo denominado **esencias ilusorias** donde al realizar la operacionalización de los constructos de las hipótesis se da un excesivo reduccionismo de los constructos complejos sin incluir todos los matices necesarios. En este sentido, como afirma Yarkoni (2020), no siempre se estaría midiendo realmente lo que se pretende medir. Esta cuestión, puede llevar a un sesgo en los resultados por una inadecuada medición, es decir, se estaría atentando contra la validez de constructo. La validez de constructo hace referencia al grado de confianza con que podemos establecer generalizaciones a constructos de orden superior a partir de las definiciones operacionales o indicadores utilizados en la investigación (Balluerka, 2011).

4.1.6. Análisis de los datos

El análisis estadístico de los datos hace referencia, por un lado, a la organización y descripción de estos (estadística descriptiva) y, por otro lado, a la estimación de parámetros y contraste de hipótesis (estadística inferencial). Para ello, es necesario poner en marcha un plan de análisis estadístico en el que utilizar las pruebas más adecuadas, dependiendo del nivel de medida de las variables que se pretenden estudiar, y el tipo de relación que se pretende analizar (Castro, 2002).

En esta fase de la investigación se pueden dar diversas QRP. Rosenthal (1979), señaló que muchos de los estudios con resultados estadísticamente no significativos se quedaban sin publicar por el sesgo de publicación o el efecto cajón (permanecer sin sacar a la luz todos los estudios sin efectos estadísticamente significativos). Por su parte, Nelson et al. (2018) afirman que no existe de manera tan generalizada ese efecto cajón y que actualmente los/as investigadores/as son más persistentes en su intento de publicar los resultados de sus investigaciones. Para asegurar la publicación algunos/as investigadores/as utilizan diferentes estrategias de análisis de sus datos para conseguir que sean estadísticamente significativos y, de esta manera ser más probable su publicación.

Algunas de las prácticas cuestionables relacionadas con el análisis de los datos son, entre otras, la eliminación de los valores extremos de manera poco transparente, sin explicar las razones de dicha eliminación (Valentine et al., 2021) o jugar con la variabilidad de los grados de libertad del/la investigador/a (Simmons et al., 2011) para llegar a resultados estadísticamente significativos. A la hora de analizar los datos, al igual que en las demás fases de la investigación, el/la investigador/a ha de tomar muchas decisiones. Tiene que decidir qué hacer con los valores perdidos (imputar o no imputar, si se decide imputar qué estrategia utilizar), qué hacer con los valores extremos, realizar los análisis con toda la muestra o haciendo subgrupos, qué puntuaciones utilizar (puede que se disponga de puntuaciones de subdimensiones y de puntuaciones totales, por ejemplo), qué estrategias analíticas utilizar, etc. Todas estas decisiones se refieren a lo que se denominan los grados de libertad del/la investigador/a, y la conducta cuestionable surgiría cuando las decisiones se toman en función de los resultados que se obtengan a raíz de estas decisiones. Es decir, realizar los análisis por subgrupos porque el objetivo es ver la relación entre las variables en diferentes grupos de edad sería adecuado, sin embargo, realizar los análisis por subgrupos porque en el análisis de la muestra total no se obtienen los resultados deseados no sería adecuado. De igual manera, existe la conducta cuestionable de realizar muchos análisis estadísticos e interpretarlos en base a la significación estadística sin realizar ningún tipo de ajuste del valor crítico de alfa (Gigerenzer, 2004; Ioannidis, 2005; Simmons et al., 2011). O incluso de interpretar los resultados en función de la significación estadística, sin tener en cuenta el tamaño del efecto, lo que puede llevar a falsos positivos (cuando se tienen muestras grandes con mucha potencia estadística), o a falsos negativos (cuando se tienen muestras reducidas con escasa potencia estadística).

Estas QRP son denominadas como *P-Hacking*, dragado de datos, expediciones de pesca o minería de datos... Todas ellas se refieren a estrategias adecuadas de análisis de datos que, llevadas al extremo o mal utilizadas, pueden dar lugar a prácticas no adecuadas que aumentarían la probabilidad de obtener resultados significativos y, consecuentemente, dificultarían la replicación. A continuación, describimos brevemente cada una de estas conductas que, tal y como acabamos de mencionar, pueden aumentar el número de resultados falsos positivos y por lo tanto incidir en la crisis de replicabilidad. El ***P-Hacking*** se caracteriza por analizar los datos obtenidos en la investigación y las variables que nos interesan desde diferentes perspectivas hasta encontrar una estrategia

analítica que asegure un resultado estadísticamente significativo. Por su parte, en las **expediciones de pesca**, se utiliza una práctica más indiscriminada donde el/la investigador/a examina de manera exhaustiva todas las posibles asociaciones entre diferentes variables, aun no siendo relevantes para su hipótesis inicial, hasta encontrar algún resultado estadísticamente significativo. Por último, el **dragado** o **la minería de datos**, consiste en expediciones de pesca extremas en las que se buscan relaciones estadísticamente significativas entre cualquiera de las variables existentes en la base de datos (Andrade, 2021). Dichas prácticas, no son procedimientos inadecuados en sí mismos, siempre y cuando se especifiquen y fundamenten, pero es cierto que no siempre están especificadas y pueden muy posiblemente ir seguidas por prácticas de HARKing.

En consecuencia, resulta difícil estimar cuántos de los estudios publicados pueden sufrir *P*-Hacking o dragado de datos, evidenciando que en ocasiones es igual de fácil publicar un positivo verdadero que uno falso, llegando a dudar de si el conocimiento existente es realmente fiable. Además, los intentos de replicaciones masivas como el mencionado anteriormente de Open Science Collaboration (2015) ponen en evidencia que es difícil realizar replicaciones que muestren los mismos efectos, lo que puede hacer pensar que en alguna medida ha podido haber algunas de estas QRP asociadas al análisis de datos.

4.1.7. Elaboración de informes, publicación y divulgación

Esta fase se caracteriza por la comunicación de los resultados encontrados en la investigación, además de la fundamentación y diseño de esta mediante un informe. Este informe conlleva una comunicación oral o escrita para que se den a conocer las aportaciones del estudio y que, además, proporciona nuevas líneas de trabajo, así como posibilidad de replicaciones por otros/as investigadores/as (Castro, 2002).

Posteriormente, con la publicación del informe los/as investigadores/as tratan de publicar los resultados del estudio en las revistas científicas más acordes a las temáticas del mismo. La divulgación de la publicación hace referencia al momento en el que la investigación se presenta como una forma de validación de la misma en la comunidad científica (Méndez y Pérez-Acosta, 2011).

En esta crisis nos encontramos con una falta de unificación y falta de transparencia a la hora de elaborar los informes y transmitir la información, la cual da lugar a dificultades para poder realizar replicaciones de las investigaciones iniciales. A menudo,

nos encontramos con falta de información en los estudios iniciales respecto a los tamaños de efectos y otros detalles de la investigación indispensables para la generalización y replicación (Nosek et al., 2012). En parte, este problema puede estar originado por la limitación del número de palabras que, en general, imponen muchas revistas científicas.

Además de esta falta de información, nos encontramos con el fenómeno **Cherry picking** (Andrade, 2021), que consiste en presentar únicamente los resultados significativos encontrados por los/as investigadores/as, omitiendo las relaciones no significativas encontradas, produciendo así, un sesgo de publicación, del que hablaremos a continuación. Esta práctica tiene como consecuencia una falta de transparencia de la totalidad de resultados de la investigación, que hace que al publicarse únicamente lo estadísticamente significativo se pierda una parte importante del conocimiento sobre las relaciones estudiadas.

También, nos podemos encontrar el **sesgo retrospectivo**, definido por Arkes et al. (1988), como la tendencia a sobrestimar la probabilidad de haber predicho la ocurrencia de un evento pasado. Esto produce un exceso de confianza en las predicciones de los/as investigadores/as y hace que se pierda el tono inquisitivo necesario para generar un mayor conocimiento científico. Incluimos este sesgo en esta fase, puesto que, en la discusión se podrían encontrar afirmaciones que sobreestimen las predicciones respecto a lo encontrado en los resultados, como si fuera algo obvio antes de realizar la investigación, pero que en realidad no es así porque se afirma después de analizar los datos. Por ejemplo, siguiendo con el ejemplo anterior, después de observar en una investigación que la ansiedad aumenta en mayor medida en las mujeres durante el periodo de exámenes, resultaría fácil afirmar en la discusión que se esperaba dicho resultado.

Por último, el **sesgo de publicación**, característico del proceso final de la investigación, se describe como la publicación selectiva de estudios con resultados concretos, prioritariamente favorables, donde se encuentran resultados estadísticamente significativos y acordes con la hipótesis propuesta. Dicha práctica, puede propiciar una visión sesgada de los efectos dentro de la psicología al estar invisibilizados muchos de los estudios no publicados con resultados de fenómenos no relacionados y resultados estadísticamente no significativos (Ferguson y Brannick, 2012). En este sentido, encontramos que, en psicología, se da un mayor número de resultados positivos publicados en comparación con otras ciencias (Miguel et al., 2014; Valero, 2015).

En la Tabla 1 se resumen las características de los factores relacionados con la crisis de replicación que hemos descrito en este apartado.

Tabla 1

Resumen de los factores que afectan a la capacidad de la replicación

Fases de la investigación	Factores	Características
Planteamiento del problema	-Conflicto de interés	-Interacción de diferentes intereses en la selección del tema o en la realización de la investigación (económicos, políticos...).
Revisión de la literatura y desarrollo del marco teórico	-Sesgo de confirmación	-Tendencia de búsqueda de información sesgada por las expectativas sobre el tema.
Diseño de la investigación e hipótesis	-HARKing	-Definición de la hipótesis tras la obtención de los resultados y presentación de esta como si hubiera sido definida a priori.
	-Sesgo de apofenia	-Percepción inconsciente de patrones en fenómenos en principio no relacionados.
Definición y selección de la muestra	-Sesgo de selección	-Selección de la muestra a partir de una subpoblación acorde con los resultados que se espera obtener.
	-Falta de generalizabilidad	-Selección de una muestra poco representativa de la población a la que se desean generalizar los resultados
Recogida de los datos	-Sesgo de esencias ilusorias	-Operacionalización inadecuada de los constructos. -Uso inadecuado de los instrumentos de medición.
Análisis de los datos	-P-Hacking	-Análisis de los datos obtenidos desde diferentes perspectivas hasta encontrar un resultado estadísticamente significativo.
	-Dragado de datos	-Análisis exhaustivo de todas las posibles asociaciones entre variables aun no siendo inicialmente relevantes, hasta encontrar un resultado estadísticamente significativo.
	-Expediciones de pesca o minería de datos	-Análisis en busca de relaciones estadísticamente significativas entre cualquiera de las variables existentes en la base de datos.
Elaboración de informes, publicación y divulgación	-Falta de estandarización de informes	-Falta de información en los estudios iniciales respecto a detalles de la investigación indispensables para la generalización y replicación.
	-Cherry picking	-Presentar solo los resultados significativos.
	-Sesgo retrospectivo	-Tendencia a sobrestimar la probabilidad de haber predicho la ocurrencia de un evento.
	-Sesgo de publicación	-Publicación selectiva de estudios con resultados concretos.

4.2. Recomendaciones para fomentar la transparencia en la investigación

Tal y como afirman algunos/as de los/as investigadores/as del ámbito de la psicología en relación con la crisis de la replicación, nos encontramos desde hace unos años en una época de confrontación activa en la que la replicación va cobrando fuerza para el avance de la ciencia. En ella, se comienzan a priorizar las características de replicación, reproducción y robustez, a pesar de los obstáculos con los que los/as investigadores/as se siguen encontrando en torno a la falta de incentivación y el poco interés hacia dichas características dentro de la cultura actual de la investigación (Nosek et al., 2022).

Una vez que hemos presentado los diversos factores que pueden situarse en el origen de la crisis de replicación o que pueden influir en la misma de manera directa o indirecta, a continuación, a partir de las propuestas de diversos/as autores/as, nos centraremos en presentar una serie de recomendaciones para tratar de paliar tales efectos. Consideramos que seguir tales recomendaciones permitiría cumplir con el carácter acumulativo que debe tener el conocimiento científico de manera cada vez más transparente, veraz y, por ende, replicable por diferentes investigadores/as y equipos de investigación.

Antes de comenzar, hay que tener en cuenta que las QRP anteriormente descritas a menudo no se dan de manera aislada y que pueden converger e interaccionar entre sí (Munafò et al., 2017). Por ello, las siguientes recomendaciones tendrán un carácter más generalizado que permitiría abordar varias de ellas simultáneamente.

Con el fin de realizar una exposición más clara y didáctica, las recomendaciones se organizan en seis grandes bloques de acuerdo al proceso óptimo de una investigación transparente.

4.2.1. Pre-registro y planes de análisis

Una de las recomendaciones más compartida por los/as investigadores/as es el pre-registro, tanto del estudio, con el compromiso de publicar los resultados cuando se obtengan, como de su diseño, las variables principales y el plan de análisis, antes de realizar el estudio (Miguel et al., 2014; Munafò et al., 2017; Wicherts, 2017). Al igual que existen en otras ciencias de la salud leyes reguladoras como la de 2007 en EEUU, según la cual los ensayos médicos tienen que ser previamente incluidos en un registro público de manera obligatoria, diferentes autores/as proponen desde hace años una

posible coordinación de las instituciones para que en el ámbito de la psicología se utilicen registros similares (Miguel et al., 2014). En este sentido, cabe señalar, que en los últimos años se han estado dando pasos en esta línea dentro de la comunidad científica.

Este pre-registro de investigaciones consistiría en especificar de manera detallada el diseño del estudio, especificando el procedimiento, los resultados principales esperados y el plan de análisis (Munafò et al., 2017). En este plan de análisis se detallarían tanto los análisis estadísticos que se pretenden realizar, como los modelos en los que se basan, las variables a medir, las covariables, los posibles términos de interacción, así como los ajustes a realizar debido al uso de contrastes múltiples (Miguel et al., 2014).

Por lo tanto, se propone con ello un cambio en la manera de publicar, ya que, en primer lugar, los revisores de las revistas tendrían que analizar este protocolo, procedimiento y plan de análisis; y, en segundo lugar, una vez que la investigación se haya culminado, se revisarían los resultados obtenidos. Únicamente cuando se haya comprobado que el protocolo se ha seguido de manera fidedigna, se podría publicar dicha investigación según el procedimiento acordado (Munafò et al., 2017; Wicherts, 2017).

Mediante este planteamiento, se podría mejorar el proceso de investigación logrando una mayor transparencia y, por lo tanto, evitando que se lleven a cabo manipulaciones a posteriori, de manera que se podrían evitar de manera eficaz sesgos como el HARKing, el sesgo de publicación o el P-hacking (Munafò et al., 2017).

Uno de los mayores inconvenientes que señalan algunos/as autores/as a esta propuesta tiene que ver con el hecho de que se coarta el descubrimiento fortuito de nuevas relaciones no previstas a través de análisis exploratorios, lo que comúnmente se denomina serendipia. Con ello, se puede perder una característica importante en el proceso de investigación como es la creatividad. Otros/as autores/as argumentan que este inconveniente no sería tal, porque a pesar de realizar un análisis pre-registrado, los/as investigadores/as serían libres de poder realizar más análisis exploratorios, siempre que los detallen e indiquen que dichos análisis se realizaron posteriormente y que, por tanto, no faltarían a la transparencia y claridad del trabajo (Miguel et al., 2014; Munafò et al., 2017).

Otro de los inconvenientes que señalan los/as investigadores/as ante esta recomendación sería que el tiempo invertido en el proceso de pre-registro supondría hacer de la investigación un trabajo más burocratizado y que consume más tiempo. Frente a tal

inconveniente cabe argumentar que, aunque puede resultar costoso implementarlo de manera rutinaria en todas las investigaciones, debería ir implementándose paulatinamente, hasta que se incluya dentro de la cultura de la investigación, lo que a largo plazo podría producir mayores beneficios y un mayor prestigio del proceso científico (Miguel et al., 2014). Asimismo, hay que reseñar que poco a poco muchas de las revisiones sistemáticas y ensayos aleatorios controlados, ya se están pre-registrando de manera bastante generalizada.

Algunas de las plataformas o páginas web en las que se pueden pre-registrar los estudios en nuestro campo son Open Science Framework (<https://osf.io/>), AsPredicted (<https://aspredicted.org/>) y Preregistration Challenge (Center of Open Science, <https://www.cos.io/>), entre otras.

4.2.2. Recomendaciones metodológicas

En este apartado vamos a agrupar las recomendaciones que cabría denominar estrictamente metodológicas, aunque todas las recomendaciones podrían incluirse bajo esta denominación, porque todas apuestan de una manera u otra por mejorar la metodología global de las investigaciones. En este subapartado, denominamos recomendaciones metodológicas a todas las relacionadas con los procedimientos de análisis de datos, así como a los referidos al diseño de la investigación.

Se han propuesto diferentes maneras de plantear los diseños de investigación y analizar los datos para tratar de evitar los sesgos que pueden producirse en estas fases de la investigación.

Con respecto al diseño, nos encontramos con propuestas tales como la de utilizar técnicas de ciego, es decir evitar que las personas que realicen la evaluación y el análisis estadístico tengan información acerca de los diversos tratamientos, tanto a la hora de la recogida de datos, como a la hora de realizar el análisis estadístico y en otros momentos de la investigación, siempre que sea posible. De esta manera, se pretende lograr una mayor objetividad y transparencia.

Asimismo, se propone implementar comités metodológicos independientes que puedan supervisar y evaluar los diseños y asesorar adecuadamente sobre la realización de las investigaciones, evitando así sesgos como los conflictos de interés (Munafò et al., 2017).

Otra opción que se propone a nivel más general es utilizar técnicas estadísticas, principalmente en los meta-análisis, que analicen el posible sesgo de publicación y no se basen únicamente en los artículos publicados, que podrían suponer una infrarrepresentación del conocimiento real (Munafò et al., 2017). Así, se podría calcular, por ejemplo, el índice de tolerancia a los resultados nulos.

A nivel puramente de análisis estadístico, autores como Hoffmann et al. (2021) señalan la relación entre la problemática existente a la hora de replicar las investigaciones y la multitud de estrategias analíticas existentes. Tales autores afirman que, a pesar del notable esfuerzo que se está realizando en las diferentes ciencias afectadas por esta crisis para innovar mediante estrategias analíticas alternativas que la puedan solucionar, únicamente están corriendo el riesgo de reinventar lo inventado. La verdadera problemática en la que se centran estos autores es la variabilidad de los grados de libertad de los/as investigadores/as a los que nos hemos referido en el apartado 4.1. La incertidumbre a la hora de responder tanto a la pregunta inicial de investigación como a la multitud de decisiones en cuanto a la estrategia estadística es muy grande, lo que puede generar que diferentes investigadores/as tomen las decisiones de maneras muy diversas y todas ellas puedan estar justificadas. Concretamente, respecto a la incertidumbre en las diferentes fases del proceso de decisiones analíticas, Hoffmann et al. (2021) proponen soluciones respecto a la necesidad de reducir, informar, integrar y aceptar dicha incertidumbre. Para ello proponen un gráfico interactivo en el que se pueden ver las diversas soluciones a la incertidumbre de manera detallada y, además, se pueden consultar las referencias correspondientes (véase captura de dicho gráfico en el Anexo A).

Por otro lado, Wicherts (2017) y Yarkoni (2020) apuestan por adoptar mejores estándares como los que nos proporcionan extraer inferencias más conservadoras e informar sobre ellas de manera detallada basándose únicamente en los datos del estudio, sin inferir enunciados generales que puedan llevar a que estas hipótesis se interpreten erróneamente.

Además, inciden en la importancia de utilizar controles estadísticos en todo el proceso, conseguir muestras más grandes para tener una mayor potencia, intentar que las medidas sean precisas y que el enfoque de los modelos sea adecuado (Wicherts, 2017).

En cuanto a la muestra, con el fin de aumentar la potencia, se propone por un lado determinar el número de participantes en el estudio en función de cálculos de potencia a

priori; y por otro utilizar estrategias de control estadístico o diseños de medidas repetidas. Del mismo modo, se propone utilizar técnicas de muestreo donde una parte de esta se reserva para poder comprobar en ella los resultados obtenidos con la otra parte de la muestra (Wicherts, 2017).

Por último, se señala la propuesta de realizar predicciones más arriesgadas, puesto que, si estas se confirman, proporcionarían una mayor aportación a la teoría sometida a prueba (Yarkoni, 2020).

En cuanto al procedimiento a seguir a la hora de realizar las replicaciones, se propone que estas sean independientes y ajenas a los autores originales, con el fin de evitar conflicto de intereses (Wicherts, 2017). Asimismo, se recomienda proponer espacios en las revistas destinados únicamente a realizar replicaciones, propuestas que ya se están comenzando a llevar a cabo como veremos en los siguientes subapartados. De todas maneras, a la hora de realizar replicaciones cabe tener en cuenta que la relevancia del estudio original (aquel que se quiere replicar) es una característica clave a la hora de seleccionar qué estudio replicar (Yarkoni, 2020). Por ello, se propone valorar la trascendencia de las posibles replicaciones mediante el impacto que han ocasionado las publicaciones originales.

4.2.3. Mejoras en la formación sobre metodología de la investigación

En lo que concierne a las mejoras en la formación de la metodología de la investigación, se apuesta por una mayor formación que genere una cultura de investigación transparente, comenzando a nivel de formación universitaria. De esta manera, se podría concienciar de la problemática a los/as investigadores/as que están a punto de iniciar sus carreras. Además, si los estudiantes se familiarizan con la cultura de la transparencia desde el inicio de sus carreras investigadoras, será más fácil cambiar los viejos hábitos (Munafò et al., 2017; Yarkoni, 2020).

Así pues, en el ámbito de la formación se proponen mejoras concretas, como una formación continua de calidad, ya que la metodología de investigación está en constante revisión y avance, con módulos accesibles y fáciles de asimilar como los propuestos por Experimental Design Assistant para diseñar experimentos con animales (NC3Rs, <https://www.nc3rs.org.uk/our-portfolio/experimental-design-assistant-eda>).

Otra práctica que se podría realizar en la formación universitaria podría consistir en participar en proyectos de ciencia colaborativa que se llevan a cabo, por ejemplo, desde OSF o Many Labs (Munafò et al., 2017).

4.2.4. Investigación colaborativa

Con el fin de mejorar la potencia estadística de las investigaciones, implementar la diversidad de culturas de investigación y fomentar la generalización de los resultados nos encontramos con propuestas de investigación colaborativa. Estas propuestas consiguen solventar el problema de recursos limitados y consiguen una mayor potencia estadística. Del mismo modo, mejoran la representatividad, se unifican los protocolos y se obtiene una mayor precisión en los tamaños de los efectos encontrados (Munafò et al., 2017). Además, luchan contra los conflictos de interés, ya que en estos estudios se da un pre-registro para evitar intereses de financiación u otros motivos (Nosek et al., 2012).

Uno de los proyectos que actualmente consideramos más interesante en nuestra disciplina sería el acelerador de las ciencias psicológicas, dándose el avance del conocimiento en la psicología a través de una red colaborativa distribuida. En ella, se realizan investigaciones con una mayor potencia estadística y muestras más representativas. Son investigaciones colaborativas a gran escala que analizan cuestiones novedosas y que, a su vez, realizan replicaciones. Su fin es el de acelerar la acumulación de evidencia confiable y generalizable.

El acelerador, es un proyecto cuyas principales características son: investigar de manera continua, eficiente (reutilización de recursos), diversa (temas e investigadores/as diversos) e inclusiva (abierta a propuestas). En el proceso de dichas investigaciones se da un examen sistemático generalizado. Además, investigadores/as independientes o equipos se pueden implicar en las diferentes fases, en las que se cuenta con supervisores/as que velan por que el proceso se realice correctamente y en la publicación se incluyen los nombres de todos/as los/as participantes. El proceso de investigación de este tipo de colaboraciones cuenta con cuatro fases importantes: 1ª fase: presentación y evaluación; 2ª fase: preparación; 3ª fase: implementación; y 4ª fase: análisis y difusión (Moshontz et al., 2018). Es posible inscribirse gratuitamente en la página web del proyecto (<https://psysciacc.org/>) y colaborar a nivel individual o como equipo en el punto del proceso que se decida. Existen diversas investigaciones abiertas simultáneamente y se van publicando las necesidades de perfiles profesionales con las que cuentan en cada momento.

Otro de los proyectos que propone investigaciones colaborativas es el Many Labs, que está más enfocado a los proyectos relacionados con el cambio climático. Tiene como objetivo, facilitar a estudiantes y científicos/as la recopilación y análisis de datos de valores ambientales del mundo físico. A través de él se proporcionan herramientas tecnológicas para los experimentos, tales como herramientas de monitoreo de carbono o de sustancias biológicas. Todo ello estaría en la línea de flujo de datos abiertos como comentaremos posteriormente (<http://www.manylabs.org/>).

4.2.5. Unificación de criterios de presentación de informes, publicación y métodos de difusión

Tal y como se ha señalado previamente, una de las causas de esta crisis es la falta de transparencia y de unificación de criterios a la hora de realizar los informes y las publicaciones, no especificando siempre la información importante de la investigación para poder replicarla. En esta ocasión, las propuestas de mejora van en la línea de que los informes o artículos se hagan de manera sistemática con la misma estructura, siguiendo diferentes pautas estandarizadas y protocolos generalizados como son CONSORT (Consolidated Standards of Reporting Trials), que consta de una guía de 22 puntos y un diagrama de flujo a seguir para poder redactar informes y mejorar la calidad de la comunicación (Schulz et al., 2010; véase Anexo B); las pautas STROBE (Strengthening the Reporting of Observational studies in Epidemiology), con recomendaciones para la transmisión de la información de estudios observacionales (Von Elm et al., 2008; véase Anexo C); el protocolo PRISMA (Elementos de informes preferidos para revisiones sistemáticas y meta-análisis), para determinar la integridad y transparencia de la revisión sistemática y meta-análisis a través de la consecución de 27 ítems (Urrútia y Bonfill, 2010; véase Anexo D); los JARS (Journal Article Reporting Standards) donde se encuentran pautas para estudios de diseño cualitativo, cuantitativo y mixto (<https://www.lluiscodina.com/jars-apa/>); las pautas TREND (Transparent Reporting of Evaluations with Nonrandomized Designs), para estudios con metodología cuasi-experimental (Des Jarlais et al., 2004; véase Anexo E); el aula virtual con recursos CASPE (Critical Appraisal Skills Programme Español), para proporcionar guías y metodologías para la lectura crítica de los estudios (<https://aulavirtual.redcaspe.org/moodle/>); o las directrices TOP (Transparency and Openness Promotion), para mejorar la transparencia a la hora de publicar por parte de las revistas (Nosek et al., 2016; véase Anexo F).

Otra de las propuestas, de fácil implementación y que, sin embargo, podría contribuir en gran medida a la transparencia a la hora de comunicar los resultados sería que, tanto los/as autores/as como los/as revisores/as y los/as editores/as, utilizaran listas de verificación para comprobar si aparecen reflejados datos tan importantes como la integración de los tamaños del efecto en los resultados, los procedimientos de análisis de datos que se han utilizado, la justificación de los valores p utilizados, entre otros (Nosek et al., 2012). A día de hoy, nos encontramos que son muchas las revistas que ponen a disposición de autores/as y revisores/as este tipo de listas de verificación (véase Anexo G para un ejemplo).

En cuanto a los criterios a la hora de la publicación y con el fin de abordar el sesgo de publicación, se plantea la opción de publicar el estudio sin saber los resultados, conociendo únicamente el diseño de la investigación y su justificación. Una vez que el estudio es aceptado y posteriormente publicado con sus resultados, la comunidad científica puede realizar feedbacks sobre él y decidir abiertamente si este estudio realiza aportaciones trascendentales para el avance del conocimiento o si se detecta algún tipo de QRP para ser revisado (Munafò et al., 2017). En este caso el reconocimiento a los/as investigadores/as no se obtendría al publicar en una revista de alto impacto, sino al ser aprobado por la comunidad científica.

4.2.6. Flujo de datos, materiales y condiciones experimentales abiertas

Una de las soluciones que se han propuesto consiste en la apertura de los datos, los materiales y las condiciones experimentales mediante repositorios públicos. El poner estos elementos a disposición de la comunidad científica favorecería la replicación y la detección de errores (Miguel et al., 2014), además de la transparencia y la accesibilidad a los datos y materiales, fomentando tanto el acceso abierto como el flujo de trabajo abierto (Munafò et al., 2017).

La mayor propuesta para la apertura de datos y a favor de la ciencia abierta se da a través de la organización sin ánimo de lucro Center for Open Science (COS). Esta organización tiene como misión la reestructuración de la cultura actual de investigación hacia una cultura con más valores de transparencia y buenas prácticas investigadoras, proporcionando recursos y lugares en los que esto se hace posible (<https://www.cos.io/>). Sus inicios se dan tras la necesidad de fomentar la replicabilidad en psicología y su principal proyecto es el Open Science Framework (OSF, <https://osf.io/>), una aplicación web a la que poder subir las publicaciones, sus datos, métodos y otros archivos, para

favorecer la replicación y la ciencia abierta. Adicionalmente, esta aplicación web también ofrece la posibilidad de publicar los llamados *preprints*, concretamente dentro de OSF el repositorio para las ciencias psicológicas se llama Psyarxiv (<https://psyarxiv.com/>). Los preprints o las prepublicaciones son las versiones de un artículo científico previas al proceso de revisión por pares y publicación por parte de una revista. Las prepublicaciones permiten por un lado estar al tanto de las investigaciones y datos actuales, aun no estando publicados oficialmente en ese momento, lo que agiliza el proceso de divulgación científica; y, por otro lado, abren una oportunidad para debatir sobre el contenido de la publicación de manera abierta y dinámica, ya que generalmente los sitios web que albergan estas prepublicaciones suelen disponer de foros o mecanismos para fomentar el intercambio de opiniones.

En este mismo sentido, cada vez existen más iniciativas para compartir los datos como el Dataverse Network Project de Harvard (<https://psr.iq.harvard.edu/the-dataverse-network>), donde se pueden compartir libremente datos cuantitativos de las investigaciones, y DRYAD (<https://datadryad.org/stash/>), donde se pueden compartir los datos, así como realizar comentarios para su avance, revisión y mejora. En esta apertura de los datos siempre hay que tener en cuenta excepciones como la confidencialidad, las barreras legales (derechos de autor) y la apertura de bases de datos masivas donde se encuentran datos de muchos estudios y algunos de ellos aún no han sido publicados (Nosek et al., 2012).

En la Tabla 2 se presenta una síntesis de las recomendaciones presentadas en este apartado.

Tabla 2

Resumen de las recomendaciones

Bloque	Recomendaciones	Recursos
Pre-registro y planes de análisis	-Pre-registro de los estudios con sus características principales (diseño, plan de análisis, resultados principales, procedimiento...).	-Open Science Framework, https://osf.io/ -As Predicted, https://aspredicted.org/ -Center of Open Science, https://www.cos.io/
Recomendaciones metodológicas	-Reducir, informar, integrar y aceptar la incertidumbre en las decisiones estadísticas. -Técnicas de ciego.	https://shiny.psy.lmu.de/multiplicity/index.html

	<ul style="list-style-type: none"> -Comités metodológicos independientes. -Técnicas que analicen el sesgo de publicación (meta-análisis). -Estándares más rigurosos. -Controles estadísticos. -Muestras más grandes. -Medidas precisas y enfoque adecuado. -Diseños de medidas repetidas. -Métodos de remuestreo o reserva de parte de la muestra. -Predicciones más arriesgadas. -Replicaciones independientes. -Espacios para las replicaciones. 	
Mejoras en la formación sobre metodología de la investigación	-Formación en investigación transparente.	-NC3Rs, https://www.nc3rs.org.uk/our-portfolio/experimental-design-assistant-eda
Investigación colaborativa	-Investigación colaborativa entre diferentes equipos de investigación.	-Acelerador de las ciencias psicológicas, https://psysciacc.org/ -Many Labs, http://www.manylabs.org/
Unificación de criterios de presentación de informes, publicación y métodos de difusión	-Estandarización de informes y publicaciones.	-Estándares CONSORT -Pautas STROBE (estudios observacionales) -Protocolo PRISMA (revisión sistemática y meta-análisis) -Directrices TOP -JARS -Pautas TREND -Aula virtual CARPE
	-Listas de verificación de los datos más relevantes. -Publicación de los estudios sin resultados.	
Flujo de datos, materiales y condiciones experimentales abiertas	-Apertura de los datos de la investigación.	-Center for Open Science, https://www.cos.io/ -OSF, https://osf.io/ -Harvard University, https://psr.iq.harvard.edu/the_dataverse_network -DRYAD, https://datadryad.org/stash/
	-Publicación de los preprints.	-Psyarxiv, https://psyarxiv.com/

5. Conclusiones

El primer objetivo de este trabajo ha sido revisar los factores que afectan a que de una manera u otra nos encontremos dentro de la crisis de la replicación, la cual pone en tela de juicio parte del conocimiento adquirido en el ámbito de la Psicología. Por otro lado, el segundo objetivo de este trabajo ha sido recoger las propuestas más relevantes para paliar esta crisis. El hecho de que paulatinamente se vayan implementando algunas de las propuestas que se han realizado en la última década para mejorar tanto la calidad

como la transparencia de la investigación en nuestro ámbito, hace que tengamos una mayor esperanza en el futuro. Sin embargo, cabe señalar que hay algunas conductas cuestionables que aún se siguen produciendo y, por lo tanto, habría que seguir implementando las propuestas o generando nuevas que traten de mejorar esta situación. Como ya se ha mencionado anteriormente, actuar contra este tipo de prácticas ayudará a que la literatura científica sea cada vez más transparente y de mayor calidad metodológica, lo que propiciará entre otras cosas, una mayor identificación de los tratamientos basados en la evidencia, dándoles un mayor peso y credibilidad.

Concretamente, en este trabajo se han identificado una serie de factores de diversa índole, que pueden darse a lo largo de todo el proceso de investigación. Entre los factores identificados podemos citar, diferentes sesgos como el de confirmación, apofenia, selección y esencias ilusorias, que pueden propiciar que las creencias preestablecidas de los/as investigadores/as puedan influir en la selección de la información, planteamiento de las hipótesis, selección de la muestra y uso de instrumentos, respectivamente. Además, cuando los resultados encontrados no son acordes con las hipótesis planteadas se puede incurrir en otro tipo de prácticas cuestionables como el HARKing, adaptando la hipótesis inicial a los resultados para que se verifique dicha hipótesis o el P-Hacking, dragado de datos y expediciones de pesca o minería de datos, analizando diferentes relaciones dentro de los datos existentes para poder presentar resultados estadísticamente significativos. En cuanto a la última fase de la investigación, relacionada con la publicación y difusión, encontramos aún una falta de estandarización en la realización de informes y un sesgo característico, como es el de publicación que produce una interferencia a la hora de conocer muchas de las investigaciones que se han realizado, que generalmente cuentan con hipótesis principales refutadas. Por ello, se incurre en prácticas como el Cherry picking, donde a la hora de publicar los resultados los/as investigadores/as visibilizan únicamente los resultados estadísticamente significativos.

En la línea del segundo objetivo que se ha planteado este trabajo, hemos recogido diferentes recomendaciones como la de la apertura de los datos, el pre-registro de las investigaciones, diversas recomendaciones metodológicas, mejoras en la formación, un fomento de la investigación colaborativa y la unificación de criterios de investigación. Algunas de estas recomendaciones ya se aplican en mayor o menor medida en algunos de los ámbitos de la psicología. Por ejemplo, hoy en día la mayoría de los artículos meta-analíticos y los que analizan ensayos clínicos suelen ser pre-registrados en repositorios

específicos. Del mismo modo, cada vez más revistas ponen a disposición de autores/as y revisores/as pautas concretas con el objetivo de unificar los apartados o datos mínimos que se han de indicar en los estudios que se envían a publicación. Sin embargo, consideramos que otras recomendaciones como el fomento de la investigación colaborativa y la apertura de los datos no son tan ampliamente respaldados hoy en día por la comunidad científica.

En relación a la implementación de dichas recomendaciones, a continuación, se plantean una serie de ideas, cambios o propuestas para reflexionar sobre cómo incentivar estos aspectos (el pre-registro, los datos en abierto, etc...) que hoy en día puede que no tengan la valoración suficiente. Esta tarea no depende únicamente de los/as investigadores/as, sino que existen muchos agentes implicados como revisores/as, directores/as y editores/as de revistas, instituciones como universidades y centros de investigación, políticos/as... por lo que sería importante que las distintas instituciones implicadas tuvieran una buena coordinación para lograr fomentar dichas actitudes (Munafò et al., 2017). Tales propuestas e ideas, podrían servir para hacer reflexionar y concienciar en cierta manera a la comunidad científica, ya que parte de esta es partícipe en los procesos de decisión y revisión de las revistas, siendo parte activa de tales propuestas.

Una de las propuestas, consiste en que las revistas puedan fomentar con incentivos de diversa índole ciertas prácticas transparentes de las investigaciones, como el pre-registro y la apertura de datos (Munafò et al., 2017), preferentemente, en repositorios públicos. Por ejemplo, podría ser que se reservaran un número de artículos de cada volumen para estudios con pre-registros, o que tal y como se ha mencionado anteriormente, que hubiera un espacio específico para estudios de replicación. Además, se plantea que sí las instituciones que financian las investigaciones utilizaran estas recomendaciones como requisito, esto podría originar una cierta cultura en los/as investigadores/as para poder incluir el pre-registro y la apertura de datos como un paso más dentro del proceso de investigación. Es importante indicar que algunas revistas ya están incentivando actualmente este tipo de prácticas. Un caso concreto en nuestra disciplina sería la revista *Psychological Science*, donde dichas prácticas se incentivan mostrando símbolos asociados a las publicaciones, como, por ejemplo, los símbolos de publicación pre-registrada, datos abiertos, materiales abiertos... Del mismo modo, cada vez más las

instituciones públicas que subvencionan los proyectos de investigación exigen que los estudios que se realicen con financiación pública sean publicados en abierto.

Cabe destacar, que además de los factores anteriormente expuestos, en esta crisis subyacen factores contextuales que también la alimentan. Estos factores están relacionados con el sistema de incentivos a través de las publicaciones dentro del mundo de la investigación y la ciencia (García-Garzón et al., 2018; Nosek et al., 2012; Yarkoni, 2020). Este sistema, se basa en gran medida en premiar a nivel laboral a los/as investigadores/as en función del número de publicaciones en revistas de prestigio y es, actualmente, la vía fundamental para poder desarrollar una carrera profesional dentro del ámbito de la investigación. Se puede pensar que dicho sistema podría ser una de las razones por las que existen estas prácticas cuestionables, ya que hace que la necesidad de publicar sea real. Encontramos, además, que se premia en mayor medida la novedad de los temas en las publicaciones, siendo más atractivas para las revistas los estudios originales sobre temas novedosos que las replicaciones (García-Garzón et al., 2018). Si bien es cierto que existen vías de promoción de la carrera investigadora como son los sexenios en el ámbito nacional, que no se basan únicamente en el número de publicaciones, la mayoría de procesos de promoción a puestos y vías de financiación sí funciona a través de la cantidad de publicaciones, entre sus principales requisitos.

Dada esta situación, podemos llegar a entender, en parte, el origen de estas prácticas cuestionables que pueden surgir tras las necesidades a las que se enfrentan los/as investigadores/as en el sistema actual. Por ello, se propone la necesidad de un cambio global en la manera de evaluar y valorar la carrera investigadora. En la actualidad existe un sistema que fomenta la publicación de los resultados científicos, aspecto positivo en sí mismo, ya que conocer lo que investigan los demás, sus resultados y conclusiones es indispensable para evitar la duplicidad de proyectos y esfuerzos y hacer que la ciencia avance. Sin embargo, el problema surge cuando la publicación deja de ser un instrumento al servicio de la investigación y se convierte en un objetivo en sí misma (Aznar y Tudela, 2013). En este aspecto, es difícil encontrar una respuesta clara para este cambio global, puesto que surgen ideas contradictorias sobre la necesidad de publicar toda investigación realizada, aunque algunas de ellas no cuenten con trascendencia en el ámbito científico con el fin de evitar el sesgo de publicación o la realización futura de esas investigaciones. Mientras que, por otro lado, se puede pensar que lo ideal sería publicar únicamente lo que se considera que puede aportar al conocimiento científico y así no generar infinidad de

publicaciones difíciles de integrar, ya que unas pueden desvirtuar a las otras. Por ello, se muestra fundamental tanto a nivel de investigadores/as como del sistema, fomentar la autocrítica continua para la búsqueda del equilibrio entre valorar el número de publicaciones y la relevancia de estas.

En síntesis, tras vislumbrar la complejidad que entraña la crisis de la replicación en la que se encuentra inmersa la psicología junto con otras ciencias sociales y de la salud, y dada la confluencia e interrelación de numerosos factores y agentes involucrados en ella, se proponen recomendaciones que poco a poco se van extendiendo dentro de la comunidad científica y van creando una nueva cultura de investigación que aboga por una mayor transparencia. Dichos avances hacen que alberguemos cierto optimismo hacía el abordaje de esta crisis, donde podemos encontrar como primer hito la propia conciencia de esta crisis en nuestra ciencia y segundo, las propuestas que se van integrando dentro de la comunidad científica gracias a los diferentes autores/as involucrados en su análisis y su esfuerzo por la búsqueda de soluciones y su difusión.

6. Referencias Bibliográficas

- Andrade, C. (2021). HARKing, cherry-picking, P-hacking, fishing expeditions, and data dredging and mining as questionable research practices. *The Journal of Clinical Psychiatry*, 82(1) 20f13804. <https://doi.org/10.4088/JCP.20f13804>
- Arkes, H. R., Faust, D., Guilmette, T. J. y Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, 73(2), 305-307. <https://doi.org/10.1037/0021-9010.73.2.305>
- Attorresi, H. F., García Díaz, A. M. y Pralong, H. O. (2006). *Identificación de la Falacia del Jugador en una situación típica de juego de azar*. XIII Jornadas de Investigación y Segundo Encuentro de Investigadores en Psicología del Mercosur, Buenos Aires, Argentina. <https://www.aacademica.org/000-039/100>
- Balluerka, N. (2011). *Planificación de la investigación. La validez del diseño (2ª ed. corregida)*. Amarú.
- Bargh, J. A., Chen, M. y Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71(2), 230-244. <https://doi.org/10.1037//0022-3514.71.2.230>
- Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100(3), 407-425. <https://doi.org/10.1037/a0021524>
- Bretz, F., Maurer, W. y Xi, D. (2019). Replicability, reproducibility, and multiplicity in drug development. *Chance*, 32(4), 4-11. <https://doi.org/10.1080/09332480.2019.1695432>
- Blomquist, K. B. (1986). Replication of research. *Research in Nursing & Health*, 9(3), 193-194. <https://doi.org/10.1002/NUR.4770090302>
- Brick, C., Hood, B., Ekroll, V. y De-Wit, L. (2022). Illusory essences: A bias holding back theorizing in psychological science. *Perspectives on Psychological Science*. 17(2) 491–506. <https://doi.org/10.1177/1745691621991838>
- Buetow, S. (2019). Apophenia, unconscious bias and reflexivity in nursing qualitative research. *International Journal of Nursing Studies*, 89, 8-13. <https://doi.org/10.1016/j.ijnurstu.2018.09.013>
- Callaway, E. (2011). Report finds massive fraud at Dutch universities. *Nature*, 479, 15.

- Carey, B. (27 de agosto de 2015). Many psychology findings not as strong as claimed, study says. *The New York Times*.
<http://www.nytimes.com/2015/08/28/science/many-social-science-findings-not-as-strong-asclaimed-study-says.html>
- Castro Fernández, J. J. (2002). *Métodos, diseños y técnicas de investigación psicológicos*. Servicio de Publicaciones y Difusión Científica de la ULPGC.
- Des Jarlais, D. C., Lyles, C., Crepaz, N. y Trend Group. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *American Journal of Public Health*, 94(3), 361-366.
- Doyen, S., Klein, O., Pichon, C. L. y Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE*, 7(1), e29081.
<https://doi.org/10.1371/journal.pone.0029081>
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4 (5), e5738.
<https://doi.org/10.1371/journal.pone.0005738>
- Ferguson, C. J. y Brannick, M. T. (2012). Publication bias in psychological science: prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, 17(1), 120-128.
<https://doi.org/10.1037/a0024445>
- Fernández Liria, C., García Fernández, O. y Galindo Ferrández, E. (2017). *Escuela o barbarie: entre el neoliberalismo salvaje y el delirio de la izquierda*. Ediciones Akal.
- Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, 19, 151–156.
<http://dx.doi.org/10.3758/s13423-012-0227-9>
- García-Garzón, E., Lecuona, O. y Carbajal, G. V. (2018). Estudios de replicación, pre-registros y ciencia abierta en Psicología. *Apuntes de Psicología*, 36(1-2), 75-83.
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33(5), 587–606. <https://doi.org/10.1016/j.socec.2004.09.033>
- Glick, M. (2017). Creer es ver: sesgo de confirmación. *El Diario de la Asociación Dental Americana*, 148(3), 131-132. <https://doi.org/10.1016/j.adaj.2017.01.009>
- Gopalakrishna, G., Ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M. y Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and

- their potential explanatory factors: A survey among academic researchers in The Netherlands. *PLoS One*, 17(2), e0263023. <https://doi.org/10.1371/journal.pone.0263023>
- Hoffmann, S., Schönbrodt, F., Elsas, R., Wilson, R., Strasser, U. y Boulesteix, A. L. (2021). The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. *Royal Society Open Science*, 8(4), 201925. <https://doi.org/10.1098/rsos.201925>
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Kerlinger, F. N., Lee, H. B., Pineda, L. E. y Mora Magaña, I. (2002). *Investigación del Comportamiento*. McGraw-Hill Interamericana.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196-217. https://doi.org/10.1207/s15327957pspr0203_4
- Losada, J. L. y López-Feal Ramil, R. (2003). *Métodos de investigación en ciencias sociales y humanas*. Paraninfo Cengage Learning.
- Makel, M. C., Plucker, J. A. y Hegarty, B. (2012). Replications in psychology research: How often do they really occur?. *Perspectives on Psychological Science*, 7(6), 537-542. <https://doi.org/10.1177/1745691612460688>
- 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. y Van der Laan, M. (2014). Promoting transparency in social science research. *Science*, 343(6166), 30-31. <https://doi.org/10.1126/science.1245317>
- Méndez Pardo, L. F. y Pérez-Acosta, A. M. (2011). Cuando la divulgación ayuda a la investigación científica: Reflexión sobre un caso sui generis. *Revista Latinoamericana de Psicología*, 43(2), 369-378. <https://doi.org/10.14349/RLP.V43I2.729>
- Montserrat, J. (1983). *Epistemología evolutiva y teoría de la ciencia*. Universidad Pontificia de Comillas.
- Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B.,... y Chartier, C. R. (2018). The Psychological Science

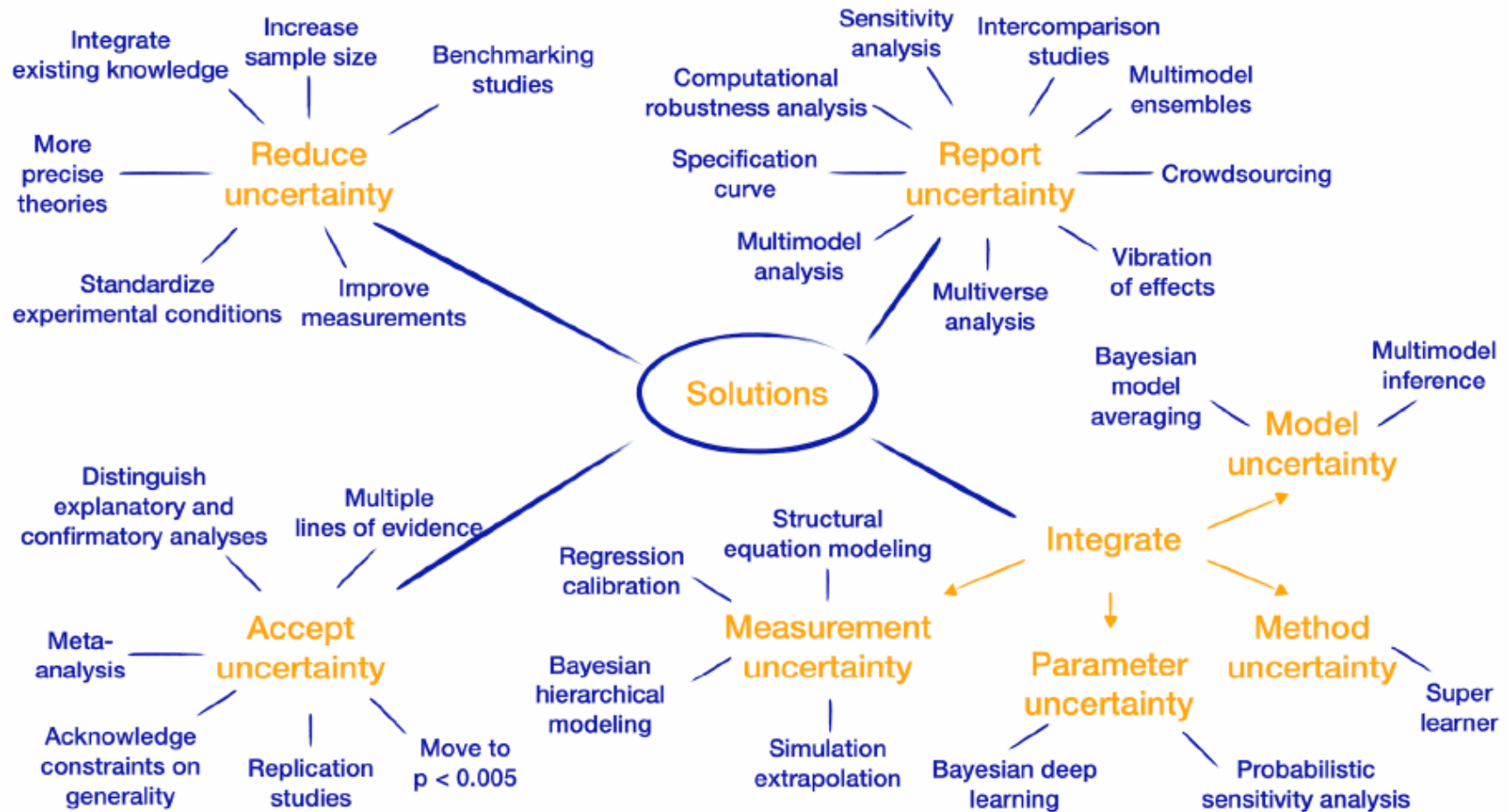
- Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, 1(4), 501-515. <https://doi.org/10.1177/2515245918797607>
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E., Ware, J. J. y Ioannidis, J. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 1-9. <https://doi.org/10.1038/s41562-016-0021>
- Nelson, L. D., Simmons, J. y Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, 69(1), 511–534. <https://doi.org/10.1146/annurev-psych-122216-011836>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175-220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nosek, B. A., Spies, J. R. y Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6), 615-631. <https://doi.org/10.1177/1745691612459058>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S., Breckler, S., Buck, S., Chambers, C., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Fich, E., Freese, J., Glennester, R., Goroff, D., Green, D., Hesse, B.,... DeHaven, A. C. (2016). *Transparency and openness promotion (TOP) guidelines*. <https://doi.org/10.31219/osf.io/vj54c>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D. y Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719-748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Padmanabhan, S. (Ed.). (2014). *Handbook of pharmacogenomics and stratified medicine*. Academic Press.
- Pashler, H. y Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528-530.

<https://doi.org/10.1177/1745691612465253>

- Popper, K. R. (1990). *La lógica de la investigación científica*. Tecnos.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638-641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, 17(4), 551-566. <https://doi.org/10.1037/a0029487>
- Schmidt, S. (2016). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90-100. <https://doi.org/10.1037/a0015108>
- Schulz, K. F., Altman, D. G. y Moher, D. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Trials*, 11(1), 1-8. <https://doi.org/10.1136/bmj.c332>
- Simmons, J. P., Nelson, L. D. y Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366. <https://doi.org/10.1177/0956797611417632>
- Stroebe, W. y Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9, 59-71. <https://doi.org/10.1177/1745691613514450>
- Tudela, J. y Aznar, J. (2013). ¿Publicar o morir? El fraude en la investigación y las publicaciones científicas. *Persona y Bioética*, 17(1), 12-27. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-31222013000100002
- Urrútia, G. y Bonfill, X. (2010). Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Medicina Clínica*, 135(11), 507- 511. <https://doi.org/10.1016/j.medcli.2010.01.015>
- Valentine, K. D., Buchanan, E. M., Cunningham, A., Hopke, T., Wikowsky, A. y Wilson, H. (2021). Have psychologists increased reporting of outliers in response to the reproducibility crisis? *Social and Personality Psychology Compass*, 15(5), e12591. <https://doi.org/10.1111/spc3.12591>
- Valero, L. (2015). *La replicación como base de la ciencia. ¿Qué pasa en Psicología?* [Conferencia]: IV Congreso SAVECC, Córdoba, España.

- Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C. y Vandembroucke, J. P. (2008). Declaración de la Iniciativa STROBE (Strengthening the Reporting of Observational studies in Epidemiology): directrices para la comunicación de estudios observacionales. *Revista Española de Salud Pública*, 82, 251-259. <https://doi.org/10.1157/13119325>
- Wicherts, J. M. (2017). The weak spots in contemporary science (and how to fix them). *Animals*, 7(12), 90. <https://doi.org/10.3390/ani7120090>
- Xie, Y., Wang, K. y Kong, Y. (2021). Prevalence of research misconduct and questionable research practices: A systematic review and meta-analysis. *Science and Engineering Ethics*, 27(4), 1-28. <https://doi.org/10.1007/s11948-021-00314-9>
- Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 1-37. <https://doi.org/10.1017/S0140525X20001685>

Anexo A. Gráfico Interactivo (Hoffmann et al., 2021)



Anexo B. CONSORT (Schulz et al., 2010)

Section/Topic	Item No	Checklist Item
Title and abstract	1a	Identification as a randomised trial in the title
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts ^{21,36})
Introduction		
Background and objectives	2a	Scientific background and explanation of rationale
	2b	Specific objectives or hypotheses
Methods		
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons
Participants	4a	Eligibility criteria for participants
	4b	Settings and locations where the data were collected
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed
	6b	Any changes to trial outcomes after the trial commenced, with reasons
Sample size	7a	How sample size was determined
	7b	When applicable, explanation of any interim analyses and stopping guidelines
Randomisation:		
Sequence generation	8a	Method used to generate the random allocation sequence
	8b	Type of randomisation; details of any restriction (such as blocking and block size)
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how
	11b	If relevant, description of the similarity of interventions
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses
Results		
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome
	13b	For each group, losses and exclusions after randomisation, together with reasons
Recruitment	14a	Dates defining the periods of recruitment and follow-up
	14b	Why the trial ended or was stopped
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms ²⁸)
Discussion		
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses
Generalisability	21	Generalisability (external validity, applicability) of the trial findings
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence
Other information		
Registration	23	Registration number and name of trial registry
Protocol	24	Where the full trial protocol can be accessed, if available
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders

²¹We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration³⁷ for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials,³⁸ non-inferiority and equivalence trials,³⁹ non-pharmacological treatments,⁴⁰ herbal interventions,⁴¹ and pragmatic trials.⁴² Additional extensions are forthcoming for those and for up to date references relevant to this checklist, see www.consort-statement.org.

Anexo C. STROBE (STROBE Statement, 2008)¹

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses

¹ STROBE Statement (2008). STROBE statement--checklist of items that should be included in reports of observational studies (STROBE initiative). *International Journal of Public Health*, 53(1), 3-4. <https://doi.org/10.1007/s00038-007-0239-9>

Results

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses

Discussion

Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results

Other information


Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based
---------	----	---

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Anexo D. Declaración PRISMA (Urrútia y Bonfill, 2010)

Sección/tema	Número	Ítem
Título		
Título	1	Identificar la publicación como revisión sistemática, metaanálisis o ambos
Resumen		
Resumen estructurado	2	Facilitar un resumen estructurado que incluya, según corresponda: antecedentes; objetivos; fuente de los datos; criterios de elegibilidad de los estudios, participantes e intervenciones; evaluación de los estudios y métodos de síntesis; resultados; limitaciones; conclusiones e implicaciones de los hallazgos principales; número de registro de la revisión sistemática
Introducción		
Justificación	3	Describir la justificación de la revisión en el contexto de lo que ya se conoce sobre el tema
Objetivos	4	Plantear de forma explícita las preguntas que se desea contestar en relación con los participantes, las intervenciones, las comparaciones, los resultados y el diseño de los estudios (PICOS)*
Métodos		
Protocolo y registro	5	Indicar si existe un protocolo de revisión al que se pueda acceder (por ej., dirección web) y, si está disponible, la información sobre el registro, incluyendo su número de registro
Criterios de elegibilidad	6	Especificar las características de los estudios (por ej., PICOS, duración del seguimiento) y de las características (por ej., años abarcados, idiomas o estatus de publicación) utilizadas como criterios de elegibilidad y su justificación
Fuentes de información	7	Describir todas las fuentes de información (por ej., bases de datos y períodos de búsqueda, contacto con los autores para identificar estudios adicionales, etc.) en la búsqueda y la fecha de la última búsqueda realizada
Búsqueda	8	Presentar la estrategia completa de búsqueda electrónica en, al menos, una base de datos, incluyendo los límites utilizados, de tal forma que pueda ser reproducible
Selección de los estudios	9	Especificar el proceso de selección de los estudios (por ej., el cribado y la elegibilidad incluidos en la revisión sistemática y, cuando sea pertinente, incluidos en el metaanálisis)
Proceso de extracción de datos	10	Describir los métodos para la extracción de datos de las publicaciones (por ej., formularios pilotado, por duplicado y de forma independiente) y cualquier proceso para obtener y confirmar datos por parte de los investigadores
Lista de datos	11	Listar y definir todas las variables para las que se buscaron datos (por ej., PICOS, fuente de financiación) y cualquier asunción y simplificación que se hayan hecho
Riesgo de sesgo en los estudios individuales	12	Describir los métodos utilizados para evaluar el riesgo de sesgo en los estudios individuales (especificar si se realizó al nivel de los estudios o de los resultados) y cómo esta información se ha utilizado en la síntesis de datos
Medidas de resumen	13	Especificar las principales medidas de resumen (por ej., razón de riesgos o diferencia de medias)
Síntesis de resultados	14	Describir los métodos para manejar los datos y combinar resultados de los estudios, cuando esto es posible, incluyendo medidas de consistencia (por ej., ítem 2) para cada metaanálisis
Riesgo de sesgo entre los estudios	15	Especificar cualquier evaluación del riesgo de sesgo que pueda afectar la evidencia acumulativa (por ej., sesgo de publicación o comunicación selectiva)
Análisis adicionales	16	Describir los métodos adicionales de análisis (por ej., análisis de sensibilidad o de subgrupos, metarregresión), en el caso de que se hiciera, indicar cuáles fueron preespecificados
Resultados		
Selección de estudios	17	Facilitar el número de estudios cribados, evaluados para su elegibilidad e incluidos en la revisión, y detallar las razones para su exclusión en cada etapa, idealmente mediante un diagrama de flujo
Características de los estudios	18	Para cada estudio presentar las características para las que se extrajeron los datos (por ej., tamaño, PICOS y duración del seguimiento) y proporcionar las citas bibliográficas
Riesgo de sesgo en los estudios	19	Presentar datos sobre el riesgo de sesgo en cada estudio y, si está disponible, cualquier evaluación del sesgo en los resultados (ver ítem 12)
Resultados de los estudios individuales	20	Para cada resultado considerado en cada estudio (beneficios o daños), presentar: a) el dato resumen para cada grupo de intervención y b) la estimación del efecto con su intervalo de confianza, idealmente de forma gráfica mediante un diagrama de bosque (<i>forest plot</i>)
Síntesis de los resultados	21	Presentar los resultados de todos los metaanálisis realizados, incluyendo los intervalos de confianza y las medidas de consistencia
Riesgo de sesgo entre los estudios	22	Presentar los resultados de cualquier evaluación del riesgo de sesgo entre los estudios (ver ítem 15)
Análisis adicionales	23	Facilitar los resultados de cualquier análisis adicional, en el caso de que se hayan realizado (por ej., análisis de sensibilidad o de subgrupos, metarregresión [ver ítem 16])
Discusión		
Resumen de la evidencia	24	Resumir los hallazgos principales, incluyendo la fortaleza de las evidencias para cada resultado principal; considerar su relevancia para grupos clave (por ej., proveedores de cuidados, usuarios y decisores en salud)
Limitaciones	25	Discutir las limitaciones de los estudios y de los resultados (por ej., riesgo de sesgo) y de la revisión (por ej., obtención incompleta de los estudios identificados o comunicación selectiva)
Conclusiones	26	Proporcionar una interpretación general de los resultados en el contexto de otras evidencias, así como las implicaciones para la futura investigación
Financiación		
Financiación	27	Describir las fuentes de financiación de la revisión sistemática y otro tipo de apoyos (por ej., aporte de los datos), así como el rol de los financiadores en la revisión sistemática

Anexo E. Pautas TREND (Des Jarlais et al., 2004)

Paper Section/ Topic	Item No	Descriptor	Reported?	
				Pg #
Title and Abstract				
Title and Abstract	1	<ul style="list-style-type: none"> • Information on how unit were allocated to interventions • Structured abstract recommended • Information on target population or study sample 		
Introduction				
Background	2	<ul style="list-style-type: none"> • Scientific background and explanation of rationale • Theories used in designing behavioral interventions 		
Methods				
Participants	3	<ul style="list-style-type: none"> • Eligibility criteria for participants, including criteria at different levels in recruitment/sampling plan (e.g., cities, clinics, subjects) • Method of recruitment (e.g., referral, self-selection), including the sampling method if a systematic sampling plan was implemented • Recruitment setting • Settings and locations where the data were collected 		
Interventions	4	<ul style="list-style-type: none"> • Details of the interventions intended for each study condition and how and when they were actually administered, specifically including: <ul style="list-style-type: none"> ○ Content: what was given? ○ Delivery method: how was the content given? ○ Unit of delivery: how were the subjects grouped during delivery? ○ Deliverer: who delivered the intervention? ○ Setting: where was the intervention delivered? ○ Exposure quantity and duration: how many sessions or episodes or events were intended to be delivered? How long were they intended to last? ○ Time span: how long was it intended to take to deliver the intervention to each unit? ○ Activities to increase compliance or adherence (e.g., incentives) 		
Objectives	5	<ul style="list-style-type: none"> • Specific objectives and hypotheses 		
Outcomes	6	<ul style="list-style-type: none"> • Clearly defined primary and secondary outcome measures • Methods used to collect data and any methods used to enhance the quality of measurements • Information on validated instruments such as psychometric and biometric properties 		
Sample Size	7	<ul style="list-style-type: none"> • How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules 		
Assignment Method	8	<ul style="list-style-type: none"> • Unit of assignment (the unit being assigned to study condition, e.g., individual, group, community) • Method used to assign units to study conditions, including details of any restriction (e.g., blocking, stratification, minimization) • Inclusion of aspects employed to help minimize potential bias induced due to non-randomization (e.g., matching) 		

Blinding (masking)	9	<ul style="list-style-type: none"> Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to study condition assignment; if so, statement regarding how the blinding was accomplished and how it was assessed. 		
Unit of Analysis	10	<ul style="list-style-type: none"> Description of the smallest unit that is being analyzed to assess intervention effects (e.g., individual, group, or community) If the unit of analysis differs from the unit of assignment, the analytical method used to account for this (e.g., adjusting the standard error estimates by the design effect or using multilevel analysis) 		
Statistical Methods	11	<ul style="list-style-type: none"> Statistical methods used to compare study groups for primary methods outcome(s), including complex methods of correlated data Statistical methods used for additional analyses, such as a subgroup analyses and adjusted analysis Methods for imputing missing data, if used Statistical software or programs used 		
Results				
Participant flow	12	<ul style="list-style-type: none"> Flow of participants through each stage of the study: enrollment, assignment, allocation, and intervention exposure, follow-up, analysis (a diagram is strongly recommended) <ul style="list-style-type: none"> Enrollment: the numbers of participants screened for eligibility, found to be eligible or not eligible, declined to be enrolled, and enrolled in the study Assignment: the numbers of participants assigned to a study condition Allocation and intervention exposure: the number of participants assigned to each study condition and the number of participants who received each intervention Follow-up: the number of participants who completed the follow-up or did not complete the follow-up (i.e., lost to follow-up), by study condition Analysis: the number of participants included in or excluded from the main analysis, by study condition Description of protocol deviations from study as planned, along with reasons 		
Recruitment	13	<ul style="list-style-type: none"> Dates defining the periods of recruitment and follow-up 		
Baseline Data	14	<ul style="list-style-type: none"> Baseline demographic and clinical characteristics of participants in each study condition Baseline characteristics for each study condition relevant to specific disease prevention research Baseline comparisons of those lost to follow-up and those retained, overall and by study condition Comparison between study population at baseline and target population of interest 		
Baseline equivalence	15	<ul style="list-style-type: none"> Data on study group equivalence at baseline and statistical methods used to control for baseline differences 		

Numbers analyzed	16	<ul style="list-style-type: none"> Number of participants (denominator) included in each analysis for each study condition, particularly when the denominators change for different outcomes; statement of the results in absolute numbers when feasible Indication of whether the analysis strategy was “intention to treat” or, if not, description of how non-compliers were treated in the analyses 		
Outcomes and estimation	17	<ul style="list-style-type: none"> For each primary and secondary outcome, a summary of results for each estimation study condition, and the estimated effect size and a confidence interval to indicate the precision Inclusion of null and negative findings Inclusion of results from testing pre-specified causal pathways through which the intervention was intended to operate, if any 		
Ancillary analyses	18	<ul style="list-style-type: none"> Summary of other analyses performed, including subgroup or restricted analyses, indicating which are pre-specified or exploratory 		
Adverse events	19	<ul style="list-style-type: none"> Summary of all important adverse events or unintended effects in each study condition (including summary measures, effect size estimates, and confidence intervals) 		
DISCUSSION				
Interpretation	20	<ul style="list-style-type: none"> Interpretation of the results, taking into account study hypotheses, sources of potential bias, imprecision of measures, multiplicative analyses, and other limitations or weaknesses of the study Discussion of results taking into account the mechanism by which the intervention was intended to work (causal pathways) or alternative mechanisms or explanations Discussion of the success of and barriers to implementing the intervention, fidelity of implementation Discussion of research, programmatic, or policy implications 		
Generalizability	21	<ul style="list-style-type: none"> Generalizability (external validity) of the trial findings, taking into account the study population, the characteristics of the intervention, length of follow-up, incentives, compliance rates, specific sites/settings involved in the study, and other contextual issues 		
Overall Evidence	22	<ul style="list-style-type: none"> General interpretation of the results in the context of current evidence and current theory 		

Anexo F. Directrices TOP (Nosek et al., 2016)

	Not Implemented	Level I	Level II	Level III
Citation Standards	No mention of data citation.	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 providing appropriate citation for data and materials following journal's author guidelines.
Data Transparency	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 prior to publication.
Analytic Methods (Code) Transparency	Journal encourages code sharing, or says nothing.	Article states whether code is available, and, if so, where to access it.	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 prior to publication.
Research Materials Transparency	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 prior to publication.
Design and Analysis Transparency	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.
Study Preregistration	Journal says nothing.	Article states whether preregistration of study exists, and, if so, where to access it.	Article states whether preregistration of study exists, and, if so, allows journal access during peer review for verification.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Analysis Plan Preregistration	Journal says nothing.	Article states whether preregistration of study with analysis plan exists, and, if so, where to access it.	Article states whether preregistration with analysis plan exists, and, if so, allows journal access during peer review for verification.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies, or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts results blind review.	Journal uses Registered Reports as a submission option for replication studies with peer review prior to observing the study outcomes.

Anexo G. Ejemplo Lista de Verificación
(<https://editorresources.taylorandfrancis.com>)

FIRST READ-THROUGH

- Is it clear what the authors want to communicate and the direction of the manuscript?
- Is it reporting original research or is it another type of article? How does this change your report?
- What contribution does the article make to the field of study?
- Is the manuscript original?
- Is the overall study design and approach appropriate?
- Are you concerned about the language? Are revisions needed to make it possible to review?

DETAILED REVIEW – RESEARCH ARTICLES

TITLE

- Does it express clearly what the manuscript is about?
- Does it highlight the importance of the study?
- Does it contain any unnecessary description?

ABSTRACT

- Is it a short and clear summary of the aims, key methods, important findings and conclusions?
- Does it include enough information to stand alone?
- Does it contain unnecessary information?

INTRODUCTION

- Does it clearly summarize the current state of the topic?
 - Does it address the limitations of current knowledge in this field?
 - Does it clearly explain why the study was necessary?
 - Does it clearly define the aim of the study and is this consistent with the rest of the manuscript?
 - Is the research question clear and appropriate?
-

METHODS

- Are the study design and methods appropriate for the research question?
 - Is there enough detail to repeat the experiments?
 - Is it clear how samples were collected or how participants were recruited?
 - Is there any potential bias in the sample or in the recruitment of participants?
 - Are the correct controls/ validation included?
 - Are any potential confounding factors considered?
 - Has any randomization been done correctly?
 - Is the time-frame of the study sufficient to see outcomes?
 - Is there sufficient power and appropriate statistics?
 - Do you have any ethical concerns?
-

RESULTS

- Are the results presented clearly and accurately?
 - Do the results presented match the methods?
 - Have all the relevant data been included?
 - Is there any risk of patients or participants being identified?
 - Is the data described in the text consistent with the data in the figures and tables?
-

DISCUSSION AND CONCLUSION

- Do the authors logically explain the findings?
 - Do the authors compare the findings with current findings in the research field?
 - Are the implications of the findings for future research and potential applications discussed?
 - Are the conclusions supported by the data presented?
 - Are any limitations of the study discussed?
 - Are any contradictory data discussed?
-

TABLES AND FIGURES

- Are data presented in a clear and appropriate manner?
 - Is the presentation of tables and figures consistent with the description in text?
 - Do the figure legends and table headings clearly explain what is shown?
 - Do the figures and tables include measures of uncertainty, such as standard error or confidence intervals, where required as well as the sample size?
 - Do you have any concerns about the manipulation of data?
-

REFERENCES

- Are there any key references missing?
 - Do the authors cite the initial discoveries where suitable?
 - Are there places where the authors cite a review but should cite the original paper?
 - Do the cited studies represent current knowledge?
-

FINAL CHECKS – BEFORE YOU SUBMIT YOUR REPORT

- Have you given a brief summary of the article and highlighted the key messages?
 - Have you given positive feedback as well as constructive criticism?
 - Have you made it clear which of your concerns are major (significant points, essential for publication) or minor (smaller issues, may not be essential for publication)?
 - Are your concerns specific, with examples where possible?
 - Have you numbered your comments and referred to page/ line numbers in the article to make it easy for the authors to address your points?
 - Is your feedback constructive, and focused on the research?
 - If you were the authors, would you understand how to improve the manuscript?
 - If you were the Editor, would the comments be detailed enough to help you make a decision?
 - Have you checked the spelling and grammar in your report?
 - Have you included your comments in the correct places in the online system – checking that any confidential comments for editors are in the right place – and have you answered all the questions?
-

SYSTEMATIC REVIEWS

- Are the search terms and inclusion/ exclusion criteria clearly described?
 - Are the search terms and criteria correct to ensure all the relevant articles are included?
 - If a meta-analysis has been done, were previous studies combined appropriately?
-

CASE REPORTS

- Does the diagnosis appear to be correct?
 - Was the treatment reasonable for the diagnosis?
 - Are the treatment and outcomes clearly described?
 - As far as possible, is the patient anonymous?
 - Are the conclusions reasonable and not attempting to generalize to wider population?
-

METHODOLOGY ARTICLES

- Is the new method clearly described?
 - Is it possible to replicate the new method?
 - Is there a rationale for why the new method is needed?
 - Is the new method compared to existing approaches?
 - Usually there should not be any experimental results, other than to demonstrate the utility of the methods.
-

REVIEW ARTICLES

- Is there any content which has been previously presented in a review?
 - Does it focus on recent advances in research?
 - Is it a balanced and unbiased overview of current understanding?
 - Are any recent or important references missing?
 - Is it too focused on the author's own research?
 - Is the interpretation and presentation of results of previous studies accurate and precise?
 - Has it a valuable contribution to the research field?
 - Is it understandable for non-expert readers?
-

OPINION ARTICLES (also called Editorials or Commentaries)

- Does the article add to the discussion on a research topic?
 - Is the opinion of the author well-argued?
 - Is the opinion based on current knowledge, or if it makes a big leap from current knowledge then is this logical? What supports the opinion presented?
-