

Reducir la difusión de noticias falsas sobre vacunas con ciencias del comportamiento¹

Reducing the Spread of Vaccine Misinformation through Behavioral Science

David Gómez-Valencia*

Universidad EAFIT. Medellín, Colombia

docvavi@gmail.com

ORCID: 0009-0008-5353-4627

Citar como: Gómez-Valencia, D. (2026). Reducir la difusión de noticias falsas sobre vacunas con ciencias del comportamiento. *Desde el Sur*, 18(3), e0072.

RESUMEN

El objetivo fue diagnosticar y reducir el comportamiento de compartir noticias falsas sobre vacunas en redes sociales, mediante un enfoque basado en ciencias del comportamiento. Para ello, se aplicó un diagnóstico mixto fundamentado en el modelo COM-B y los dominios TDF, que combinó una encuesta virtual ($n = 260$) y análisis de registros de cuentas de *fact-checking* colombianas. Posteriormente, se diseñaron, testearon y evaluaron a través de un A/B/C *testing* dos intervenciones (*prebunking* y *accuracy nudge*) integradas con el marco Behaviour Change Wheel (BCW). El *prebunking* mostró mejores resultados en contenidos neutros, mientras que el *accuracy nudge* destacó frente a noticias emocionales-morales. Este estudio valida el potencial de estas intervenciones para reducir la desinformación sanitaria y plantea lineamientos prácticos para la salud pública e investigaciones interdisciplinarias, incluyendo enfoques de ciencias sociales computacionales (CSC).

PALABRAS CLAVE

Desinformación sanitaria, emociones morales, noticias falsas, vacunas, redes sociales, comportamiento, *prebunking*, *accuracy nudge*, esquemas argumentativos

¹ Este trabajo forma parte de una tesis en la que participaron, además del autor de este artículo, Andrea Catalina Isaza Duarte y Simón Pineda Alarcón.

* Autor corresponsal: David Gómez-Valencia, Universidad EAFIT. Medellín, Colombia. Correo: docvavi@gmail.com

ABSTRACT

Objective: To diagnose and reduce the behavior of sharing fake news about vaccines on social media through a Behavioral Science approach. **Method:** A mixed diagnostic study was conducted based on the COM-B model and TDF domains, combining a virtual survey ($n = 260$) with an analysis of posts from Colombian fact-checking accounts. Subsequently, two interventions—prebunking and accuracy nudge—were designed, tested, and evaluated using A/B/C testing, integrated within the Behaviour Change Wheel (BCW) framework. **Results and conclusions:** *Prebunking* showed better outcomes with neutral content, while the *accuracy nudge* proved more effective for emotionally moralized news. This study validates the potential of these behavioral interventions to reduce health misinformation and proposes practical guidelines for public health and interdisciplinary research, including approaches from Computational Social Sciences (CSS).

KEYWORDS

Health misinformation, moral emotions, fake news, vaccines, social media, behaviour, prebunking, accuracy nudge, argumentation scheme

Introducción

Contexto y situación actual del problema

La desinformación sanitaria constituye hoy uno de los principales desafíos para la salud pública a nivel global. El Foro Económico Mundial (FEM, 2018) definió la desinformación en línea como uno de los principales riesgos globales para los sistemas ambientales, económicos, tecnológicos e institucionales de los que depende nuestro futuro. La Organización Mundial de la Salud (OMS, 2020) ha llegado a hablar de una «infodemia», aludiendo a la sobreabundancia de información —tanto veraz como falsa— que circula en entornos digitales y que dificulta la toma de decisiones informadas en la ciudadanía. La pandemia de covid-19 visibilizó de manera dramática las consecuencias de esta situación: teorías conspirativas sobre el origen del virus, dudas infundadas respecto a la eficacia de las vacunas, y la difusión masiva de remedios no comprobados proliferaron en redes sociales y canales de mensajería, lo que afecta la confianza en las autoridades sanitarias, incrementa conductas de riesgo y genera un estado de ánimo de desconfianza (Chou *et al.*, 2020; Murillo, 2020;

Roozenbeek *et al.*, 2023; Al Khaja *et al.*, 2018; Nan *et al.*, 2022; Peng *et al.*, 2022; Xue *et al.*, 2024).

Dentro de este panorama, la desinformación sobre vacunas ocupa un lugar crítico. La literatura ha documentado consistentemente que las dudas, los rumores y los discursos antivacunas han contribuido a la reducción de la cobertura inmunológica en distintos contextos, con consecuencias tangibles como el resurgimiento de enfermedades prevenibles (Dubé *et al.*, 2013). En América Latina, donde los sistemas de salud enfrentan desigualdades estructurales y recursos limitados, los efectos de la desinformación se combinan con otros factores sociales y culturales que erosionan la confianza pública en medidas de salud pública (FEM, 2018; Morejón-Llamas, 2023). En países latinoamericanos, este problema se traduce en una disminución de la cobertura vacunal infantil (OPS, 2019), y en un aumento de la reticencia hacia vacunas recientes como la del virus del papiloma humano (VPH) o las desarrolladas contra la covid-19 (Díaz *et al.*, 2025)

Las redes sociales juegan un papel decisivo en este fenómeno (Vosoughi *et al.*, 2018; Feng, 2024), pues se han consolidado como espacios centrales para la circulación de información en salud, especialmente entre las generaciones más jóvenes (Newman *et al.*, 2023). Sin embargo, la arquitectura algorítmica de estas plataformas favorece la amplificación de contenidos llamativos y emocionalmente cargados, independientemente de su veracidad (Van Bavel *et al.*, 2021). El estudio de Brady *et al.* (2017) ha mostrado que los mensajes que apelan a emociones morales como el miedo, la indignación o el asco tienen mayor probabilidad de viralizarse en comparación con aquellos que presentan información neutra o técnica. En este sentido, las dinámicas propias de las redes sociales no solo facilitan la propagación de noticias falsas, sino que además refuerzan patrones de consumo y compartición que favorecen su persistencia (Wardle y Derakhshan, 2017).

En este escenario, resulta prioritario comprender los factores sociales, culturales y psicológicos que motivan a las personas a creer y compartir noticias falsas sobre vacunas en entornos digitales. Identificar las barreras cognitivas y motivacionales que influyen en este comportamiento es un primer paso para diseñar intervenciones efectivas que no se limiten a corregir información errónea, sino que actúen preventivamente para reducir la probabilidad de difusión.

Objetivo del trabajo

Los objetivos principales de este trabajo son comprender, desde un enfoque comportamental, factores psicológicos, contextuales y

motivacionales que influyen en que las personas creen y compartan noticias falsas sobre vacunas en redes sociales. Además, se propone diseñar, implementar y evaluar soluciones basadas en ciencias del comportamiento orientadas a reducir su propagación (Gómez *et al.*, 2025).

Bases teóricas

Para abordar el fenómeno de la desinformación, se empleó el modelo COM-B (Capacidad, Oportunidad, Motivación y Comportamiento), ampliamente utilizado en estudios sobre cambio de comportamiento (Michie *et al.*, 2014). Este modelo permite entender la conducta como resultado de la interacción entre tres componentes fundamentales:

- **Capacidad:** las habilidades psicológicas y físicas necesarias para llevar a cabo el comportamiento; en este caso, las habilidades psicológicas asociadas a la alfabetización mediática y sanitaria (Berkman *et al.*, 2011; Norman y Skinner, 2006).
- **Oportunidad:** los factores contextuales, tanto físicos (y digitales, por extensión) como sociales, que posibilitan o dificultan el comportamiento, como el diseño de las plataformas, el lenguaje de los contenidos, las normas sociales o el *encuadre* del lenguaje de los contenidos. El encuadre (o *framing* en inglés) es el modo en que se presenta la información a las personas e implica la influencia en sus percepciones, opciones y decisiones (Aruguete *et al.*, 2023). Es el caso de las apelaciones a emociones morales en las noticias, pues se ha demostrado que cada palabra con carga emocional-moral incrementa la probabilidad de que un mensaje sea compartido (Van Bavel *et al.*, 2021; Brady *et al.*, 2017).
- **Motivación:** procesos internos —reflexivos y automáticos— que orientan la acción, como el deseo de proteger a otros, las creencias sobre la credibilidad de las fuentes, la percepción de las propias capacidades o la valoración de las consecuencias de compartir desinformación sanitaria. Acá va a ser particularmente importante la motivación altruista: cuando las personas, aun reconociendo que un mensaje podría ser desinformativo, lo comparten conscientemente por considerar que tiene un alto nivel de utilidad social o que puede proteger a otras personas cercanas de sufrir daños (Ardèvol-Abreu *et al.*, 2020; Duffy *et al.*, 2020; Metzger *et al.*, 2021).

Estos tres componentes (capacidad, oportunidad y motivación) interactúan dinámicamente para dar lugar al comportamiento de interés.

La desinformación es entendida como cualquier tipo de información falsa o engañosa, incluyendo aquella que, si bien no es objetivamente incorrecta, puede distorsionar los hechos, carecer de contexto o recurrir a falacias lógicas (List *et al.*, 2024), abarca contenidos digitales diversos

como noticias falsas (*fakes news*), rumores, propaganda y teorías de conspiración (Roizenbeek *et al.*, 2023). Este trabajo se centró en las noticias falsas sobre vacunas, que se inscriben dentro del ámbito de la desinformación sanitaria, entendida como cualquier contenido relacionado con la salud que ha sido demostrado como falso (Bautista *et al.*, 2021; Feng, 2024) o que se aleja de la evidencia científica objetiva y puede inducir a error al receptor (Xue *et al.*, 2024).

Un elemento central que la literatura reciente ha subrayado es el papel de las emociones morales en la viralidad de la información. Van Bavel *et al.* (2021) y Brady *et al.* (2017) han documentado que los mensajes cargados con apelaciones emocionales relacionadas con moralidad tienden a recibir más interacciones y a propagarse más rápidamente. Esta característica de la comunicación digital plantea un reto adicional: no basta con corregir información errónea, sino que se requiere considerar los procesos emocionales y motivacionales que impulsan la decisión de compartir.

En este sentido, la integración de marcos conductuales (COM-B, TDF, BCW) con la literatura sobre desinformación y emociones morales ofrece un abordaje sólido y novedoso para enfrentar el problema. El presente estudio se inscribe en esta línea, pues busca aportar tanto evidencia empírica sobre el fenómeno en el contexto colombiano como propuestas prácticas de intervención aplicables en campañas de salud pública.

Materiales y métodos

Descripción de la problemática a analizar

El fenómeno de la desinformación sanitaria, particularmente en el campo de las vacunas, ha cobrado una relevancia crítica en los últimos años debido a sus efectos directos sobre la confianza pública y las decisiones individuales en torno a la inmunización (FEM, 2018; Morejón-Llamas, 2023). La literatura especializada muestra que la propagación de noticias falsas puede erosionar la percepción de seguridad y eficacia de las vacunas, alimentar la reticencia vacunal y, en consecuencia, poner en riesgo los logros alcanzados en materia de salud pública (Díaz *et al.*, 2025; Dubé *et al.*, 2013; OPS, 2019).

En este contexto, se justifica la necesidad de estudiar no solo la presencia de estas noticias falsas en redes sociales, sino también los factores que influyen en que las personas decidan compartirlas. Las explicaciones más simplistas, que reducen el problema a la falta de información, han demostrado ser insuficientes (Pennycook *et al.*, 2020). El comportamiento de compartir está atravesado por motivaciones diversas: desde la excesiva confianza en la precisión de los juicios propios (el sesgo de sobrepresión [Murad, 2024]), hasta el deseo de advertir o proteger a otros a través del

acto de compartir, aun cuando la información sea dudosa (motivación altruista [Ardèvol-Abreu *et al.*, 2020; Duffy *et al.*, 2020; Metzger *et al.*, 2021]).

Así, el presente estudio partió de la premisa de que para comprender de manera integral el fenómeno es necesario considerar no solo factores cognitivos (como la alfabetización mediática y sanitaria), sino factores motivacionales (entre ellos, los creencias en torno a la desinformación, la sobreconfianza, y la motivación altruista) y factores contextuales (como el uso de lenguaje emocional-moral con el que se encuadra una noticia).

Metodología empleada

Con el propósito de abordar el problema de las noticias falsas sobre vacunas y proponer soluciones comportamentales, se desarrolló un diseño mixto de tipo concurrente², con propósitos descriptivos y aplicados. El estudio integró técnicas cuantitativas y cualitativas, de modo que el análisis diagnóstico fundamentara el diseño de las intervenciones.

Diagnóstico

En la fase diagnóstica, el modelo COM-B se operacionalizó a través del Theoretical Domains Framework (TDF), que desglosa las dimensiones del comportamiento en dominios observables y medibles, tales como el conocimiento, las habilidades, influencia social, los factores ambientales/contextuales, las creencias sobre capacidades, las creencias sobre consecuencias, entre otras (Michie *et al.*, 2014). Este marco facilitó la construcción de los instrumentos para el desarrollo de una encuesta virtual.

La encuesta estuvo compuesta por tres cuestionarios de preguntas cerradas y abiertas validadas mediante *think-aloud protocol*, y fue complementada con un análisis de registros de los últimos doce meses de contenidos digitales de los principales proyectos de *fact-checking* en Colombia³, lo que permitió un acercamiento mixto a fin de mapear las barreras y facilitadores del comportamiento de compartir noticias falsas sobre vacunas en redes sociales.

Los cuestionarios distribuyeron aleatoriamente a los participantes entre versiones con lenguaje emocional-moral o lenguaje neutro, diferenciadas por el tono del titular y las imágenes. Cada uno incluyó una simulación digital controlada⁴ de noticias falsas sobre vacunas, creadas

2 De acuerdo con Hernández *et al.* (2022), un diseño concurrente implica la recolección y el análisis simultáneo de datos cuantitativos y cualitativos.

3 Colombia Check y La Silla Vacía.

4 Tanto la encuesta como la simulación controlada fue revisada y asesorada por Beatriz Vallejo, asesora de la tesis de maestría sobre la que se basa este artículo. Beatriz Vallejo es CEO de Ethos BT y directora del Detox Information Project (DIP). Utiliza evidencia y las

específicamente para este estudio. A través de esta herramienta experimental se presentaron a los participantes contenidos falsos que integraban titulares, fuentes fabricadas e imágenes que variaban en el tipo de lenguaje utilizado. Esta simulación permitió medir comportamientos a través de ítems conductuales (con opción única de respuesta «me gusta», «compartir», «ignorar», «otro»⁵). Los dos cuestionarios cuantitativos tenían secciones Likert, de preguntas múltiple y dicotómica, que buscaban recabar información cuantitativa y cualitativa en torno a tres dimensiones: alfabetización mediática, sanitaria y nivel de creencias frente a noticias falsas sobre vacunas; mientras que el cuestionario cualitativo indagaba por las razones para compartir o no, así como las principales preocupaciones por compartir desinformación sanitaria.

La encuesta fue respondida por 386 personas colombianas, de las cuales se descartaron 126 casos (33 %) por no aprobar los controles de atención incorporados en los cuestionarios. Así, la muestra final quedó con 260 participantes; 144 de ellos respondieron el cuestionario cualitativo y 116 respondieron los dos cuestionarios cuantitativos. Se trató de una muestra por conveniencia, autoseleccionada y compuesta mayoritariamente por personas con acceso a internet y familiaridad con el entorno digital. Esto implica posibles sesgos de autoselección.

En general, encontramos en la muestra una población colombiana predominantemente adulta joven y de mediana edad, con una mayor concentración entre los 30 y 39 años (29 %), seguida de los grupos de 25 a 29 años (23 %) y 50 a 64 años (22 %). A su vez, la participación de personas entre 18 y 24 años (11 %) y mayores de 65 años (2 %) fue más reducida. Predomina el estado civil soltero (50 %) y una mayoría femenina (67 %). Además, se observa un nivel educativo alto, con mayoría de participantes con estudios de pregrado (38 %), especialización (22 %) y maestría (19 %). La mayoría se encontraba empleada (54 %) al momento de diligenciar el cuestionario, residía en zonas urbanas (94 %) y pertenecía a estratos socioeconómicos medios y altos, entre los que destacaba el estrato 4 (31 %), seguido por los estratos 5 y 3 (22 % cada uno)⁶ (Gómez *et al.*, 2025).

El análisis de registro se hizo en torno a las publicaciones que La Silla Vacía y ColombiaCheck, dos importantes proyectos de chequeo de datos

ciencias del comportamiento para generar alfabetización digital, reducir la polarización, la violencia y la desinformación (<https://www.somosdip.com/>).

5 La codificación emergente del análisis cualitativo permitió identificar que «Otro» estaba asociado a verificar o leer la noticia.

6 En Colombia los estratos son una clasificación socioeconómica que se asigna así: estrato 1: bajo-bajo, estrato 2: bajo, estrato 3: medio-bajo, estrato 4: medio, estrato 5: medio-alto, estrato 6: alto.

en Colombia, en sus cuentas de Instagram y páginas de web durante los últimos doce meses. Se examinaron variables asociadas al uso de lenguaje emocional-moral, con el fin de identificar en qué medida estos proyectos usaban este tipo de lenguaje para combatir la desinformación.

El análisis consistió en el conteo de palabras emocionales-morales presentes en las publicaciones de estas cuentas. Para lograr esto, se construyó un instrumento de conteo basado en un diccionario de palabras emocional-morales creado previamente por los investigadores adaptado del trabajo de Brady *et al.* (2017). Operativamente, una palabra emocional-moral fue definida como aquella que aparece tanto en un diccionario de términos emocionales como en uno de términos morales. Para las palabras emocionales, se usó el NRC Emotion Lexicon desarrollado por Mohammad y Turney (2013), una herramienta ampliamente utilizada en análisis computacional del lenguaje afectivo. Para las palabras morales, se acudió a la categorización propuesta por Graham *et al.* (2009) en su estudio sobre fundamentos morales.

Diseño, testeo y evaluación de las intervenciones

A partir del diagnóstico, se empleó el Behaviour Change Wheel (BCW) (Michie *et al.*, 2014). como guía para el diseño de intervenciones. El BCW vincula los hallazgos del diagnóstico con funciones de intervención —categorías de estrategias dirigidas a modificar el comportamiento, como educar, persuadir, en este caso—y técnicas validadas empíricamente (Behaviour Change Techniques [BCT]).

La literatura sobre desinformación digital ha evidenciado que las estrategias correctivas tradicionales, como los *fact-checking*, tienen un efecto limitado (List *et al.*, 2024). En contraste, los enfoques preventivos como el *prebunking* han demostrado mayor eficacia, pues preparan cognitivamente a los individuos para resistir la persuasión de información falsa antes de exponerse a ella (Traberg *et al.*, 2022). Por su parte, los *accuracy nudges* se han mostrado prometedores al inducir de forma sutil a los usuarios a priorizar la precisión por encima del beneficio colectivo percibido del contenido o el impacto emocional en sus decisiones de compartir contenido (Kozyreva *et al.*, 2024).

Considerando este respaldo científico, así como los resultados del diagnóstico, se diseñaron dos estrategias de intervención fundamentadas en literatura internacional sobre desinformación digital y validadas en estudios previos:

- **Prebunking:** basado en el principio de inoculación psicológica, consiste en exponer de manera preventiva a las personas a versiones débiles

de desinformación, junto con explicaciones que revelen las técnicas manipulativas empleadas. De esta forma, se busca «inocular» cognitivamente a las personas dotándolas de defensas mentales frente a futuros intentos de persuasión con información falsa (Mcguire, 1964, citado en Traberg *et al.*, 2022). La intervención combinó los dos componentes esenciales de la inoculación psicológica —advertencia y refutación preventiva— para fortalecer capacidades psicológicas asociadas al reconocimiento de contenidos engañosos (Traberg *et al.*, 2022) a través de un video humorístico⁷ grabado con celular en plano cenital que parodiaba una «receta» donde se mostraban «ingredientes» engañosos de las noticias falsas sobre vacunas: la apelación emocional-moral y la fabricación de fuentes pseudoautorizadas.

Además de ello, incorporó novedades al buscar incidir en el componente motivacional a través de mensajes que operaban sobre barreras mapeadas, tales como el exceso de confianza para identificar fuentes confiables (sesgo de sobreprecisión) y la creencia en que compartir es igual que informar a otros (asociada a la motivación altruista), así como facilitadores asociados a las preocupaciones por desinformar y las figuras de cuidado (abuelos) identificadas en el diagnóstico.

- **Accuracy nudges:** consiste en recordatorios sutiles que llaman la atención de los usuarios sobre la importancia de priorizar la precisión al momento de decidir qué contenidos compartir (Pennycook *et al.*, 2020). Estudios previos han demostrado que incluso un breve recordatorio sobre la relevancia de la precisión puede reducir significativamente la difusión de desinformación (Kozyreva *et al.*, 2024; Sultan *et al.*, 2024). Esta intervención, por su parte, consistió en un video breve en bucle con mensajes secuenciales; su forma y contenido fueron diseñados para promover un reencuadre cognitivo en torno al significado de «compartir»⁸: «Compartir no es lo mismo que informar», «Compartir información verificada Sí lo es», «Verifica. Informa bien a los demás». Junto con estímulos visuales salientes (efectos de neón, íconos y emojis), buscaron generar una asociación automática entre «verificar» e «informar» y reforzar el foco atencional en la precisión. Con esto, se buscó cuestionar la asociación automática del «compartir» con «informar» y sugerir que solo compartir información verificada cumple realmente a ese fin. Este reencuadre no invalida el valor que los usuarios otorgan al acto de compartir, sino que reorganiza sus condiciones, ya que incorpora la verificación como un criterio central y eleva el estándar al apelar a un comportamiento moralmente deseable: «informar

7 El video de elaboración propia puede verse en este enlace: <https://youtube.com/shorts/8b012UZ5aTk?si=vABVb6pYUJlil7q5>. Especial agradecimiento a Diego Mosquera, egresado de la maestría en Estudios del Comportamiento de la Universidad EAFIT, por prestar su voz para la narración del video.

8 El video de elaboración propia puede verse en este enlace: <https://www.youtube.com/shorts/t0wJYgzlvMo>

bien». Esto se vincula con valores prosociales identificados en el diagnóstico —como el altruismo, el cuidado y la responsabilidad—, e incorpora recomendaciones derivadas de la revisión bibliográfica sobre el uso de encuadres emocionales positivos como estrategia efectiva para contrarrestar la desinformación (Aruguete *et al.*, 2023).

Ambas intervenciones fueron seleccionadas y ajustadas después de un proceso de testeo de los tres pilotos: historieta (versión preliminar del video *prebunkig*), cartel digital (versión preliminar del *accuracy nudge*) y video tipo *norm nudge* que fue descartado. El proceso se realizó con 121 estudiantes de la Universidad EAFIT (Sede Medellín Colombia) en dos iteraciones, bajo un esquema de doble aleatorización según tipo de intervención y tipo de lenguaje (emocional-moral o neutro). Se evaluaron dos variables: porcentaje que compartía noticias falsas sobre vacunas y nivel de creencia en noticias falsas sobre vacunas.

En la primera iteración se comparó cartel con historieta. Ningún participante compartió noticias falsas con lenguaje emocional-moral; frente a contenido neutro, lo hizo el 13 % del grupo del cartel y 0 % del de la historieta. La historieta destacó además por un 40 % de participantes con bajo nivel de creencia, lo que superó el 33 % alcanzado por el cartel en ambos lenguajes. En la segunda iteración, se comparó historieta con *norm nudge*. En este caso, el cartel redujo más la difusión (8 % compartieron frente a 17 % en el *norm nudge*) y mostró niveles de creencia más bajos (59 % versus 50 % en emocional-moral; 100 % versus 41 % en neutro). Con base en estos resultados, se seleccionaron el cartel y la historieta para la intervención final, y se descartó el *norm nudge*.

La población finalmente intervenida ($n = 229$) con las versiones definitivas estuvo conformada por estudiantes la Universidad de Antioquia (79 %, $n = 181$) y la Universidad EAFIT (15 %, $n = 34$) en Medellín, con un 6 % adicional sin institución reportada. Predominó una población joven (71 % entre 18 y 24 años), de estratos 2 y 3 (72 %) (Gómez *et al.*, 2025). Ninguno de los participantes había diligenciado la encuesta del diagnóstico. Esta población se distribuyó aleatoriamente entre los tres grupos experimentales (*prebunking*, *accuracy nudge* y control). Los grupos tratamiento eran expuestos aleatoriamente a una de las dos intervenciones antes de presentarles las noticias falsas, mientras que al grupo control se le presentaban estas noticias sin ninguna intervención previa.

Buscando que los resultados observados fueran atribuibles a las intervenciones, se implementó un control metodológico buscando fortalecer la validez interna del estudio. Este control se centró, en primer lugar, en la construcción de los estímulos replicando y ajustando la simulación digital

usada para el diagnóstico, donde se elaboraron noticias simuladas para aislar factores clave de difusión: se varió su veracidad (noticias verdaderas versus falsas), el tema (vacunas versus genérico), y el lenguaje (apelación emocional-moral y neutro). Además, para garantizar la coherencia interna de los estímulos emocionales, se procuró que el lenguaje de apelación emocional-moral se enfocara preferentemente en dos categorías morales específicas: Daño-Cuidado y Pureza-Impureza (Graham *et al.*, 2009). En segundo lugar, se integraron controles como el uso de redacción no dirigido en los ítems de la simulación, el registro sistemático del contexto de aplicación a través de bitácoras de observación, así como un protocolo de fidelidad de entrega que estandarizaba la forma en que se presentaba y ejecutaba el experimento.

Las intervenciones fueron evaluadas con este mismo diseño cuasiexperimental aleatorizado tipo *A/B/C testing* aplicado mediante la plataforma Typeform, que implementó, como se dijo, dos niveles de aleatorización: asignación de participantes a uno de tres grupos (*prebunking*, *accuracy nudge* o grupo control) y el tipo de lenguaje (emocional-moral o neutro) (Gómez *et al.*, 2025).

Lineamientos éticos

El presente trabajo se desarrolló bajo lineamientos éticos, en conformidad con las normas nacionales e internacionales para investigaciones con seres humanos y en concordancia con las recomendaciones establecidas por el Código de Conducta para Editores de Revistas (COPE).

Aunque este estudio no fue sometido a la evaluación de un comité de ética institucional, se adoptaron rigurosamente principios éticos fundamentales para la investigación con personas. Así, los procedimientos fueron realizados respetando las normas éticas concordantes con la Declaración de Helsinki. Los participantes fueron debidamente informados sobre los objetivos del estudio, el uso de sus respuestas y su derecho a desistir en cualquier momento sin repercusiones. Para proteger su privacidad, los datos personales fueron tratados con estricta confidencialidad y las encuestas fueron anónimas. La información recolectada fue utilizada exclusivamente con fines académicos y de publicación, respetando en todo momento la dignidad, la autonomía y el bienestar de los participantes. Además, se prestó atención a evitar cualquier desinformación derivada de la exposición a noticias falsas, incluyendo la provisión de aclaraciones inmediatas sobre el carácter ficticio del material empleado en la simulación.

Finalmente, se adoptó un compromiso explícito con la transparencia científica, tanto en el manejo de los datos como en la presentación de los resultados. Esto incluyó la obligación de reportar los hallazgos sin

sesgos, reconocer las limitaciones del estudio y poner a disposición los aprendizajes obtenidos para contribuir de manera responsable al avance de la investigación sobre desinformación digital y salud pública. Estos principios fueron comunicados a los participantes antes de someterse a las intervenciones.

Resultados

Resultados del diagnóstico

El análisis del diagnóstico reveló hallazgos que cuestionan explicaciones comunes sobre la difusión de noticias falsas en salud. Una de las observaciones más relevantes fue la escasa relación entre la alfabetización mediática y sanitaria y la propensión a compartir desinformación. Aunque la literatura especializada hacía suponer que mayor alfabetización mediática o sanitaria protegería frente a la difusión de contenido falso (Berkman *et al.*, 2011; Norman y Skinner, 2006), los datos mostraron que estos niveles de alfabetización no se tradujeron en comportamientos más cautelosos: los valores de las correlaciones obtenidos fueron cercanos a 0 en alfabetización sanitaria ($r = 0,01$) y alfabetización mediática ($r = -0,03$); por esta razón, no se identificaron barreras ni facilitadores relevantes asociados al componente de capacidad psicológica.

Este hallazgo sugiere que, si bien las habilidades y capacidades cognitivas pueden ser relevantes, no son por sí solas determinantes del comportamiento de compartir noticias falsas sobre vacunas, lo cual refuerza la necesidad de ampliar la mirada hacia factores motivacionales y contextuales⁹.

La codificación abierta cualitativa de las respuestas también aportó información valiosa. Mediante el uso del *software* ATLAS.ti, se identificaron dos barreras motivacionales centrales. La primera fue la excesiva confianza de los participantes en su capacidad para reconocer fuentes confiables, pese a que en la práctica no lograron hacerlo. Esta barrera, asociada al sesgo de sobreprecisión, hace referencia a la excesiva certeza en la precisión de los juicios propios (Murad, 2024), se manifestó en la brecha entre la autopercepción y la conducta observada en los participantes. Al presentarles una noticia falsa con una fuente fabricada¹⁰, si el participante la compartía, se le preguntaba: «¿Habría alguna razón por la que no hubieras compartido esa publicación?», muchos mencionaron señales de alerta

9 Además, las bajas correlaciones también pueden estar relacionadas con la baja cantidad de encuestados.

10 Las fuentes fabricadas fueron: «viralexpress.info», «elojocritico.news», «Universidad de York» y «Colegio de Médicos de Colombia».

que, paradójicamente, ya estaban explícitamente presentes en el contenido, como la ausencia de respaldo científico o la falta de certeza sobre la fuente. Así, dieron respuestas como: «Que la fuente no sea confiable», «Si no tiene soporte investigación», «Si me doy cuenta de que la fuente no es confiable», «Que sea falsa» y «Si tal vez tuviese alguna pista de ser scam», lo que revela una sobreestimación de sus propias habilidades de detección.

Otra barrera identificada fue una dinámica de motivación altruista (Ardèvol-Abreu *et al.*, 2020; Duffy *et al.*, 2020; Metzger *et al.*, 2021), donde la decisión de compartir se justificó como un medio legítimo para «informar» o responder a un «interés público», incluso por encima de la duda sobre la veracidad de la noticia. Se codificaron 27 citas en las que las respuestas estaban asociadas a la motivación por «informar» a los demás. Por ejemplo: «Para informar», «Para que más personas se enteren», «Para informar a mis seguidores», «Informar a más personas», «Para que las personas se enteren», «Para que más personas estén informadas», «Importante que la gente esté enterada», «Para que más personas lo lean», etc. Desde esta perspectiva, el compartir no era concebido como un comportamiento riesgoso, sino como un gesto solidario. El patrón de motivación altruista se ve reforzado por una correlación negativa entre el nivel de creencia en las noticias falsas sobre vacunas y el acto de compartirlas ($r = -0,24$), lo que sugiere que los usuarios pueden difundir noticias falsas aun sin estar convencidos de ellas, siempre que las perciban como útiles para otros.

Otro aspecto clave del diagnóstico estuvo relacionado con el tipo de lenguaje utilizado en las noticias falsas. En la simulación inicial del diagnóstico, las piezas que contenían un lenguaje emocional-moral fueron menos compartidas que aquellas con lenguaje neutro. Cuando la intensidad del lenguaje fue ajustada y moderada, se observó un patrón diferente: las noticias con apelaciones emocionales comenzaron a ser más compartidas que las neutras, por lo que se alinearon así con la evidencia previa sobre el rol amplificador de las emociones morales en redes sociales. Se profundizará más adelante.

Al mismo tiempo, el análisis cualitativo también permitió identificar preocupaciones por los posibles efectos negativos de compartir desinformación, como generar miedo innecesario, inducir a decisiones de salud equivocadas, provocar pánico colectivo o sentirse responsable por haber difundido un contenido falso¹¹.

11 Estas categorías son el resultado de la codificación de las respuestas que dieron los encuestados que compartieron alguna noticia falsa sobre vacunas en la simulación controlada, a la pregunta abierta «¿Qué es lo que más te preocuparía si te dieras cuenta

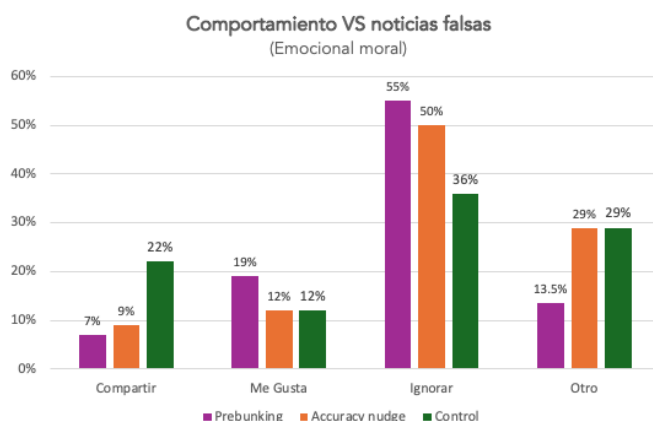
Resultados de la intervención

El *prebunking* mostró resultados más prometedores en contenidos neutros en comparación con los de lenguaje emocional-moral. Solo el 4 % de los participantes expuestos al *prebunking* compartieron noticias falsas sobre vacunas en lenguaje neutro, mientras el 14 % compartieron en grupo control en este mismo lenguaje. Este resultado refuerza el potencial preventivo de la inoculación psicológica en contextos donde el contenido es emocionalmente neutro.

En contraste, el *accuracy nudge* tuvo resultados similares ante los contenidos emocionales-morales. En este tipo de lenguaje, el grupo que recibió esta intervención llegó al 7 %, en tanto que el 22 % del grupo control la compartió.

No obstante, también se registraron posibles efectos colaterales. En los grupos expuestos al *prebunking*, se observó una disminución en la opción «otro» (asociada a verificar, leer). En lenguaje emocional-moral solo el 13,5 % eligió esta opción alternativa, frente al 29 % en el grupo control. De igual forma, en lenguaje neutro 12,5 % eligió «otro», frente al 17 % en el grupo control.

Otro hallazgo relevante fue el incremento de «me gusta» frente a noticias falsas, especialmente en el grupo expuesto al *prebunking* con contenidos emocionales-morales (19 % frente al 12 % del grupo control) y, en menor medida, en el de *accuracy nudge* con contenidos neutros (23 % frente al 22 % del grupo control).



de que has compartido información falsa sobre salud en tus redes sociales?». En total se identificaron 25 citas: 11 citas se asignaron a generar miedo innecesario; 8 citas, a inducir a otros a tomar decisiones de salud equivocadas; y 6 citas, a sentirse responsable por haber compartido contenido falso.

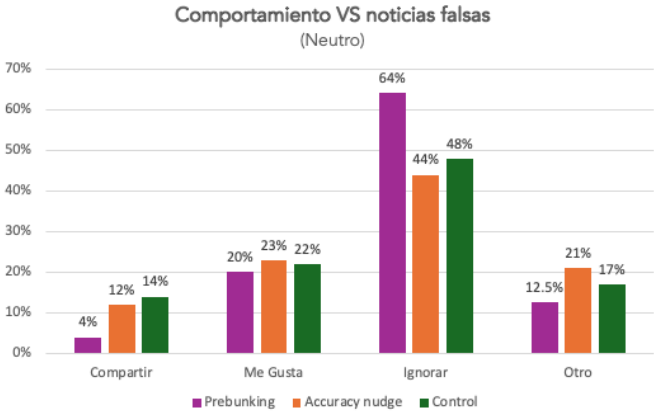


FIGURA 1. Resultados de intervención (noticias falsas sobre vacunas en lenguaje emocional-moral)

FIGURA 2. Resultados de intervención (noticias falsas sobre vacunas en lenguaje neutro)

Discusión y conclusiones

Los resultados de este trabajo permiten afirmar que las intervenciones conductuales implementadas —*prebunking* y *accuracy nudges*— mostraron un importante potencial para la reducción del comportamiento de compartir noticias falsas sobre vacunas en redes sociales, tanto en lenguaje de apelación emocional moral como neutro. Esta evidencia se alinea con la literatura internacional que ha documentado el potencial de este tipo de estrategias para contener la desinformación en línea (Van Bavel *et al.*, 2021; Kozyreva *et al.*, 2024). Los resultados diferenciales observados según el tipo de lenguaje —el *prebunking* mostró mejores resultados en contenidos neutros y los *nudges* en aquellos con apelaciones emocionales-morales— refuerza la importancia de diseñar intervenciones que consideren esta variable.

Sin embargo, una lectura más profunda de los resultados revela dimensiones menos evidentes del fenómeno. Por ejemplo, el análisis cualitativo reveló que una de las principales motivaciones detrás del comportamiento de compartir era el deseo de «informar» a los demás. En ese sentido, el compartir era visto como un gesto altruista orientado al bien común. Este hallazgo coincide con estudios previos que señalan que, cuando un contenido es percibido como socialmente útil, las personas tienden a privilegiar esta utilidad por encima de su veracidad (Duffy *et al.*, 2020; Metzger *et al.*, 2021). Esta motivación altruista, si bien opera como una barrera para la contención de la desinformación, también puede servir como facilitador si las intervenciones se diseñan apelando y dirigiendo a esta prosocialidad.

Otro hallazgo relevante emergió en relación con el tipo de lenguaje utilizado en las piezas desinformativas. En la fase de diagnóstico, las noticias falsas sobre vacunas con lenguaje emocional-moral fueron menos compartidas que las formuladas en lenguaje neutro. Este resultado fue contrario a los hallazgos de la literatura internacional que señalan la carga emocional tiende a amplificar la viralidad (Brady *et al.*, 2017).

Cuando la intensidad del lenguaje fue moderada en la aplicación de la intervención (con diferentes participantes a los del diagnóstico), el patrón cambió: las noticias con apelaciones emocionales-morales fueron más compartidas que las neutras, en línea con la evidencia que sitúa a la carga emocional y moral como motores de la viralidad digital (Brady *et al.*, 2017). Una posible explicación es que el tono excesivamente «amarillista» del diseño inicial generó desconfianza entre los participantes y atenuó el efecto esperado. Queda abierta la pregunta si esta posible desconfianza ocurre cuando el amarillismo se usa para hablar de temas de salud y vacunas, o si existe una suerte de umbral de carga emocional-moral que, una vez superado, genera alerta en las personas.

De igual forma, el hecho de que ambas intervenciones mostraran posibles efectos colaterales plantea preguntas sobre la profundidad del cambio comportamental alcanzado. Por ejemplo, en los grupos expuestos al *prebunking* se registraron porcentajes más bajos de respuestas en la categoría «otro» (asociada a verificar, leer). Esto podría sugerir que, aunque disminuyó el compartir, no necesariamente aumentó la adopción de comportamientos alternativos como la verificación activa o la búsqueda de información adicional. En otras palabras, parte de los resultados observados podría deberse a un aumento de la pasividad, más que a un cambio crítico en la toma de decisiones informadas.

Finalmente, otro resultado digno de atención fue el incremento en la cantidad de «me gusta» dados a contenidos falsos sobre vacunas, lo cual se interpreta en contraste con el comportamiento del grupo control. Este comportamiento introduce una paradoja: aunque la acción de compartir se redujo, el aumento de reacciones positivas puede seguir contribuyendo a la amplificación algorítmica de la desinformación. Dado que los «me gusta» actúan como indicadores de relevancia dentro de la lógica de las plataformas digitales, su incremento puede ampliar la exposición del contenido, incluso si no se comparte activamente. Este hallazgo invita a repensar la efectividad de las intervenciones no solo en términos de reducción del compartir, sino también en su capacidad para modificar otras formas de interacción que sostienen la circulación de noticias falsas.

Conclusiones y recomendaciones

En síntesis, los resultados de esta investigación permiten concluir que las intervenciones conductuales basadas en *prebunking* y *accuracy nudges* constituyen herramientas prometedoras para reducir la propensión a compartir noticias falsas sobre vacunas en redes sociales. No obstante, su eficacia depende del tipo de lenguaje de la desinformación y su impacto se limita a una de las múltiples formas de interacción digital.

Con base en estos hallazgos, se plantean las siguientes recomendaciones (Gómez *et al.*, 2025):

1. **Incorporar el *prebunking* como «prevacuna» psicológica en campañas de salud pública:** en los contextos latinoamericanos donde se busque fortalecer modelo de atención primaria en salud orientado a la prevención y la promoción, integrar estrategias de *prebunking* en las prácticas educativas de los equipos sanitarios podría resultar especialmente prometedor. Esta estrategia se alinea con la metodología IEC (Información, Educación, Comunicación), ampliamente utilizada para fomentar comportamientos saludables en el ámbito comunitario, y ofrece una herramienta programáticamente viable para la salud pública. Su aplicación preventiva en campañas como las Jornadas Nacionales de Vacunación o aquellas dirigidas a grupos de riesgo (por ejemplo, contra la influenza o la fiebre amarilla [Ministerio de Salud, 2024]) podría funcionar como una «prevacuna» psicológica que refuerce la confianza en la inmunización y limite la propagación de la desinformación sanitaria en momentos críticos como epidemias o pandemias.
2. **Segmentar las intervenciones según perfiles de reticencia vacunal:** las estrategias para contrarrestar la desinformación sobre vacunas podrían beneficiarse del conocimiento acumulado sobre la reticencia vacunal (MacDonald, 2015), al diferenciar entre reticencia parcial (frente a una vacuna específica) y rechazo generalizado (a la vacunación en su conjunto). Se pueden formular intervenciones segmentadas con mayor precisión, lo que aumentaría su efectividad al responder a perfiles más definidos de comportamientos y creencias asociados a tipos de vacuna.
3. **Integrar perspectivas de machine behaviour y ciencias sociales computacionales:** abordar la amplificación algorítmica de la desinformación exige considerar que interacciones aparentemente inocuas (como los «me gusta») también actúan como señales de relevancia que aumentan la visibilidad de los contenidos falsos. En futuras fases de testeo, sería clave incorporar controles que mitiguen estos efectos colaterales y ampliar el análisis con enfoques interdisciplinarios, como el *machine behaviour* (Rahwan *et al.*, 2019), que estudia a los algoritmos y sistemas de Inteligencia Artificial (IA) como agentes con patrones de acción que influyen en las decisiones humanas (y son influidos

por estas). Asimismo, el uso de modelos basados en agentes desde las ciencias sociales computacionales permitiría simular perfiles comportamentales, propagar artificialmente una noticia falsa y observar el impacto de distintas estrategias contra la desinformación, al modificar parámetros de viralidad de la noticia, tamaño y composición de la población, entre otros, lo que favorece diseños más adaptados, sensibles a la heterogeneidad poblacional y con un control experimental más robusto.

4. **Integrar la vulnerabilidad a la desinformación como un determinante emergente de la salud:** los resultados de este trabajo aportan insumos para las líneas que exploran la relación entre vulnerabilidad a la desinformación y salud, y abren la pregunta sobre si dicha vulnerabilidad debería considerarse dentro de los determinantes sociales de la salud (Castiel y Sanz-Valero, 2010). Integrar dimensiones como la alfabetización mediática y la relación humano-tecnología en los análisis de riesgo sanitario permitiría actualizar las categorías desde las cuales se comprende el bienestar, al incorporar factores emergentes del ecosistema digital.

En conclusión, este estudio aporta evidencia empírica y modelos que validan la pertinencia de las intervenciones comportamentales para abordar la desinformación digital. Al mismo tiempo, abre nuevas preguntas sobre la relación entre cultura digital y salud, así como sobre los horizontes de aplicabilidad, escalabilidad y sensibilidad cultural de estas estrategias. Afrontar estos desafíos requiere una articulación interdisciplinaria entre profesionales de la salud, comunicadores, científicos del comportamiento y actores de la sociedad civil, orientada a fortalecer la capacidad colectiva de respuesta frente a uno de los retos más complejos de la era digital.

Contribución de autoría

David Gómez-Valencia cumplió con todas las fases CRediT.

Fuente de financiamiento

Autofinanciado.

Potenciales conflictos de interés

Ninguno.

Declaración de uso de IA

Durante la preparación de este manuscrito, el autor utilizó la IA únicamente para mejorar la claridad, la gramática, la morfosintaxis y la concisión del texto, así como para sintetizar contenido proveniente de la tesis de maestría sobre la que se basa este manuscrito. En ningún caso se utilizó como fuente directa de información.

Agradecimientos

A Carla Maldonado, quien se adentró en los manuscritos para darme recomendaciones tanto cuidadosas como certeras, y a Susana Torres, por su apoyo en el desarrollo del trabajo de grado.

REFERENCIAS BIBLIOGRÁFICAS

- Al Khaja, K. A. J., AlKhaja, A. K. y Sequeira, R. P. (2018). Drug information, misinformation, and disinformation on social media: A content analysis study. *Journal of Public Health Policy*, 39(3), 343-357. <https://doi.org/10.1057/s41271-018-0131-2>
- Ardèvol-Abreu, A., Gil de Zúñiga, H. y McCombs, M. E. (2020). Orígenes y desarrollo de la teoría de la agenda setting en Comunicación. Tendencias en España (2014-2019). *Profesional de la información*, 29(4), 1-23. <https://doi.org/10.3145/epi.2020.jul.14>
- Aruguete, N. y Calvo, E. (2023). *Nosotros contra ellos. Cómo trabajan las redes para confirmar nuestras creencias y rechazar las de los otros*. Siglo XXI.
- Bautista, J. R., Zhang, Y. y Gwizdka, J. (2021). Healthcare professionals' acts of correcting health misinformation on social media. *International Journal of Medical Informatics*, 148, 104375. <https://doi.org/10.1016/j.ijmedinf.2021.104375>
- Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J. y Crotty, K. (2011). Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine*, 155(2), 97-107. <https://doi.org/10.7326/0003-4819-155-2-201107190-00005>
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A. y Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313-7318. <https://doi.org/10.1073/pnas.1618923114>
- Castiel, L. D. y SanzValero, J. (2010). Access to information as a social determinant of health. *Nutrición Hospitalaria*, 25(Supl. 3), 26-30. http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0212-16112010000900004&lng=es&tlng=en
- Chou, W. Y. S., Gaysynsky, A. y Cappella, J. N. (2020). Where we go from here: Health misinformation on social media. *American Journal of Public Health*, 110(S3), S273-S275. <https://doi.org/10.2105/AJPH.2020.305905>
- Díaz, L., Martínez Villarreal, D., Márquez, K. y Scartascini, C. (2025). Combating vaccine hesitancy: The case of HPV vaccination. *Social Science & Medicine*, 381, 118081. <https://doi.org/10.1016/j.socscimed.2025.118081>
- Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R. y Bettinger, J. (2013). Vaccine hesitancy: An overview. *Human Vaccines & Immunotherapeutics*, 9(8), 1763-1773. <https://doi.org/10.4161/hv.24657>
- Duffy, A., Tandoc, E. y Ling, R. (2020). Too good to be true, too good not to share: The social utility of fake news. *Information, Communication & Society*, 23(13), 1965-1979. <https://doi.org/10.1080/1369118X.2019.1623904>

- Feng, B. (2024). Gaming with health misinformation: a social capital-based study of corrective information sharing factors in social media. *Frontiers in Public Health*, 12, 1351820. <https://doi.org/10.3389/fpubh.2024.1351820>
- Foro Económico Mundial (FEM). (2018). *The global risks report 2018*. (13.ª ed.). <https://www.weforum.org/reports/the-global-risks-report-2018>
- Gómez Valencia, D., Isaza Duarte, A. C. y Pineda Alarcón, S. (2025). *Entender y reducir la difusión de noticias falsas sobre vacunas en redes sociales: Un enfoque desde los estudios del comportamiento*. [Tesis de maestría, Universidad EAFIT].
- Graham, J., Haidt, J. y Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029-1046. <https://doi.org/10.1037/a0015141>
- Hernández S, R., Fernández C, C. y Baptista L., P. (2022). *Metodología de la investigación*. (7.ª ed.). McGraw-Hill Education.
- Kozyreva, A., Lorenz-Spreen, P., Herzog, S. M., Ecker, U. K. H., Lewandowsky, S., Hertwig, R., Ali, A., Bak-Coleman, J., Barzilai, S., Basol, M., Berinsky, A. J., Betsch, C., Cook, J., Fazio, L. K., Geers, M., Guess, A. M., Huang, H., Larreguy, H., Maertens, R., Panizza, F., Pennycook, G., Rand, D. G., Rathje, S., Reifler, J., Schmid, P., Smith, M., Swire-Thompson, B., Szewach, P., van der Linden, S. y Wineburg, S. (2024). Toolbox of individual-level interventions against online misinformation. *Nature Human Behaviour*, 8(6), 1044-1052. <https://doi.org/10.1038/s41562-024-01881-0>
- List, J. A., Ramirez, L. M., Seither, J., Unda, J. y Vallejo, B. H. (2024). Critical thinking and misinformation vulnerability: Experimental evidence from Colombia, *PNAS Nexus*, 3(10), 361. <https://doi.org/10.1093/pnasnexus/pgae361>
- MacDonald, N. E. (2015). Vaccine hesitancy: Definition, scope and determinants. *Vaccine*, 33(34), 4161-4164. <https://doi.org/10.1016/j.vaccine.2015.04.036>
- Metzger, M. J., Flanagin, A. J. y Medders, R. B. (2021). Social and heuristic approaches to credibility evaluation online. *Media and Communication*, 9(1), 189-199. <https://doi.org/10.17645/mac.v9i1.3409>
- Michie, S., Atkins, L. y West, R. (2014). *The Behavior Change Wheel. A Guide to Designing Interventions*. Silverback Publishing.
- Ministerio de Salud y Protección Social. (2024). *Lineamientos para la gestión y administración del Programa Ampliado de Inmunizaciones - PAI - 2024*. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/PP/PAI/lineamiento-gestion-administracion-pai-2024.pdf>
- Mohammad, S. M. y Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436-465. <https://doi.org/10.1111/j.1467-8640.2012.00460.x>

Morejón-Llamas, N. (2023). Características y ejes discursivos de la desinformación y el proceso de fact-checking sobre las vacunas COVID-19 en Latinoamérica. *Revista Española de Comunicación en Salud*, 14, 47-61. <https://doi.org/10.20318/recs.2023.7005>

Murad, Z. (2024). Confidence and relative performance. SSRN. <https://doi.org/10.2139/ssrn.4986692>

Murillo, A. (2020). Desconfianza, regímenes de verdad, conspiraciones y montajes en el contexto de la covid-19 en México. *Desde el Sur*, 12(2), 547-571. <https://revistas.cientifica.edu.pe/index.php/desdeelsur/article/view/767/733>

Nan, X., Wang, Y. y Thier, K. (2022). Why people believe health misinformation and who are at risk? A systematic review of individual differences in susceptibility to health misinformation. *Social Science & Medicine*, 315, 115398. <https://doi.org/10.1016/j.socscimed.2022.115398>

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K. y Nielsen, R. K. (2023). *Reuters Institute Digital News Report 2023*. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023>

Norman, C. D. y Skinner, H. A. (2006). eHEALS: The eHealth literacy scale. *Journal of Medical Internet Research*, 8(4), e27. <https://doi.org/10.2196/jmir.8.4.e27>

Organización Panamericana de la Salud (OPS). (2019). Inmunización en las Américas. Resumen 2019. <https://www.paho.org/es/documentos/inmunizacion-americas-resumen-2019>

Peng, W., Lim, S. y Meng, J. B. (2022). Persuasive strategies in online health misinformation: A systematic review. *Information, Communication & Society*, 26(11), 2131-2148. <https://doi.org/10.1080/1369118X.2022.2085615>

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. y Rand, D. (2020). Fighting COVID-19 Misinformation on social media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, 31, 770-780. <https://doi.org/10.1177/0956797620939054>

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. S., Roberts, M. E., Shariff, A., Tenenbaum, J. B. y Wellman, M. (2019). Machine behaviour. *Nature*, 568(7753), 477-486. <https://doi.org/10.1038/s41586-019-1138-y>

Roozenbeek, J., Culloty, E. y Suiter, J. (2023). Countering Misinformation. Evidence, Knowledge Gaps, and Implications of Current Interventions. *European Psychologist*, 28(3), 135-224. <https://doi.org/10.1027/1016-9040/a000492>

- Sultan, M., Tump, A. N., Ehmann, N., LorenzSpreen, P., Hertwig, R., Gollwitzer, A. y Kurvers, R. H. J. M. (2024). Susceptibility to online misinformation: A systematic meta-analysis of demographic and psychological factors. *Proceedings of the National Academy of Sciences of the United States of America*, 121(47), e2409329121. <https://doi.org/10.1073/pnas.2409329121>
- Traberg, C. S., Roozenbeek, J. y van der Linden, S. (2022). Psychological Inoculation against Misinformation: Current Evidence and Future Directions. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 136-151. <https://doi.org/10.1177/00027162221087936>
- Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C. y Tucker, J. A. (2021). Political psychology in the digital (mis)information age: A model of news belief and sharing. *Social Issues and Policy Review*, 15(1), 84-113. <http://dx.doi.org/10.1111/sipr.12077>
- Vosoughi, S., Roy, D. y Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. <https://doi.org/10.1126/science.aap9559>
- Wardle, C. y Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe. <https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77>
- Xue, X., Ma, H., Zhao, Y. C., Zhu, Q. y Song, S. (2024). Mitigating the influence of message features on health misinformation sharing intention in social media. *Social Science & Medicine*, 356, 117136. <https://doi.org/10.1016/j.socscimed.2024.117136>

David Gómez-Valencia es sociólogo y magíster en Estudios del Comportamiento. Su trayectoria interdisciplinaria abarca la investigación aplicada, diseño de intervenciones de cambio cultural y comportamental en áreas de salud, aprendizaje social e innovación tecnológica. Ha sido galardonado con el Premio Nacional a la Mejor Trabajo de Grado de la Universidad Nacional de Colombia, autor y ponente en temas de cuerpo, emociones y cultura.

Recepción: 31/8/2025
Aceptación: 13/7/2026