



Artículo de revisión

**Heterogeneidad en las estimaciones de prevalencia de
síntomas depresivos con PHQ-8/PHQ-9: revisión
sistemática y metaanálisis**

**Heterogeneity in Prevalence Estimates of Depressive
Symptoms Measured with PHO-8/ PHO-9: A Systematic
Review and Meta-Analysis**

Alberto Guevara Tirado¹  

Gerard Guardia Barra¹ 

Wilson Leonardo Nolasco Elias¹ 

Jazmin Camila Borja Tineo¹ 

¹Facultad de Medicina Humana, Universidad Científica del Sur, Lima, Perú

Recibido: 14/04/2026

Aceptado: 13/05/2026

Resumen

Introducción: las estimaciones poblacionales de síntomas depresivos obtenidas con instrumentos de tamizaje difieren por instrumento, punto de corte, diseño muestral, contexto y periodo de recolección, lo que limita su comparabilidad.

Objetivo: cuantificar y explorar la heterogeneidad de la prevalencia de tamizaje positivo para síntomas depresivos en población general mediante PHQ-8 o PHQ-9.

Método: revisión sistemática según PRISMA 2020. Se incluyeron muestras poblacionales no clínicas, probabilísticas o representativas, con PHQ-8/PHQ-9 ≥ 10 , sin restricción por periodo de recolección. Se aplicó metaanálisis de proporciones con transformación logit, REML, ajuste Hartung-Knapp e intervalo de predicción. Se realizaron subgrupos y metarregresiones por instrumento, periodo y características del estudio.

Resultados: se incluyeron 20 estimaciones no superpuestas (N=651 509). La prevalencia combinada fue 10,40 % (IC95%: 8,08-13,31; $I^2=99,5$ %; $\tau^2=0,351$), con intervalo de predicción de 3,14-29,38 %. Las estimaciones prepandémicas fueron 7,78 % (IC95%: 6,59-9,16; k=14) y las obtenidas desde marzo de 2020, 19,68 % (IC95%: 14,14-26,72; k=6). El periodo se asoció con la prevalencia en metarregresión ($p<0,001$) y explicó 70,3 % de la varianza entre estudios, aunque persistió heterogeneidad elevada dentro de ambos periodos.

Conclusiones: el periodo de recolección fue una fuente importante de heterogeneidad. La variabilidad residual respalda una interpretación específica del contexto y desaconseja considerar PHQ-8/PHQ-9 ≥ 10 como prevalencia diagnóstica universal de depresión.

Palabras clave: depresión; cuestionario de salud del paciente; prevalencia; tamizaje masivo; salud mental; metaanálisis como asunto

Abstract

Introduction: Population estimates of depressive symptoms obtained with screening instruments differ by instrument, cut-off point, sampling design, context and collection period, which limits their comparability.

Objective: Quantify and explore the heterogeneity of the prevalence of the positive screening for depressive symptoms in the general population measured with PHQ-8 or PHQ-9.

Methods: Systematic review according to PRISMA 2020. Non-clinical, probabilistic or representative population samples using PHQ-8/PHQ-9 ≥ 10 were included without restriction of collection period. Meta-analyses of proportions using logit transformation, REML, Hartung-Knapp adjustment and prediction interval were applied. Subgroups and meta-regressions analyses were conducted by instrument, period and characteristics of the study.

Results: Twenty non-overlapping estimates (N=651,509) were included. The combined prevalence was 10.40% (95%CI: 8.08- 13.31; $I^2=99.5$ %; $\tau^2=0.351$), with a prediction interval of 3.14-29.38%. Pre-pandemic estimates were 7.78% (95%CI: 6.59-9.16; K=14),

and those from March 2020 onwards were 19.68% (95%CI:14.14-26.72; K=6). The period was associated with prevalence in meta-regression ($P < 0.001$), and explained 70.3% of the variance between studies, although high heterogeneity persisted within both periods.

Conclusions: The data collection period was an important source of heterogeneity. The residual variability supports a context-specific interpretation and advises against considering PHQ-8/PHQ-9 ≥ 10 as a universal diagnostic prevalence of depression.

Keywords: Depression; Patient Health Questionnaire; Prevalence; Mass Screening; Mental Health; Meta-Analysis as Topic

Introducción

Los trastornos depresivos contribuyen de forma importante a la carga mundial de enfermedad y discapacidad.^(1,2) En epidemiología poblacional, sin embargo, una estimación derivada de un instrumento de tamizaje no equivale necesariamente a la prevalencia de un trastorno depresivo diagnosticado mediante entrevista clínica. En particular, el uso de PHQ-9 ≥ 10 (Patient Health Questionnaire) puede sobreestimar la prevalencia de depresión mayor cuando se compara con entrevistas estructuradas,⁽³⁾ y análisis poblacionales europeos han mostrado que corregir la precisión diagnóstica del PHQ-8 reduce sustancialmente la estimación atribuible a depresión mayor.⁽⁴⁾

La comparación entre encuestas también depende del instrumento, el punto de corte, el modo de administración, el marco temporal y el diseño de muestreo. Incluso con la misma familia PHQ, el modo de entrevista puede modificar la prevalencia observada,⁽⁵⁾ aunque análisis de funcionamiento diferencial sugieren que el efecto del modo de administración sobre la puntuación total puede ser pequeño en algunos contextos.⁽⁶⁾ La estructura interna, la invariancia de medida, las normas poblacionales y la adaptación de instrumentos de tamizaje han sido examinadas en muestras amplias y diversos contextos, lo que refuerza la necesidad de considerar las propiedades de medición al comparar prevalencias.⁽⁷⁻¹²⁾ Estudios poblacionales recientes han documentado, además, variabilidad geográfica y temporal aun con instrumentos de la familia PHQ y criterios comparables.⁽¹³⁻¹⁶⁾

Ante la heterogeneidad atribuible a diferencias metodológicas y temporales, el presente estudio tuvo como objetivo estimar y explorar la variabilidad de la prevalencia de tamizaje positivo para síntomas depresivos en población general mediante PHQ-8 o PHQ-9 con punto de corte ≥ 10 , se consideró el periodo de recolección como fuente potencial de heterogeneidad. Se espera que este enfoque contribuya a mejorar la comparabilidad de las estimaciones poblacionales y a orientar la estandarización de la medición de síntomas depresivos en estudios epidemiológicos y sistemas de vigilancia.

Método

Diseño y reporte

Se realizó una revisión sistemática con metaanálisis de estudios de prevalencia. El reporte se estructuró según PRISMA 2020, su documento de explicación y elaboración, y PRISMA-S.⁽¹⁷⁻¹⁹⁾ La síntesis cuantitativa tuvo un propósito principalmente exploratorio, orientado a caracterizar y explicar la heterogeneidad entre estimaciones, no a establecer una prevalencia universal.

Registro y protocolo

La revisión no fue registrada prospectivamente. Al no haberse depositado un protocolo en un registro público antes de iniciar la selección de estudios, no se realizó un registro retrospectivo, ya que este no habría documentado prospectivamente las decisiones metodológicas. No existió un protocolo público previo y, en consecuencia, no hubo desviaciones respecto de un protocolo registrado.

Fuentes de información y estrategia de búsqueda

La búsqueda sistemática se realizó el 10 de agosto de 2026 en PubMed/MEDLINE, Embase, Scopus, PsycINFO y LILACS, y se complementó con Google Scholar y la revisión manual de referencias. La identificación total comprendió 23 705 registros, correspondientes a 1 552 en PubMed/MEDLINE, 9 153 en Embase, 10 555 en Scopus, 1 570 en PsycINFO y 875 en LILACS. Las sintaxis se muestran en el Material suplementario S1.

Criterios de elegibilidad

Se incluyeron estudios observacionales que: (1) evaluaran población general adulta no clínica, con muestreo probabilístico o diseño explícitamente representativo; (2) utilizaran PHQ-8 o PHQ-9; (3) definieran tamizaje positivo mediante punto de corte ≥ 10 ; y (4) reportaran prevalencia o datos suficientes para derivarla. No se aplicó restricción por periodo de recolección. Se excluyeron muestras clínicas o de grupos ocupacionales/específicos, instrumentos distintos del PHQ-8/PHQ-9, puntos de corte diferentes, algoritmos diagnósticos no equivalentes, muestras opt-in no representativas y estimaciones superpuestas cuando ya se había retenido una estimación de la misma muestra.

Selección y extracción de datos

Los registros se cribaron por título y resumen y posteriormente a texto completo. La selección y la extracción se realizaron mediante criterios preespecificados y un formulario estandarizado; antes de la síntesis se efectuó una segunda verificación de elegibilidad, duplicación de muestras y consistencia de los datos extraídos. Se recopilaron país, periodo de recolección, tamaño muestral, representatividad, instrumento, punto de corte, modo de administración, prevalencia e IC95 %. Cuando existían varias publicaciones sobre la misma encuesta se compararon periodos y marcos muestrales para evitar doble conteo.

Riesgo de sesgo

El riesgo de sesgo se evaluó con la JBI Critical Appraisal Checklist for Studies Reporting Prevalence Data, que examina marco muestral, muestreo, tamaño de muestra, descripción de participantes, cobertura analítica, validez de la medición, estandarización, análisis y respuesta.⁽²⁰⁾ No se utilizó una suma de puntos como sustituto del juicio por dominios; los elementos no verificables se clasificaron como inciertos.

Análisis estadístico

La medida de síntesis fue la proporción de participantes con PHQ-8/PHQ-9 ≥ 10 . Las proporciones se transformaron a escala logit y se combinaron mediante un modelo de efectos aleatorios; τ^2 se estimó por máxima verosimilitud restringida (REML) y los intervalos de confianza del efecto promedio se calcularon con ajuste de Hartung-Knapp. Cuando el estudio publicó un IC95 % del estimador ponderado, su error estándar se derivó en escala logit a partir de ese intervalo; cuando no se dispuso de un IC95% verificable, se utilizó una aproximación binomial basada en n y prevalencia. Se calcularon Q , I^2 , τ^2 e intervalo de predicción del 95%. Como sensibilidad, se repitió el análisis con varianzas binomiales para todas las estimaciones.

Se realizaron subgrupos por PHQ-8 frente a PHQ-9 y por periodo de recolección (prepandémico frente a datos obtenidos desde marzo de 2020), además de meta regresiones exploratorias por versión del instrumento, periodo, año de recolección y alcance subnacional frente a nacional/multinacional. Se efectuaron modelos conjuntos para valorar la contribución independiente de estas variables, así como análisis leave-one-out. Los small-study effects se exploraron mediante gráfico de embudo y una regresión del efecto en escala logit sobre su error estándar; este análisis se consideró exploratorio por la heterogeneidad extrema y las particularidades de los metaanálisis de prevalencia. Todos los análisis

estadísticos se realizaron en Python 3.13, con las bibliotecas NumPy, pandas, SciPy y statsmodels; las figuras se generaron con Matplotlib.

Certeza de la evidencia

La confianza en la estimación global se valoró con un enfoque GRADE informado por dominios de riesgo de sesgo, inconsistencia, indirectitud, imprecisión y posible sesgo de publicación. Dado que el desenlace es una positividad de tamizaje y no una prevalencia diagnóstica, la interpretación se centró en la certeza sobre la proporción de tamizaje positivo.

Consideraciones éticas

El estudio utiliza exclusivamente datos agregados publicados y no contiene información individual identificable; por ello no requirió aprobación de un comité de ética ni consentimiento informado.

Resultados

Selección y características

La búsqueda en PubMed/MEDLINE, Embase, Scopus, PsycINFO y LILACS identificó 23 705 registros. Tras eliminar 8 746 duplicados, se cribaron 14 959 registros y se excluyeron 14 686 por título y resumen. Se buscaron 273 informes para recuperación y 18 no pudieron recuperarse. Se evaluaron 255 textos completos. Finalmente, 20 estudios que aportaron 20 estimaciones no superpuestas cumplieron los criterios de elegibilidad y se incluyeron en la revisión sistemática y en el metaanálisis (figura 1; tabla I; material suplementario S2).

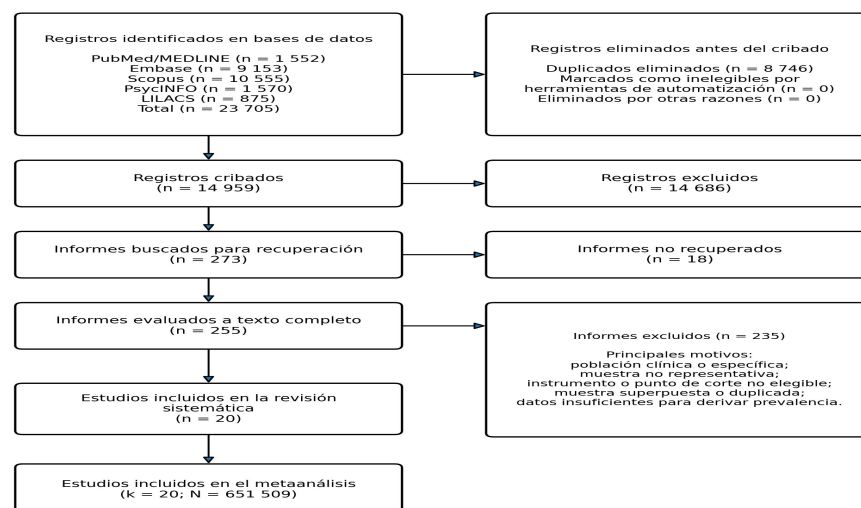


Figura 1. Diagrama de flujo de identificación, cribado, elegibilidad e inclusión de estudios según PRISMA 2020

Tabla 1. Características de las 20 estimaciones incluidas en el análisis cuantitativo

Estudio	País/área	Instrumento	Fuente/período	N	Prev. %	IC95%	Alcance
Kroenke et al., 2009 ⁽²¹⁾	Estados Unidos	PHQ-8	BRFSS 2006	198 678	8,60	8,48-8,72†	Nacional
Li et al., 2012 ⁽⁵⁾	Estados Unidos	PHQ-8	NHANES 2005-2006	4 800	5,15	4,56-5,81†	Nacional
Kocalevent et al., 2013 ⁽²²⁾	Alemania	PHQ-9	Encuesta 2003-2008	5 018	5,60	5,00-6,27†	Nacional
Kader Maideen et al., 2014 ⁽²³⁾	Malasia (Selangor)	PHQ-9	Encuesta 2012	1 460	10,30	8,74-11,86	Subnacional
Shidhaye et al., 2016 ⁽²⁴⁾	India (Vidarbha)	PHQ-9	Encuesta 2014	1 516	14,60	12,80-16,40	Subnacional
Shin et al., 2017 ⁽²⁵⁾	Corea del Sur	PHQ-9	KNHANES 2014	4 949	6,70	5,70-7,60	Nacional
Kepler et al., 2018 ⁽²⁶⁾	EE.UU. (Nueva York)	PHQ-9	NYC HANES 2013-2014	1 459	8,30	6,80-10,10	Subnacional
Khaled, 2019 ⁽²⁷⁾	Qatar	PHQ-9	Encuesta 2017	2 424	6,60	5,40-7,90	Nacional
Ssebunnya et al., 2019 ⁽²⁸⁾	Uganda (Kamuli)	PHQ-9	Encuesta rural 2015	1 290	6,40	5,19-7,87†	Subnacional
Cao et al., 2020 ⁽²⁹⁾	Estados Unidos	PHQ-9	NHANES 2015-2018	10 257	8,00	7,49-8,54†	Nacional
Arias-de la Torre	Europa (27 países)	PHQ-8	EHIS-2 2014-	258	6,38	6,24-	Multinacional

et al., 2021 ⁽³⁰⁾				2015	888		6,52	
Shields et al., 2021 (CCHS) ⁽³¹⁾	Canadá (8 provincias)	PHQ-9	CCHS 2015-2019	31	6,70	6,20-7,20	Nacional parcial	
Melo et al., 2023 [32]	Brasil	PHQ-9	PNS 2019	88	10,80	10,40-11,00	Nacional	
Xu et al., 2022 ⁽³³⁾	China (Guangdong)	PHQ-9	GSPHS 2019	13	8,60	8,00-9,30	Subnacional	
Ettman et al., 2020 ⁽³⁴⁾	Estados Unidos	PHQ-9	CLIMB 2020	1	27,80	24,90-30,90	Nacional	
Dawel et al., 2020 ⁽³⁵⁾	Australia	PHQ-9	COVID-MHBRC 2020	1	20,30	18,11-22,49†	Nacional	
Shields et al., 2021 (SCMH) ⁽³¹⁾	Canadá	PHQ-9	SCMH 2020	12	15,20	14,20-16,20	Nacional	
Hajek y König, 2021 ⁽³⁶⁾	Alemania	PHQ-9	Encuesta 2021	3	20,00	18,59-21,41†	Nacional	
Craig et al., 2022 ⁽³⁷⁾	Sudáfrica	PHQ-9	Encuesta 2021	3	25,70	24,23-27,17†	Nacional	
Jiang et al., 2025 ⁽³⁸⁾	Estados Unidos	PHQ-9	NHANES 2021-2023	4	12,40	10,60-14,10	Nacional	

Nota: IC95%: intervalo de confianza del 95%; PHQ: Patient Health Questionnaire; BRFSS: Behavioral Risk Factor Surveillance System; NHANES: National Health and Nutrition Examination Survey; KNHANES: Korea National Health and Nutrition Examination Survey; NYC HANES: New York City Health and Nutrition Examination Survey; EHIS: European Health Interview Survey; CCHS: Canadian Community Health Survey; GSPHS: Guangdong Sleep and Psychosomatic Health Survey; PNS: Pesquisa Nacional de Saúde; CLIMB: COVID-19 and Life Stressors Impact on Mental Health and Well-being; COVID-MHBRC: Australian National COVID-19 Mental Health, Behavior and Risk Communication Survey; SCMH: Survey on COVID-19 and Mental Health. El símbolo † identifica IC95% aproximados mediante varianza binomial cuando la fuente no proporcionó el intervalo del estimador ponderado.

Fuente: elaboración propia

Síntesis cuantitativa

Las 20 estimaciones comprendieron 651 509 participantes y oscilaron entre 5,15 % y 27,8 %. El modelo REML con Hartung-Knapp estimó una prevalencia combinada de 10,40 % (IC95%: 8,08-13,31), con heterogeneidad extrema ($Q=3647,6$; $gl=19$; $p<0,001$; $I^2=99,5$ %; $\tau^2=0,351$) e intervalo de predicción del 95% de 3,14-29,38 %. En el subgrupo prepandémico, la prevalencia combinada fue 7,78 % (IC95%: 6,59-9,16; $k=14$; $N=624$ 958; $I^2=98,9$ %), mientras que en las estimaciones recolectadas desde marzo de 2020 fue 19,68 % (IC95%: 14,14-26,72; $k=6$; $N=26$ 551; $I^2=97,7$ %). En metarregresión, el periodo se asoció con mayor prevalencia (coeficiente logit=1,067; $p<0,001$) y redujo τ^2 de 0,351 a 0,104, equivalente a 70,3 % de la varianza entre estudios. El año de recolección también mostró asociación en un modelo univariable (coeficiente logit=0,075 por año; $p=0,001$). En el modelo conjunto con

periodo y versión del PHQ, el periodo mantuvo su asociación ($p<0,001$), mientras que PHQ-9 frente a PHQ-8 no mostró una asociación estadísticamente clara ($p=0,297$). En un modelo adicional, el alcance subnacional tampoco mostró asociación estadísticamente clara ($p=0,177$). La sensibilidad con varianzas binomiales produjo 10,40 % (IC95%: 8,08-13,31), y en el leave-one-out la prevalencia osciló entre 9,84 % y 10,78 %, con I^2 persistentemente >99 % (tabla 2; figura 2).

Tabla 2. Metaanálisis y análisis de heterogeneidad

Análisis	k	Prevalencia % (IC95%)	I^2 %	τ^2	IP95%
Principal	20	10,40 (8,08-13,31)	99,5	0,351	3,14-29,38
Periodo prepandémico	14	7,78 (6,59-9,16)	98,9	0,091	4,09-14,30
Periodo desde marzo de 2020	6	19,68 (14,14-26,72)	97,7	0,139	7,41-42,88
PHQ-8	3	6,60 (3,44-12,29)	99,6	0,074	0,13-79,50
PHQ-9	17	11,27 (8,54-14,72)	99,2	0,352	3,34-31,84
Varianza binomial (sensibilidad)	20	10,40 (8,08-13,31)	99,7	0,352	3,14-29,41
Leave-one-out	19 por iteración	9,84-10,78*	99,3-99,5	0,288-0,371	Ver Tabla S3

Nota: k: número de estimaciones; IC95%: intervalo de confianza del 95%; I^2 : porcentaje de heterogeneidad; τ^2 : varianza entre estudios; IP95%: intervalo de predicción del 95%; PHQ: Patient Health Questionnaire; REML: máxima verosimilitud restringida. El contraste por periodo se definió como recolección antes de marzo de 2020 frente a recolección desde marzo de 2020. *Rango de las prevalencias combinadas obtenidas al excluir secuencialmente cada estudio. El IP95% del subgrupo PHQ-8 debe interpretarse con cautela debido al reducido número de estudios ($k=3$) y a la heterogeneidad elevada. Fuente: elaboración propia

Los resultados completos del análisis leave-one-out se presentan en la tabla 3.

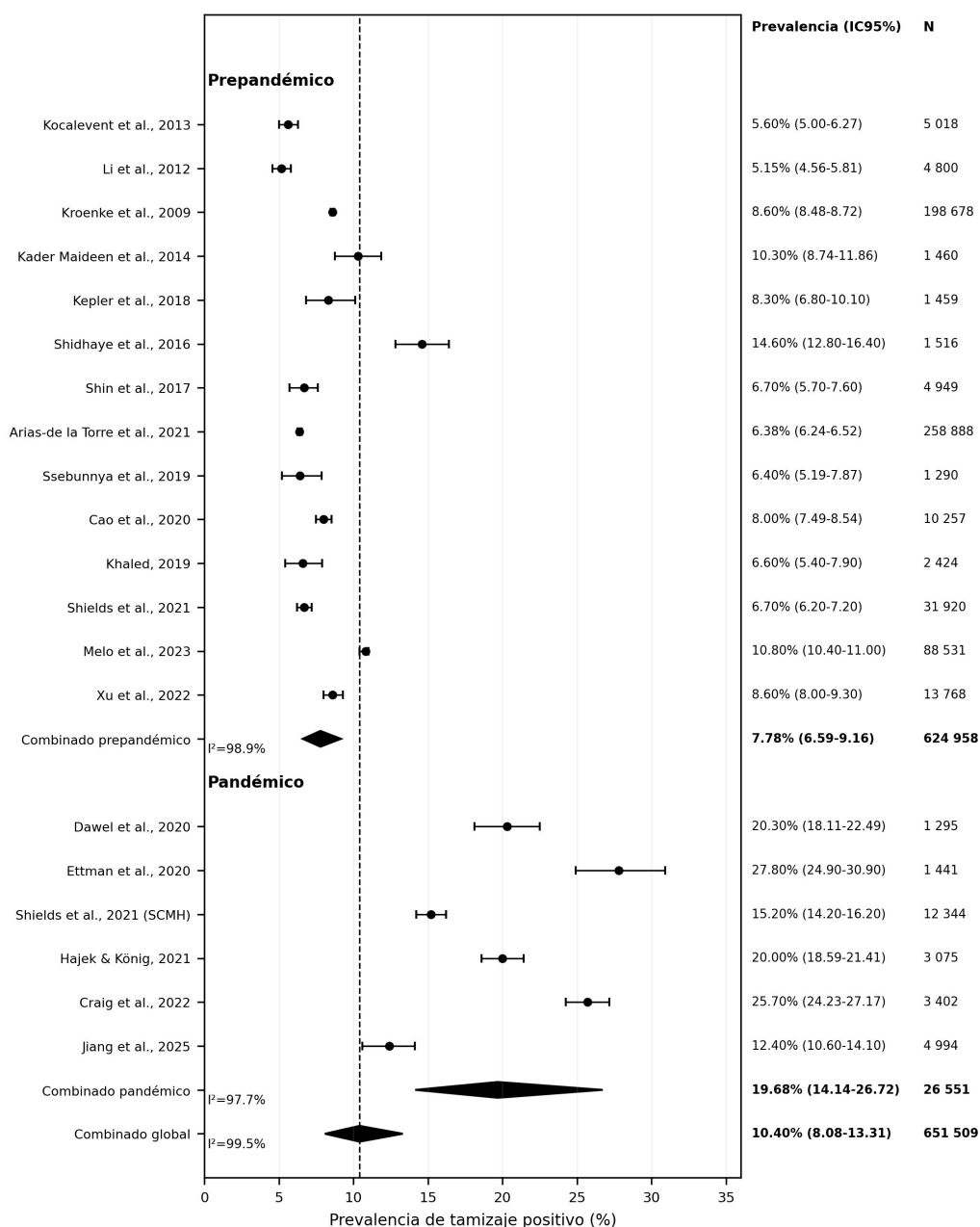


Figura 2. Forest pilot de las estimaciones de prevalencia de tamizaje positivo con PHQ-8/PHQ-9 ≥ 10 , ordenadas por periodo de recolección
 Fuente: elaboración propia

Riesgo de sesgo y certeza

La evaluación individual con los dominios JBI mostró riesgo de sesgo bajo en 10 estudios y algunas preocupaciones en 10, principalmente por cobertura subnacional, modalidad de muestreo o respuesta. No se identificaron estudios con limitaciones globales suficientes para clasificarlos como riesgo alto (tabla 3).

Tabla III. Evaluación del riesgo de sesgo de los estudios incluidos con dominios JBI

Estudio	Juicio global	Fundamento principal
Kroenke et al., 2009	Bajo	BRFSS nacional, gran tamaño muestral, instrumento estandarizado y ponderación poblacional.
Li et al., 2012	Bajo	NHANES con muestreo probabilístico complejo, procedimientos estandarizados y ponderación.
Kocalevent et al., 2013	Algunas preocupaciones	Muestra nacional representativa; información limitada para algunos dominios de respuesta y cobertura.
Kader Maideen et al., 2014	Algunas preocupaciones	Muestreo aleatorio multietápico, pero cobertura subnacional y respuesta moderada.
Shidhaye et al., 2016	Algunas preocupaciones	Encuesta poblacional subnacional; generalización territorial limitada.
Shin et al., 2017	Bajo	KNHANES nacional con muestreo probabilístico, estandarización y ponderación.
Kepler et al., 2018	Algunas preocupaciones	NYC HANES ponderada; respuesta global relativamente baja aunque ajustada por diseño.
Khaled, 2019	Algunas preocupaciones	Muestra poblacional probabilística con posibles limitaciones de cobertura y no respuesta.
Ssebunnya et al., 2019	Algunas preocupaciones	Muestra poblacional rural y subnacional; representatividad geográfica limitada.
Cao et al., 2020	Bajo	NHANES nacional probabilística, ponderada y con medición estandarizada.
Arias-de la Torre et al., 2021	Bajo	EHIS multinacional representativa y ponderada; instrumento y umbral homogéneos.
Shields et al., 2021 (CCHS)	Algunas preocupaciones	Muestreo probabilístico y ponderado; la estimación comparable cubrió ocho provincias.
Xu et al., 2022	Algunas preocupaciones	Muestreo estratificado multietápico representativo de Guangdong; inferencia provincial.
Melo et al., 2023	Bajo	PNS nacional con muestreo probabilístico complejo y ponderación.
Ettman et al., 2020	Algunas preocupaciones	Panel probabilístico nacional ponderado; posibles limitaciones propias de panel y no respuesta.
Dawel et al., 2020	Algunas preocupaciones	Muestra por cuotas alineada con edad, sexo y región; no corresponde a muestreo probabilístico puro.
Shields et al., 2021 (SCMH)	Bajo	Muestreo aleatorio de viviendas, ponderación y cobertura nacional.
Hajek y König, 2021	Bajo	Encuesta nacional representativa con medición PHQ-9 estandarizada.
Craig et al., 2022	Bajo	Encuesta nacional representativa, ponderada y con PHQ-9 ≥ 10 preespecificado.
Jiang et al., 2025	Bajo	NHANES con muestreo probabilístico multietápico, ponderación y medición estandarizada.

Nota: JBI: Joanna Briggs Institute. Los juicios se basaron en los dominios de la lista para estudios de prevalencia y no se transformaron en una puntuación global numérica.

Fuente: elaboración propia

El gráfico de embudo mostró dispersión compatible con la heterogeneidad observada. La regresión exploratoria del efecto sobre el error estándar no detectó una asociación estadísticamente significativa ($p=0,736$). Este resultado no demuestra ausencia de sesgo de publicación, pues la heterogeneidad extrema y las grandes diferencias de tamaño y diseño limitan la interpretación. La evaluación GRADE indicó una certeza global muy baja para la estimación promedio, principalmente por inconsistencia muy seria e indirecta entre contextos y periodos (figura 3; tabla 4).

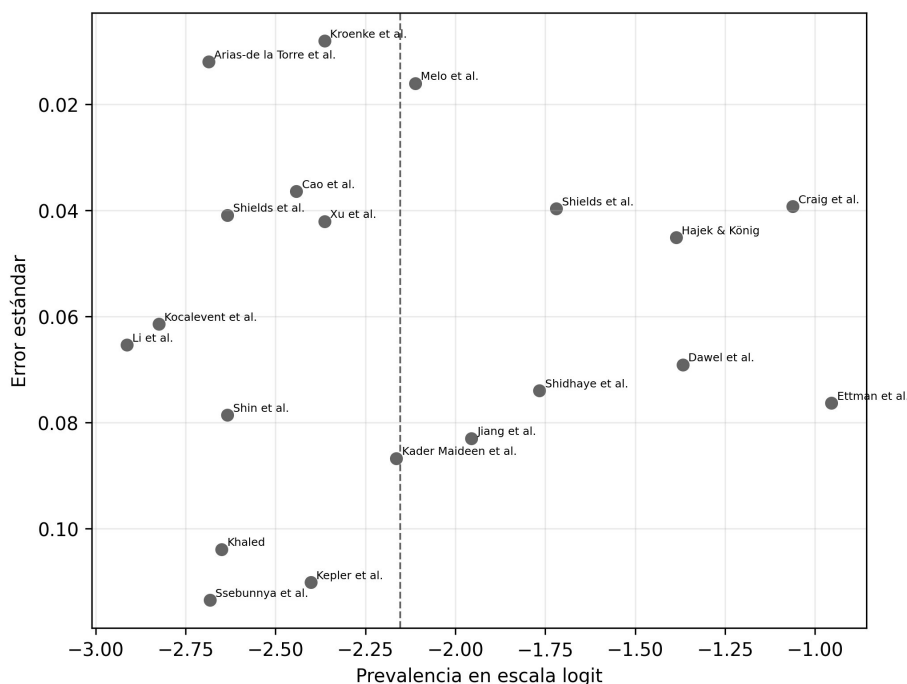


Figura 3. Gráfico de embudo exploratorio del metaanálisis

Fuente: elaboración propia

La asimetría debe interpretarse con cautela por la heterogeneidad extrema y las diferencias de tamaño y diseño entre encuestas

Tabla 4. Evaluación GRADE-informada de la confianza en la estimación global de tamizaje positivo

Dominio	Juicio	Justificación
Riesgo de sesgo	Serio	La mayoría son encuestas poblacionales sólidas, pero existen diferencias de respuesta/coertura y varios estudios son subnacionales.
Inconsistencia	Muy serio	$I^2=99,5\%$ e $IP95\%$ 3,14-29,38%; la heterogeneidad persiste dentro de los subgrupos por periodo e instrumento.
Indirectitud	Serio	Poblaciones, modos de administración, contextos socioculturales y periodos de recolección distintos; el desenlace es tamizaje positivo, no diagnóstico clínico.
Imprecisión	Serio a nivel predictivo	El IC del promedio es relativamente preciso, pero el intervalo de predicción es amplio y limita la transportabilidad.
Sesgo de publicación	No demostrado / incierto	Regresión exploratoria de small-study effects $p=0,736$; la heterogeneidad extrema limita la interpretación.
Certeza global	Muy baja	La estimación promedio no debe presentarse como prevalencia universal de depresión ni sustituir estimaciones contextuales.

Nota: GRADE: Grading of Recommendations Assessment, Development and Evaluation; IC95%: intervalo de confianza del 95%; IP95%: intervalo de predicción del 95%; I^2 : porcentaje de heterogeneidad; PHQ: Patient Health Questionnaire.

Fuente: elaboración propia

Discusión

El principal hallazgo fue que el periodo de recolección explica una proporción importante de la variabilidad entre estimaciones. La prevalencia combinada global fue 10,40 %, pero las estimaciones prepandémicas se agruparon en 7,78 %, frente a 19,68 % en las encuestas realizadas desde marzo de 2020. La metarregresión mostró una asociación clara con el periodo y una reducción aproximada de 70 % de la varianza entre estudios, lo que indica que el contexto temporal debe considerarse explícitamente al comparar prevalencias poblacionales. Aun así, la heterogeneidad permaneció elevada dentro de ambos periodos y el intervalo de predicción global fue amplio, por lo que ninguna cifra única representa adecuadamente todos los contextos.

El aumento observado en las estimaciones recolectadas desde el inicio de la pandemia es coherente con encuestas poblacionales representativas de Estados Unidos, Canadá, Australia, Alemania y Sudáfrica, que documentaron prevalencias de tamizaje positivo considerablemente mayores durante 2020-2021.^(31,34-37) Datos nacionales de Estados Unidos hasta 2023 también mostraron una prevalencia superior a la observada antes de la pandemia.⁽³⁸⁾ Sin embargo, estas diferencias temporales no deben interpretarse automáticamente como un efecto causal exclusivo de COVID-19, dado que también pueden intervenir cambios sociales, económicos, de administración de las encuestas y tendencias seculares.

La terminología es esencial. PHQ-8 y PHQ-9 son instrumentos de tamizaje y severidad; una puntuación ≥ 10 no debe denominarse automáticamente "depresión" o "trastorno depresivo mayor". Metaanálisis de exactitud diagnóstica han mostrado que las estimaciones basadas

en puntuaciones del PHQ pueden diferir de las obtenidas mediante entrevistas diagnósticas [3,39,40]. En EHIS, ajustar por la precisión diagnóstica del PHQ-8 redujo sustancialmente la estimación atribuible a depresión mayor.⁽⁴⁾ Por ello, el presente estudio utiliza "síntomas depresivos" y "tamizaje positivo" para el desenlace cuantitativo.

La coexistencia de estimaciones procedentes de distintas encuestas exige además evitar el doble conteo de muestras reutilizadas. EHIS-2, BRFSS, NHANES y otros sistemas poblacionales han generado múltiples publicaciones y periodos analíticos; por ello, solo se retuvieron estimaciones no superpuestas en participantes o periodos. Cuando una misma publicación aportó muestras independientes de distintos periodos, como en la comparación canadiense entre CCHS y SCMH, ambas pudieron incorporarse por corresponder a encuestas diferentes y no superpuestas.⁽³¹⁾

Desde la perspectiva de vigilancia, los resultados apoyan la estandarización del instrumento, el umbral, el periodo de referencia, el diseño muestral y la denominación del desenlace, pero muestran que esa estandarización no elimina la influencia del contexto temporal. Las comparaciones entre países o años deberían reportar explícitamente el periodo de recolección y evitar contrastar sin ajuste estimaciones obtenidas antes y durante grandes perturbaciones poblacionales. Cuando se sintetizan múltiples contextos, los intervalos de predicción y los análisis por periodo aportan información más útil que una prevalencia promedio aislada.

Entre las limitaciones, la heterogeneidad permaneció extrema incluso después de estratificar por periodo, lo que limita la transportabilidad del promedio global. La clasificación temporal en dos grupos resume una transición compleja y no separa fases específicas de la pandemia o del periodo posterior. Además, varias encuestas utilizaron ponderaciones complejas y algunas estimaciones requirieron aproximaciones de varianza. Las metarregresiones tienen potencia limitada con $k=20$ y deben interpretarse como exploratorias.

Conclusiones

Las estimaciones poblacionales de tamizaje positivo para síntomas depresivos con PHQ-8/PHQ-9 ≥ 10 mostraron heterogeneidad extrema. El periodo de recolección fue una fuente importante de variabilidad: las estimaciones prepandémicas fueron sustancialmente menores que las obtenidas desde marzo de 2020, aunque persistió heterogeneidad elevada dentro de ambos periodos. Por ello, la interpretación epidemiológica debe considerar conjuntamente instrumento, umbral, periodo temporal, contexto y diseño muestral, y evitar utilizar la prevalencia combinada como un valor universal o como equivalente a prevalencia diagnóstica de depresión.

Recomendaciones

Para encuestas poblacionales y vigilancia, los resultados no justifican priorizar PHQ-8 sobre PHQ-9. Cuando se utilice esta familia de instrumentos, se recomienda mantener constante la versión del PHQ, emplear el punto de corte ≥ 10 si se busca comparabilidad con la evidencia sintetizada, conservar el mismo periodo de referencia y modo de administración, utilizar ponderaciones del diseño complejo, reportar la positividad de tamizaje separadamente del diagnóstico, documentar tasas de respuesta, evitar el doble conteo de encuestas reutilizadas, estratificar o ajustar las comparaciones por periodo de recolección cuando existan cambios contextuales mayores y presentar intervalos de predicción en las síntesis entre poblaciones.

Referencias bibliográficas

1. Rong J, Wang X, Cheng P, Li D, Zhao D. Global, regional and national burden of depressive disorders and attributable risk factors, from 1990 to 2021: results from the 2021 Global Burden of Disease study. *Br J Psychiatry* [Internet]. 2025 [citado 10 ago 2026];227(4):688-697. Disponible en: <https://doi.org/10.1192/bjp.2024.266>
2. The Lancet Psychiatry. Global Burden of Disease 2021: mental health messages. *Lancet Psychiatry* [Internet]. 2024 [citado 10 ago 2026];11(8):573. Disponible en: [https://doi.org/10.1016/S2215-0366\(24\)00222-0](https://doi.org/10.1016/S2215-0366(24)00222-0)
3. Levis B, Benedetti A, Thombs BD; DEPRESSion Screening Data (DEPRESSD) Collaboration. Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: individual participant data meta-analysis. *J Clin Epidemiol* [Internet]. 2020 [citado 10 ago 2026];122:115-128.e1. Disponible en: <https://doi.org/10.1016/j.jclinepi.2020.02.002>
4. Fischer F, Zocholl D, Rauch G, Levis B, Benedetti A, Thombs B, et al. Prevalence estimates of major depressive disorder in 27 European countries from the European Health Interview Survey: accounting for imperfect diagnostic accuracy of the PHQ-8. *BMJ Ment Health* [Internet]. 2023 [citado 10 ago 2026];26:e300675. Disponible en: <https://doi.org/10.1136/bmjment-2023-300675>
5. Li C, Ford ES, Zhao G, Tsai J, Balluz LS. A comparison of depression prevalence estimates measured by the Patient Health Questionnaire with two administration modes: computer-assisted telephone interviewing versus computer-assisted personal interviewing. *Int J Public Health* [Internet]. 2012 [citado 10 ago 2026];57(1):225-233. Disponible en: <https://doi.org/10.1007/s00038-011-0253-9>
6. Harel D, Wu Y, Levis B, Fan S, Sun Y, Xu M, et al. Comparison of Patient Health Questionnaire-9, Edinburgh Postnatal Depression Scale and Hospital Anxiety and Depression-Depression subscale scores by administration mode: an individual participant

data differential item functioning meta-analysis. J Affect Disord [Internet]. 2024 [citado 11 ago 2026];361:674-683. Disponible en: <https://doi.org/10.1016/j.jad.2024.06.033>

7. Nouwen A, Deschênes SS, Balkhiyarova Z, Albertorio-Díaz JR, Prokopenko I, Schmitz N. Measurement invariance testing of the Patient Health Questionnaire-9 (PHQ-9) across people with and without diabetes mellitus from the NHANES, EMHS and UK Biobank datasets. J Affect Disord [Internet]. 2021 [citado 11 ago 2026];292:311-318. Disponible en: <https://doi.org/10.1016/j.jad.2021.05.031>

8. Bianchi R, Verkuilen J, Toker S, Schonfeld IS, Gerber M, Brähler E, et al. Is the PHQ-9 a unidimensional measure of depression? A 58,272-participant study. Psychol Assess [Internet]. 2022 [citado 11 ago 2026];34(6):595-603. Disponible en: <https://doi.org/10.1037/pas0001124>

9. Lee EH, Kang EH, Kang HJ, Lee HY. Measurement invariance of the Patient Health Questionnaire-9 depression scale in a nationally representative population-based sample. Front Psychol [Internet]. 2023 [citado 11 ago 2026];14:1217038. Disponible en: <https://doi.org/10.3389/fpsyg.2023.1217038>

10. Damiano RF, Hoffmann MS, Gosmann NP, Pan PM, Miguel EC, Salum GA. Translating measurement into practice: Brazilian norms for depressive symptom assessment with the Patient Health Questionnaire (PHQ-9). Braz J Psychiatry [Internet]. 2023 [citado 11 ago 2026];45(4):310-317. Disponible en: <https://doi.org/10.47626/1516-4446-2022-2945>

11. Cosci F, Christensen KS, Ceccatelli S, Patierno C, Carrozzino D. Patient Health Questionnaire-9: a clinimetric analysis. Braz J Psychiatry [Internet]. 2024 [citado 11 ago 2026];46:e20233449. Disponible en: <https://doi.org/10.47626/1516-4446-2023-3449>

12. Cjuno J, Moya A, Calderón-Pérez E, Quispe-Ilizarbe C, Mayon L, Livia J. Scientific output on validation and adaptation of depression screening instruments in Peruvian population. Rev Peru Med Exp Salud Publica [Internet]. 2022 [citado 11 ago 2026];39(3):357-361. Disponible en: <https://doi.org/10.17843/rpmesp.2022.393.11197>

13. Arias-de la Torre J, Vilagut G, Ronaldson A, Bakolis I, Dregan A, Martín V, et al. Prevalence and variability of depressive symptoms in Europe: update using representative data from the second and third waves of the European Health Interview Survey (EHIS-2 and EHIS-3). Lancet Public Health [Internet]. 2023 [citado 10 ago 2026];8(11):e889-e898. Disponible en: [https://doi.org/10.1016/S2468-2667\(23\)00220-7](https://doi.org/10.1016/S2468-2667(23)00220-7)

14. Lopes CS, Gomes NL, Junger WL, Menezes PR. Trend in the prevalence of depressive symptoms in Brazil: results from the Brazilian National Health Survey 2013 and 2019. Cad

Saude Publica [Internet]. 2022 [citado 10 ago 2026];38. Disponible en: <https://doi.org/10.1590/0102-311X00123421>

15. Villarreal-Zegarra D, Cabrera-Alva M, Carrillo-Larco RM, Bernabé-Ortiz A. Trends in the prevalence and treatment of depressive symptoms in Peru: a population-based study. BMJ Open [Internet]. 2020 [citado 10 ago 2026];10:e036777. Disponible en: <https://doi.org/10.1136/bmjopen-2020-036777>

16. Aydin K, Yildiz NG, Aydin HZ, Karaboğa HA, Güloğlu FK, Phiri YVA, et al. Prevalence of depressive symptoms and related factors in Türkiye: results of the 2016 and 2019 Turkish Health Survey. Turk Psikiyatri Derg [Internet]. 2025 [citado 10 ago 2026];36:23. Disponible en: <https://doi.org/10.5080/u27341>

17. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ [Internet]. 2021 [citado 10 ago 2026];372:n71. Disponible en: <https://doi.org/10.1136/bmj.n71>

18. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev [Internet]. 2021 [citado 10 ago 2026];10:39. Disponible en: <https://doi.org/10.1186/s13643-020-01542-z>

19. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ [Internet]. 2021 [citado 10 ago 2026];372:n160. Disponible en: <https://doi.org/10.1136/bmj.n160>

20. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Chapter 5: Systematic reviews of prevalence and incidence. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis [Internet]. Adelaide: JBI; 2020 [citado 10 ago 2026]. Disponible en: <https://jbi.global/critical-appraisal-tools>

21. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord [Internet]. 2009 [citado 10 ago 2026];114(1-3):163-173. Disponible en: <https://doi.org/10.1016/j.jad.2008.06.026>

22. Kocalevent RD, Hinz A, Brähler E. Standardization of the depression screener Patient Health Questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry [Internet].

2013 [citado 10 ago 2026];35(5):551-555. Disponible en: <https://doi.org/10.1016/j.genhosppsy.2013.04.006>

23. Kader Maideen SF, Mohd Sidik S, Rampal L, Mukhtar F. Prevalence, associated factors and predictors of depression among adults in the community of Selangor, Malaysia. PLoS One [Internet]. 2014 [citado 10 ago 2026];9(4):e95395. Disponible en: <https://doi.org/10.1371/journal.pone.0095395>

24. Shidhaye R, Gangale S, Patel V. Prevalence and treatment coverage for depression: a population-based survey in Vidarbha, India. Soc Psychiatry Psychiatr Epidemiol [Internet]. 2016 [citado 10 ago 2026];51(7):993-1003. Disponible en: <https://doi.org/10.1007/s00127-016-1220-9>

25. Shin C, Kim Y, Park S, Yoon S, Ko YH, Kim YK, et al. Prevalence and associated factors of depression in general population of Korea: results from the Korea National Health and Nutrition Examination Survey, 2014. J Korean Med Sci [Internet]. 2017 [citado 10 ago 2026];32(11):1861-1869. Disponible en: <https://doi.org/10.3346/jkms.2017.32.11.1861>

26. Kepler KL, Perlman SE, Chernov C, Thorpe LE, Hoenig J, Norman C. The burden of depression in New York City adults: results from the 2013-14 NYC Health and Nutrition Examination Survey. J Urban Health [Internet]. 2018 [citado 10 ago 2026];95(6):832-836. Disponible en: <https://doi.org/10.1007/s11524-018-0283-1>

27. Khaled SM. Prevalence and potential determinants of subthreshold and major depression in the general population of Qatar. J Affect Disord [Internet]. 2019 [citado 10 ago 2026];252:382-393. Disponible en: <https://doi.org/10.1016/j.jad.2019.04.056>

28. Ssebunnya J, Medhin G, Kangere S, Kigozi F, Nakku J, Lund C. Prevalence, correlates and help-seeking behaviour for depressive symptoms in rural Uganda: a population-based survey. Glob Ment Health (Camb) [Internet]. 2019 [citado 10 ago 2026];6:e27. Disponible en: <https://doi.org/10.1017/gmh.2019.25>

29. Cao C, Hu L, Xu T, Liu Q, Koyanagi A, Yang L, et al. Prevalence, correlates and misperception of depression symptoms in the United States, NHANES 2015-2018. J Affect Disord [Internet]. 2020 [citado 10 ago 2026];269:51-57. Disponible en: <https://doi.org/10.1016/j.jad.2020.03.031>

30. Arias-de la Torre J, Vilagut G, Ronaldson A, Serrano-Blanco A, Martín V, Peters M, et al. Prevalence and variability of current depressive disorder in 27 European countries: a population-based study. Lancet Public Health [Internet]. 2021 [citado 10 ago 2026];6(10):e729-e738. Disponible en: [https://doi.org/10.1016/S2468-2667\(21\)00047-5](https://doi.org/10.1016/S2468-2667(21)00047-5)

31. Shields M, Tonmyr L, Gonzalez A, Weeks M, Park SB, Robert AM, Blair DL, MacMillan HL. Symptoms of major depressive disorder during the COVID-19 pandemic: results from a representative sample of the Canadian population. *Health Promot Chronic Dis Prev Can* [Internet]. 2021 [citado 10 ago 2026];41(11):340-358. Disponible en: <https://doi.org/10.24095/hpcdp.41.11.04>
32. Melo APS, Bonadiman CSC, Andrade FM, Pinheiro PC, Malta DC. Depression screening in a population-based study: Brazilian National Health Survey 2019. *Cien Saude Colet* [Internet]. 2023 [citado 10 ago 2026];28(4):1163-1174. Disponible en: <https://doi.org/10.1590/1413-81232023284.14912022>
33. Xu WQ, Tan WY, Li XL, Huang ZH, Zheng HR, Hou CL, Jia FJ, Wang SB. Prevalence and correlates of depressive and anxiety symptoms among adults in Guangdong Province of China: a population-based study. *J Affect Disord* [Internet]. 2022 [citado 10 ago 2026];308:535-544. Disponible en: <https://doi.org/10.1016/j.jad.2022.04.089>
34. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. *JAMA Netw Open* [Internet]. 2020 [citado 10 ago 2026];3(9):e2019686. Disponible en: <https://doi.org/10.1001/jamanetworkopen.2020.19686>
35. Dawel A, Shou Y, Smithson M, Cherbuin N, Banfield M, Caelear AL, et al. The effect of COVID-19 on mental health and wellbeing in a representative sample of Australian adults. *Front Psychiatry* [Internet]. 2020 [citado 11 ago 2026];11:579985. Disponible en: <https://doi.org/10.3389/fpsy.2020.579985>
36. Hajek A, König HH. The prevalence and correlates of probable major depressive disorder and probable generalized anxiety disorder during the COVID-19 pandemic: results of a nationally representative survey in Germany. *Int J Environ Res Public Health* [Internet]. 2021 [citado 11 ago 2026];18(23):12302. Disponible en: <https://doi.org/10.3390/ijerph182312302>
37. Craig A, Rochat T, Naicker SN, Mapanga W, Mtintsilana A, Dlamini SN, et al. The prevalence of probable depression and probable anxiety, and associations with adverse childhood experiences and socio-demographics: a national survey in South Africa. *Front Public Health* [Internet]. 2022 [citado 11 ago 2026];10:986531. Disponible en: <https://doi.org/10.3389/fpubh.2022.986531>
38. Jiang Y, Deng W, Zhao M. Influence of the COVID-19 pandemic on the prevalence of depression in U.S. adults: evidence from NHANES. *Sci Rep* [Internet]. 2025 [citado 11 ago 2026];15:3107. Disponible en: <https://doi.org/10.1038/s41598-025-87593-5>

39. Negeri ZF, Levis B, Sun Y, He C, Krishnan A, Wu Y, et al. Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis. *BMJ* [Internet]. 2021 [citado 11 ago 2026];375:n2183. Disponible en: <https://doi.org/10.1136/bmj.n2183>

40. Costantini L, Pasquarella C, Odone A, Colucci ME, Costanza A, Serafini G, et al. Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. *J Affect Disord* [Internet]. 2021 [citado 11 ago 2026];279:473-483. Disponible en: <https://doi.org/10.1016/j.jad.2020.09.131>

Declaraciones

Conflictos de intereses: Los autores declaran no tener conflictos de intereses.

Financiamiento: Los autores declaran no haber recibido financiamiento específico para esta investigación.

Contribución de autoría: Alberto Guevara Tirado: conceptualización, metodología, curación de datos, análisis formal, supervisión, visualización, redacción del borrador original, revisión y edición. Gerard Guardia Barra: investigación, validación, revisión y edición. Wilson Leonardo Nolasco Elias: investigación, validación, revisión y edición. Jazmin Camila Borja Tineo: investigación, validación, revisión y edición.

Disponibilidad de datos y código: La base analítica y el script reproducible del metaanálisis están disponibles como material complementario de este manuscrito.

Uso de inteligencia artificial: Los autores declaran que no se utilizaron herramientas de inteligencia artificial generativa en la elaboración, redacción, análisis o revisión de este manuscrito.

Registro/protocolo: La revisión no fue registrada prospectivamente. No se realizó un registro retrospectivo porque no habría documentado prospectivamente las decisiones metodológicas. No existió un protocolo público previo ni desviaciones respecto de un protocolo registrado.