

Propiedades psicométricas del PHQ-9 (*Patient Health Questionnaire*) en centros de atención primaria de Chile

SANDRA SALDIVIA^{1,a}, JOSEPH ASLAN^{2,b}, FÉLIX COVA^{2,a}, BENJAMÍN VICENTE¹, CAROLINA INOSTROZA^{1,a}, PAULINA RINCÓN^{2,a}

Psychometric characteristics of the Patient Health Questionnaire (PHQ-9)

Background: Simple and brief questionnaires with adequate psychometric properties are useful for the early detection of depressive disorders. **Aim:** To analyze the psychometric behavior of the Patient Health Questionnaire (PHQ-9) in Chile. **Material and Methods:** Data were collected from 1,738 adult patients, consulting in primary health care centers. The values obtained in the PHQ-9 were compared with those of the Composite International Diagnostic Interview (CIDI). **Results:** The mean score of PHQ-9 was 4.85 ± 5.57 . Internal consistency indices were high. McDonald's ω coefficient of was 0.90 and Cronbach alpha 0.89. A confirmatory factor analysis showed a good fit of the one-dimensional model. Using a cutoff score of seven determined with the receiver operating characteristic (ROC) curve, the sensitivity and specificity of the instrument were 0.8 and 0.7, when compared with CIDI diagnoses. **Conclusions:** PHQ-9 is a useful instrument for the screening of depressive disorders in primary health care centers in Chile.

(Rev Med Chile 2019; 147: 53-60)

Key words: Depression; Patient Health Questionnaire; Primary Health Care; Sensitivity and Specificity.

¹Departamento de Psiquiatría y Salud Mental, Facultad de Medicina, Universidad de Concepción. Concepción, Chile.

²Departamento de Psicología, Facultad de Ciencias Sociales, Universidad de Concepción. Concepción, Chile.

^aDoctor/a en Psicología.

^bPsicólogo. Alumno de Doctorado en Psicología, Universidad de Concepción. Concepción, Chile.

Fuente de financiamiento:

FONDECYT 1171732.

Los autores declaran no tener conflictos de interés.

Recibido el 8 de mayo de 2018, aceptado el 8 de enero de 2019.

Correspondencia a:

Félix Cova

fecova@udec.cl

Entre los problemas de salud mental más frecuentes y con mayor impacto en el bienestar y calidad de vida de las personas están los síntomas y trastornos depresivos (TD)¹. El trastorno depresivo mayor (TDM) representa la segunda causa de discapacidad en 188 países². En Chile se ha encontrado una prevalencia de TDM, en el último año, de 7,9% para los trastornos depresivos³. La prevalencia de sintomatología depresiva es más alta, la Encuesta Nacional de Salud de Chile de 2009-2010 la estima en 17,2%⁴.

La atención primaria de salud (APS) tiene un rol fundamental en la detección oportuna y tratamiento de los trastornos depresivos. Investigaciones internacionales y nacionales convergen

en indicar que la prevalencia de TD en atención primaria es elevada, si bien las cifras precisas son objeto de discusión por consideraciones conceptuales y metodológicas; estas investigaciones también convergen en indicar la existencia de una baja tasa de detección de trastornos depresivos en APS^{5,6}. Estudios nacionales reportan prevalencias de 24,4% de trastornos mentales comunes en centros de APS en la zona centro sur del país, donde el trastorno depresivo estuvo presente en 13,4% de los usuarios (Saldivia S. Prevalencia y variables asociadas a trastornos mentales comunes en Centros de Atención Primaria de la Provincia de Concepción. Tesis de Magíster. Santiago: Universidad de Chile; 2016). Sobre la misma muestra, la

tasa de detección de trastornos mentales comunes por los médicos de APS alcanzó a 57,8% de los diagnósticos activos⁷.

Si bien, para el diagnóstico y valoración de síntomas y TD las entrevistas clínicas son indispensables, cuestionarios sencillos y breves son útiles para realizar tamizajes y para un monitoreo de la trayectoria de la sintomatología a lo largo del tiempo⁸. El uso de estos instrumentos puede contribuir a aumentar tasas de detección de trastornos depresivos en APS. Entre los instrumentos más reconocidos para estos propósitos están el *Beck Depression Inventory*, *Center for Epidemiologic Studies Depression Scale*, *Geriatric Depression Scale*, *Hospital Anxiety and Depression Scale*, *Edinburgh Postnatal Depression Scale* y el *Patient Health Questionnaire*-9⁹⁻¹¹. Instrumentos apropiados requieren mostrar evidencias de su confiabilidad en los diversos contextos de uso, sin las cuales sus puntuaciones son ininteligibles, y de indicadores de validez en relación a sus propósitos¹². Así, por ejemplo, se ha observado que instrumentos excesivamente breves, de uno a tres ítems, han mostrado no ser adecuados para tamizaje¹³.

El *Patient Health Questionnaire* (PHQ-9)¹⁴ es uno de los instrumentos que ha alcanzado mayor reconocimiento a nivel internacional debido a su capacidad para identificar la gravedad de los síntomas en personas con TD, y su sensibilidad al cambio para monitorear la respuesta al tratamiento¹⁵.

El PHQ-9 utiliza como recomendación original un puntaje de 10 para detectar un trastorno depresivo en pacientes de APS. Sin embargo, evidencia reciente relativa al puntaje de corte óptimo del PHQ-9 para tamizar depresión encontró los puntajes entre 8 y 11 con propiedades satisfactorias¹⁶. Para este estudio consideramos como valor 7, pues planteamos que es la mejor combinación de sensibilidad y especificidad. Se enfatiza la elección cuidadosa de un puntaje de corte, considerando las características de población y el contexto de aplicación. Un puntaje de corte inflexible podría generar falsos negativos en contextos hospitalarios y falsos positivos en APS¹⁷.

En Chile, el PHQ-9 ha sido estudiado en contextos de APS¹⁸, mostrando una alta consistencia interna ($\alpha = 0,84$) y una sensibilidad y especificidad de 88% y 92%, usando el puntaje de corte original, respecto del diagnóstico de trastorno

depresivo, usando la escala de Hamilton por psiquiatra. Sin embargo, el tamaño de la muestra en ese estudio fue pequeño (90 personas); se empleó solo análisis de componentes principales para el estudio de la homogeneidad de la escala, y no se analizaron los valores de sensibilidad y especificidad de diversos puntos de corte, por lo que son necesarios más estudios de validación del uso de esta escala PHQ-9 en nuestro contexto.

Este estudio tuvo por objetivo analizar las propiedades psicométricas del PHQ-9 en una muestra aleatoria de consultantes de atención primaria. Se realizó un análisis factorial confirmatorio (AFC) para analizar el supuesto de su estructura unifactorial¹⁹. Se analiza su consistencia interna, los valores de sensibilidad, especificidad, predicción positiva y negativa, y las razones de verosimilitud para distintos puntos de corte tomando como criterio el diagnóstico de trastorno depresivo obtenido a través de una entrevista estructurada.

Material y Método

Muestra

Los datos fueron obtenidos de la línea base de una población que ya había sido partícipe de otros estudios^{20,21}. Se realizó un muestreo aleatorio de pacientes entre 18 y 75 años en 10 centros de APS de la provincia de Concepción, Chile. En cada centro, un profesional contactaba e invitaba a participar a la persona; si aceptaba, se coordinaba una entrevista para la recolección de datos. Los criterios de exclusión incluyeron la inhabilidad para entender el idioma local, la presencia de psicosis, demencia o de una enfermedad física incapacitante.

La muestra del presente estudio corresponde a usuarios de APS, que pudieron ser contactados y evaluados por entrevistadores previamente entrenados que recogieron la información, que incluyó una encuesta sociodemográfica, el PHQ-9, la entrevista diagnóstica internacional compuesta (CIDI) y otras medidas no relacionadas con este estudio.

Medidas

Patient Health Questionnaire (PHQ-9)

El PHQ-9 es una medida de autoinforme de nueve ítems que evalúa la presencia de síntomas depresivos basados en los criterios del DSM-IV

para el episodio depresivo mayor²². Refiere a los síntomas experimentados por los pacientes durante las dos semanas previas a la entrevista. En este caso, se utilizó la versión en español de la escala²³. Las puntuaciones del PHQ-9 se califican con una escala Likert que va de 0 (nunca), a 1 (varios días), 2 (más de la mitad de los días) y 3 (casi todos los días), por lo que el puntaje total va en un rango de 0 a 27. La gravedad de los síntomas puede organizarse en 4 categorías: 0-4 (mínimo), 5-9 (leve), 10-14 (moderado), 15-19 (moderado a grave), 20-27 (grave)¹⁴. El PHQ-9 se desarrolló como una herramienta de tamizaje, siendo los puntajes de corte recomendados entre 8 y 11 para un probable caso de depresión mayor¹⁶.

Entrevista diagnóstica internacional compuesta (CIDI)

La CIDI²⁴ es una entrevista diagnóstica estructurada desarrollada por la Organización Mundial de la Salud (OMS). Ha sido utilizada en estudios epidemiológicos en población general, presenta una alta confiabilidad inter-jueces y test-retest y evidencias de validez en múltiples estudios internacionales y en Chile^{25,26}. La CIDI puede ser aplicada por entrevistadores legos, superando así las limitaciones de las entrevistas realizadas solo por profesionales, manteniendo el objetivo de que los diagnósticos se atuvieran estrictamente a los criterios diagnósticos establecidos²⁷.

La CIDI 2.1 entrega diagnósticos siguiendo criterios DSM-IV y CIE-10, para presencia de trastorno a lo largo de la vida, en los últimos 12 meses y últimos 30 días^{24,28}. Para este estudio se utilizaron los diagnósticos de acuerdo a criterio DSM-IV y se evaluó la presencia de este durante el último mes. En este estudio se emplearon las secciones A (datos sociodemográficos) y E (depresión).

Análisis de datos

Para el análisis de descriptivos, sensibilidad, especificidad, valores predictivos y razones de verosimilitud del PHQ-9 se construyó la curva de eficacia diagnóstica (ROC), usando como valor de criterio el diagnóstico de TDM obtenido con la CIDI (criterios DSM-IV), utilizando el *software* estadístico SPSS 20²⁹. Para el cálculo del coeficiente ω de McDonald se utilizó el JASP³⁰, y el AFC se usó el MPlus 7.11³¹. Se postuló un modelo unifactorial, utilizando el estimador de mínimos cuadrados ponderados robustos (WLSMV), que no asume

la distribución normal de las variables y es considerado como la mejor opción para modelar datos categóricos u ordinales³².

Consideraciones éticas

Cada persona que aceptó participar firmó un formulario de consentimiento informado y los protocolos del estudio fueron sometidos a la consideración de los comités de ética de la Facultad de Medicina de la Universidad de Concepción.

Resultados

La media de edad de los 1.738 participantes fue 54,52 años ($DT = 16,2$). De ellos, 1.321 eran mujeres (76%) y 417 hombres (24%). La mitad de los participantes (50,6%) estaba casado y 62,3% tenía menos de 12 años de escolaridad. La mediana fue 3,0, con una asimetría de 1,52 ($DT = 0,05$). La distribución de los puntajes del PHQ-9 se muestra en la Tabla 1. La mayoría de los participantes (83,9%) tuvieron un puntaje PHQ-9 bajo; 24,9% de los entrevistados tuvo un puntaje >7 ; 7,4% obtuvo una puntuación de 15 o más (depresión moderada o severa).

Los dos ítems más reportados fueron baja energía y problemas de sueño. El ítem menos señalado fue la ideación suicida. Como indicadores de consistencia interna se calculó tanto el coeficiente ω de McDonald, más apto para escalas con pocas opciones de respuesta³³ y el alfa de Cronbach. Ambos indicadores fueron elevados: 0,896 y 0,891, respectivamente (Tabla 2).

Análisis factorial confirmatorio

El AFC indicó un buen ajuste del modelo a los datos ($\chi^2 = 190,82$, $gl = 27$, $p < 0,001$; CFI = 0,990; TLI = 0,987; RMSEA = 0,059 [90% IC 0,051-

Tabla 1. Distribución de los puntajes del PHQ-9 (n = 1.738)

Nivel de depresión, puntaje PHQ-9	n	%
Mínimo, 0-4	1.039	59,8
Leve, 5-9	419	24,1
Moderado, 10-14	152	8,7
Moderado grave, 15-19	78	4,5
Grave, 20-27	50	2,9

Tabla 2. Valores de ítems del PHQ-9 y correlaciones ítem-total (n = 1.738)

Ítem PHQ-9	Media	DT	Correlaciones ítem-total corregidas	ω si se elimina el ítem
1. ¿Ha sentido poco interés o agrado al hacer las cosas?	0,55	0,86	0,71	0,88
2. ¿Se ha sentido triste, deprimido o desesperado?	0,61	0,85	0,74	0,87
3. ¿Ha tenido problemas para dormir, mantenerse despierto o duerme demasiado?	0,74	1,00	0,63	0,88
4. ¿Se siente cansado o tiene poca energía?	0,88	0,96	0,69	0,88
5. ¿Tiene poco apetito o excesivo apetito?	0,63	0,95	0,56	0,89
6. ¿Se ha sentido mal consigo mismo, ha sentido que usted es un fracaso o ha sentido que se ha fallado a sí mismo o a su familia?	0,42	0,78	0,67	0,88
7. ¿Ha tenido problemas para concentrarse en actividades como leer el periódico o ver televisión?	0,41	0,78	0,64	0,88
8. ¿Se mueve o habla tan despacio que otras personas pueden darse cuenta? ¿Está tan inquieto o intranquilo que da vueltas de un lugar a otro más que de costumbre?	0,42	0,77	0,68	0,88
9. ¿Ha pensado que estaría mejor muerto o ha deseado hacerse daño de alguna forma?	0,19	0,54	0,53	0,89

0,067]). El χ^2 fue significativo, pero dado el tamaño muestral, se trata de un resultado esperable; CFI, TLI y los valores de RMSEA estuvieron dentro de los estándares recomendados (i.e., CFI > 0,95, TLI > 0,95, RMSEA < 0,08)³⁴. La Figura 1 muestra la estructura factorial del PHQ-9, donde todas las

cargas factoriales fueron significativas ($p < 0,05$) y mayores a 0,698.

Convergencia con diagnóstico CIDI

De acuerdo a la CIDI, 187 pacientes (10,7%) cumplían criterios para el diagnóstico de TDM.

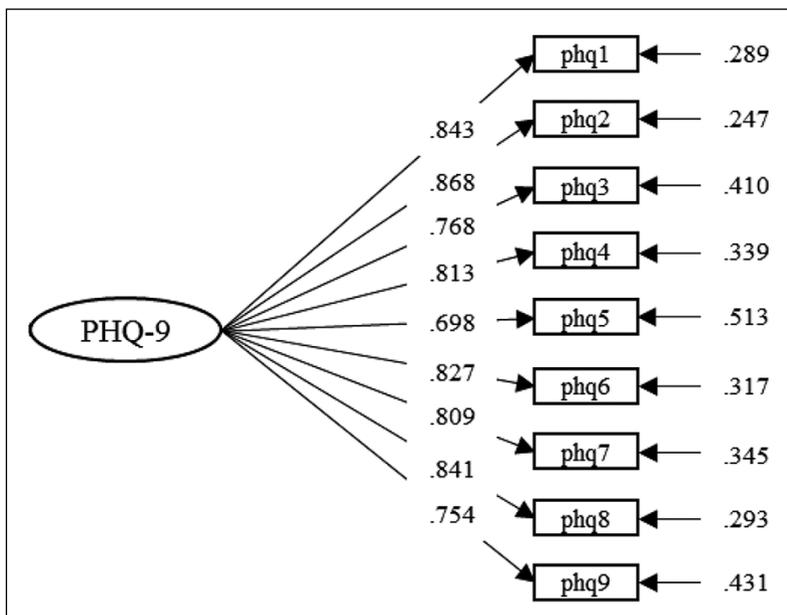


Figura 1. Diagrama de estructura factorial para los ítems del PHQ-9. Cargas estandarizadas.

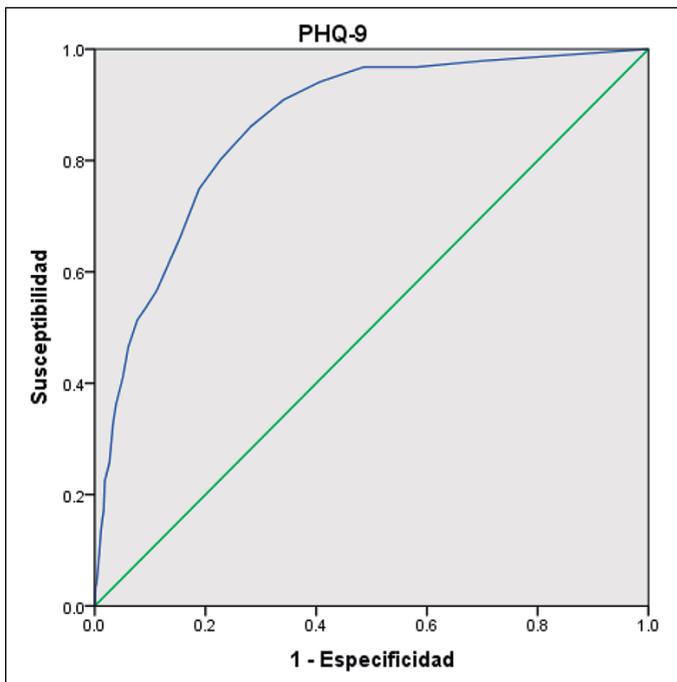


Figura 2. La curva característica operativa del receptor (ROC) para el PHQ-9 versus el CIDI para el diagnóstico de trastorno depresivo mayor.

Se observaron diferencias estadísticamente significativas en la puntuación media del PHQ-9 entre los pacientes con y sin diagnóstico de TDM ($M = 12,34$, $DT = 6,55$; $M = 3,94$, $DT = 4,70$, $t(209.699) = 17.007$, $p < 0,001$, $r = 0,76$).

La Figura 2 ilustra la curva ROC que muestra el funcionamiento del PHQ-9 en la identificación de pacientes con TDM. El área bajo la curva (AUC)

fue de 0,86 ($DT = 0,01$, 95% CI 0,83 a 0,88), lo que da cuenta de una precisión (exactitud) moderada³⁵. La Tabla 3 muestra la sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN) y la razón de probabilidad de diferentes umbrales PHQ-9 en el diagnóstico de depresión mayor. La puntuación de corte de siete puntos es la que maximiza los valores

Tabla 3. Desempeño de varios puntos de corte del PHQ-9 para trastorno depresivo mayor a un mes

Punto de corte	Sensibilidad	Especificidad	Valor predictivo positivo	Valor predictivo negativo	Razón de verosimilitud positiva	Razón de verosimilitud negativa
≥ 6	0,86	0,72	0,27	0,98	3,06	0,19
≥ 7	0,80	0,77	0,30	0,97	3,52	0,26
≥ 8	0,75	0,81	0,32	0,96	3,98	0,31
≥ 9	0,66	0,85	0,34	0,95	4,29	0,40
≥ 10	0,57	0,89	0,38	0,94	5,05	0,49
≥ 11	0,53	0,91	0,41	0,94	5,84	0,51
≥ 12	0,51	0,92	0,45	0,94	6,69	0,53
≥ 13	0,47	0,94	0,48	0,94	7,68	0,57
≥ 14	0,41	0,95	0,49	0,93	8,08	0,62
≥ 15	0,36	0,96	0,53	0,93	9,40	0,66

de sensibilidad y especificidad (sensibilidad de 0,80 y especificidad de 0,77). Con este puntaje de corte, el VPP fue de 0,30 y el VPN de 0,97, y las razones de verosimilitud positiva y negativa fueron de 3,52 y 0,26, respectivamente.

Discusión

En este estudio se analizan las propiedades psicométricas del PHQ-9 en usuarios de centros de atención primaria en Chile y su utilidad como instrumento de tamizaje de TD. Si bien este no es el único estudio del funcionamiento del PHQ-9 que se ha realizado en nuestro país¹⁸, tiene la ventaja de haber sido desarrollado sobre una muestra amplia y aleatoria de consultantes por morbilidad general que concurren a centros de APS, nivel con amplia cobertura en nuestro país; además, en este estudio se empleó como *gold standard* la CIDI, instrumento ampliamente usado en estudios epidemiológicos por su uso riguroso de los criterios diagnósticos de trastornos mentales establecidos.

Los resultados muestran un comportamiento psicométrico adecuado del PHQ-9 en la muestra estudiada. En concordancia con lo mostrado por la mayoría de los estudios, el modelo unifactorial mostró un buen ajuste^{19,36}. Todos los ítems mostraron cargas superiores a 0,69, y las correlaciones ítem-total corregidas fueron superiores a 0,56. La consistencia interna observada también fue alta.

A diferencia de estudios donde el ítem de ideación suicida ha mostrado una correspondencia menor con las demás medidas de la escala, en la muestra estudiada sus valores de convergencia fueron altos, superiores al ítem que mide aumento o disminución del apetito, que fue el que mostró los valores más bajos de convergencia. Por razones psicométricas no habría motivos para eliminar el ítem de ideación suicida que, a su vez, tiene alta relevancia clínica. Sin embargo, algunos autores sugieren eliminarlo para evitar las implicaciones éticas de su aplicación³⁷.

Asimismo, el PHQ-9 mostró adecuados valores de sensibilidad y especificidad, comparables a los obtenidos en otros estudios¹⁶. Los VPP y VPN están en los valores que confirman su utilidad como instrumento de tamizaje³⁸. Como es frecuente con instrumentos de tamizaje empleados en contextos de relativas bajas tasas de prevalencia del trastorno o enfermedad analizado (10,6% en este estudio), el

mejor rendimiento del PHQ-9 fue para descartar la presencia de un trastorno depresivo más que para confirmar su presencia. Los VPP mostrados para puntajes entre 7 y 11 están en un rango entre 0,27 y 0,41, comparables a los obtenidos internacionalmente¹⁶. Si bien son valores adecuados para un instrumento de tamizaje en salud mental, indican que un valor positivo en el instrumento no debe ser en ningún caso considerado como indicador de probable diagnóstico positivo, sino como un indicador de la conveniencia de una nueva evaluación, más profunda.

En la muestra estudiada, el puntaje que mejor combina los valores de sensibilidad y especificidad fue de 7, lo que está en el rango inferior en comparación con otros estudios. Estudios previos con otros instrumentos en nuestro país han mostrado una tendencia inversa³⁹. Podría atribuirse a un efecto de atenuación debido a que es una muestra que ha recibido sucesivas evaluaciones. La mayoría de los estudios se ha realizado con muestras que solo han recibido una única evaluación. La selección del punto de corte debe estar en función de los objetivos de quienes empleen el instrumento y la necesidad de maximizar la detección o de reducir el número de falsos positivos^{17,40}.

Aunque este estudio tuvo la ventaja de ser realizado con una muestra amplia de consultantes de APS, tiene la desventaja habitual de los estudios de instrumentos de tamizaje de trastornos depresivos: el análisis se hizo contrastando dos medidas puntuales en el tiempo (PHQ-9 y CIDI) y asumiendo que el uso de los criterios diagnósticos DSM-IV en población general es pertinente para detectar “auténticos” trastornos depresivos. Diversos autores han observado que existe una sobreestimación de TD si se usan los criterios diagnósticos DSM-IV descontextualizadamente²⁷. Ello es un tema que trasciende a los instrumentos de tamizaje y que apunta al modo de entender los TD y a la pertinencia de sus criterios diagnósticos actuales^{41,42}. Una implicación de estas consideraciones para el uso de cuestionarios de tamizaje es que podría ser conveniente, en el caso de puntajes de riesgo, que se tome más de una medida, separadas entre sí por un breve período de tiempo, de modo de descartar aquellos casos en que la sintomatología depresiva muestra un carácter situacional y transitorio, sugerencia que algunos investigadores ya están considerando⁴³.

En síntesis, con las cautelas indicadas, los

resultados obtenidos indican que el PHQ-9 es un instrumento útil para el tamizaje de trastornos depresivos en los centros de atención primaria en nuestro país.

Agradecimientos: Los autores agradecen a los Departamentos de Salud Municipal de Concepción, Talcahuano, Chiguayante y Hualpén, a los centros de atención primaria en salud y a los usuarios que participaron de este estudio.

Referencias

- Ferrari AJ, Somerville AJ, Baxter AJ, Norman R, Patten SB, Vos T, et al. Global variation in the prevalence and incidence of major depressive disorder: a systematic review of the epidemiological literature. *Psychol Med* 2013; 43 (3): 471-81.
- Vos T, Barber RM, Bell B, Bertozzi-Villa A, Biryukov S, Bolliger I, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015; 386 (9995): 743-800.
- Vicente B, Kohn R, Rioseco P, Saldivia S, Levav I, Torres S. Lifetime and 12-month prevalence of DSM-III-R disorders in the Chile psychiatric prevalence study. *Am J Psychiatry* 2006; 163 (8): 1362-70.
- MINSAL. Encuesta Nacional de Salud ENS Chile 2009-2010. 2011.
- Cano A, Salguero JM, Wood C, Dongil E, Latorre JM. La depresión en atención primaria: prevalencia, diagnóstico y tratamiento. *Papeles del Psicólogo* 2012; 33 (1): 2-11.
- Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009; 374 (9690): 609-19.
- Saldivia S, Vicente B, Kohn R, Levav I. La epidemiología de los trastornos psiquiátricos en la atención primaria de salud. In: Rodríguez J, R. K, Aguilar-Gaxiola S, editors. *Epidemiología de los trastornos mentales en América Latina y el Caribe*. Pub. Científico Técnica N° 632. Washington D.C.: Organización Panamericana de la Salud; 2009.
- Schantz K, Reighard C, Aikens JE, Aruquipa A, Pinto B, Valverde H, et al. Screening for depression in Andean Latin America: Factor structure and reliability of the CES-D short form and the PHQ-8 among Bolivian public hospital patients. *Int J Psychiatry Med* 2017; 52 (4-6): 315-27.
- Smarr KL, Keefer AL. Measures of depression and depressive symptoms: Beck Depression Inventory II (BDI II), Center for Epidemiologic Studies Depression Scale (CES D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). *Arthritis Care Res* 2011; 63 (S11): S454-S66.
- Jadresic E, Araya R, Jara C. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Chilean Postpartum Women. *J Psychosom Obstet Gynaecol* 1995; 16 (4): 187-91.
- Castañón SC, Pinto LJ. Mejorando la pesquisa de depresión posparto a través de un instrumento de tamizaje, la escala de depresión posparto de Edimburgo. *Rev Med Chile* 2008; 136: 851-8.
- Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press; 2015.
- Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *Br J Gen Pract* 2007; 57 (535): 144-51.
- Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med* 2001 (16): 606-13.
- Löwe B, Schenkel I, Carney-Doebbeling C, Göbel C. Responsiveness of the PHQ-9 to psychopharmacological depression treatment. *Psychosomatics* 2006; 47 (1): 62-7.
- Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ* 2012; 184 (3): E191-E6.
- Kroenke K, Spitzer RL, Williams JB, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *Gen Hosp Psychiatry* 2010; 32 (4): 345-59.
- Baader T, Molina JL, Venezian S, Rojas C, Fariás R, Fierro-Freixenet C, et al. Validación y utilidad de la encuesta PHQ-9 (Patient Health Questionnaire) en el diagnóstico de depresión en pacientes usuarios de atención primaria en Chile. *Rev Chil Neuro Psiquiat* 2012; 50 (1): 10-22.
- Merz EL, Malcarne VL, Roesch SC, Riley N, Sadler GR. A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among English- and Spanish-speaking Latinas. *Cultur Divers Ethnic Minor Psychol* 2011; 17 (3): 309-16.
- King M, Walker C, Levy G, Bottomley C, Royston P, Weich S, et al. Development and validation of an international risk prediction algorithm for episodes of major

- depression in general practice attendees: the PredictD study. *Arch Gen Psychiatry* 2008; 65 (12): 1368-76.
21. Saldivia S, Vicente B, Marston L, Melipillán R, Nazareth I, Bellón-Saameño J, et al. Desarrollo y validación de un algoritmo para predecir riesgo de depresión en consultantes de atención primaria en Chile. *Rev Med Chile* 2014; 142: 323-9.
 22. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Publishing; 1994.
 23. Díez-Quevedo C, Rangil T, Sánchez-Planell L, Kroenke K, Spitzer RL. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. *Psychosom Med* 2001; 63 (4): 679-86.
 24. World Health Organization. *Composite International Diagnostic Interview (CIDI)*. Version 2.1. Geneva, Switzerland: World Health Organization; 1997.
 25. Peters L, Andrews G. Procedural validity of the computerized version of the Composite International Diagnostic Interview (CIDI-Auto) in the anxiety disorders. *Psychol Med* 1995; 25 (6): 1269-80.
 26. Vielma M, Vicente B, Riosco P, Castro N, Torres S. Validación en Chile de la entrevista diagnóstica estandarizada para estudios epidemiológicos CIDI. *Revista de Psiquiatría* 1992; 9 (1): 1039-49.
 27. Gelaye B, Tadesse MG, Williams MA, Fann JR, Stoep AV, Zhou X-HA. Assessing Validity of a Depression Screening Instrument in the Absence of a Gold Standard. *Ann Epidemiol* 2014; 24 (7): 527-31.
 28. Robins LN, Wing J, Wittchen H, Helzer JE, Babor TF, Burke J, et al. The composite international diagnostic interview: An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Arch Gen Psychiatry* 1988; 45 (12): 1069-77.
 29. IBM Corp. *IBM SPSS Statistics for Windows*, version 20.0. Armonk, NY: IBM Corporation; 2011.
 30. JASP Team. *JASP (Version 0.8.4)*. 2017.
 31. Muthén LK, Muthén BO. *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén; 1998-2017.
 32. Brown TA. *Confirmatory Factor Analysis for Applied Research*. 2nd ed. New York: The Guilford Press; 2015.
 33. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation* 2012; 17 (3).
 34. Hair JF, Black WC, Babin BJ, Anderson RE. *Multivariate Data Analysis*: Pearson Education Limited; 2013.
 35. Streiner DL, Cairney J. What's under the ROC? An introduction to receiver operating characteristics curves. *Can J Psychiatry* 2007; 52 (2): 121-8.
 36. Schantz K, Reighard C, Aikens JE, Aruquipa A, Pinto B, Valverde H, et al. Screening for depression in Andean Latin America: Factor structure and reliability of the CES-D short form and the PHQ-8 among Bolivian public hospital patients. *Int J Psychiatry Med* 2017; 52 (4-6): 315-27.
 37. 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* 2009; 114 (1): 163-73.
 38. McGee S. Simplifying Likelihood Ratios. *J Gen Intern Med* 2002; 17 (8): 647-50.
 39. Araya R, Wynn R, Lewis G. Comparison of two self administered psychiatric questionnaires (GHQ-12 and SRQ-20) in primary care in Chile. *Soc Psychiatry Psychiatr Epidemiol* 1992; 27 (4): 168-73.
 40. Patel V, Araya R, Chowdhary N, King M, Kirkwood B, Nayak S, et al. Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. *Psychol Med* 2008; 38 (2): 221-8.
 41. Frances A. *¿Somos todos enfermos mentales? Manifiesto contra los abusos de la psiquiatría*. Barcelona: Ariel; 2014.
 42. Wakefield JC, Demazeux S. *Sadness or Depression?: International Perspectives on the Depression Epidemic and Its Meaning*: Springer; 2016.
 43. Araya R, Rojas G, Fritsch R, Gaete J, Rojas M, Simon G, et al. Treating depression in primary care in low-income women in Santiago, Chile: a randomised controlled trial. *Lancet* 2003; 361 (9362): 995-1000.