



UNIVERSIDAD PERUANA
CAYETANO HEREDIA
FACULTAD DE PSICOLOGÍA

VALIDACIÓN DE LA VERSIÓN
PERUANA DEL CUESTIONARIO SOBRE
LA SALUD DEL PACIENTE (PHQ-9)
PARA TAMIZAJE DE DEPRESIÓN EN
PACIENTES ADULTOS CON CÁNCER
DE UN HOSPITAL PÚBLICO
ESPECIALIZADO DE LIMA
METROPOLITANA

TESIS PARA OPTAR EL TÍTULO DE
LICENCIADO EN PSICOLOGÍA

MARTÍN ALONSO DE LA FLOR VALDIVIA

LIMA – PERÚ

2020

MIEMBROS DEL JURADO

Dra. ELIZABETH DANY ARAUJO ROBLES
Presidente

Mg. NORMA GISELLA DIAZ ARANA
Vocal

Lic. RITA ORIHUELA ANAYA
Secretaria

ASESOR DE TESIS

Mg. ELENA ESTHER YAYA CASTAÑEDA

DEDICATORIA

Dedico esta investigación a los pacientes del INEN. Siempre recordaré la
confianza que me otorgaron y las vivencias que compartieron.

AGRADECIMIENTOS

Gracias al Instituto Nacional de Enfermedades Neoplásicas por permitirme realizar este trabajo de tesis en tan prestigiosa institución.

Gracias a mi familia por siempre estar detrás de mí, impulsándome a seguir adelante.

Gracias a todos los pacientes que colaboraron con esta investigación.

TABLA DE CONTENIDOS

RESUMEN

ABSTRACT

INTRODUCCIÓN	1
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA.....	2
1. IDENTIFICACIÓN DEL PROBLEMA.....	2
2. JUSTIFICACIÓN E IMPORTANCIA	4
3. LIMITACIONES DE LA INVESTIGACIÓN.....	7
4. OBJETIVOS DE LA INVESTIGACIÓN.....	8
4.1. Objetivo general.....	8
4.2. Objetivos específicos	8
CAPÍTULO II: MARCO TEÓRICO.....	10
1. ASPECTOS CONCEPTUALES.....	10
1.1. Depresión	10
1.2. Validez	23
2. INVESTIGACIONES EN TORNO AL PROBLEMA INVESTIGADO, NACIONALES E INTERNACIONALES	39
3. DEFINICIONES CONCEPTUALES Y OPERACIONALES	42
3.1. Definición conceptual	42
3.2. Definición operacional.....	42
CAPÍTULO III: METODOLOGÍA	44
1. TIPO DE INVESTIGACIÓN	44
2. DISEÑO DE LA INVESTIGACIÓN	44
3. NATURALEZA DE LA MUESTRA	44
3.1. Participantes	44
3.2. Método de selección de los participantes.....	45
3.3. Criterios de inclusión	46
3.4. Criterios de exclusión.....	46
3.5. Criterios de eliminación	47
4. INSTRUMENTOS	48

4.1. Cuestionario sobre la Salud del Paciente (PHQ-9)	49
4.2. Escala de Ansiedad y Depresión Hospitalaria (HADS)	50
5. PROCEDIMIENTO	51
6. PLAN DE ANÁLISIS DE DATOS	52
7. CONSIDERACIONES ÉTICAS	55
7.1. Autonomía.....	55
7.2. Beneficencia.....	56
7.3. No maleficencia	56
7.4. Justicia.....	57
CAPÍTULO IV: RESULTADOS.....	58
CAPÍTULO V: DISCUSIÓN.....	76
CAPÍTULO VI: CONCLUSIÓN.....	97
RECOMENDACIONES	100
REFERENCIAS BIBLIOGRÁFICAS.....	101

RESUMEN

Aproximadamente 30% de los pacientes oncológicos en el Perú experimentan sintomatología depresiva clínicamente significativa, por lo que se recomienda que la depresión sea evaluada en todas las fases de la enfermedad y tratamiento. Se evaluó la evidencia de validez de las puntuaciones de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9) para tamizaje de depresión en una población de 500 pacientes adultos con cáncer de un hospital de atención especializada de Lima metropolitana. Se halló evidencia de validez basada en el contenido del cuestionario ($V > 0.8$). Se identificó una estructura interna simple de dos factores que explica 47% de la varianza con una adecuada bondad de ajuste (CFI=.942; TLI=.92; AGFI=.946; SRMR=.044; RMSEA=.058). El PHQ-9 y su dimensión Cognitiva/Afectiva presentaron una adecuada confiabilidad ($\omega_1=.76$; $\omega_2=.72$), pero moderada en la dimensión somática ($\omega=.69$). En cuanto al sesgo entre grupos, el instrumento tiene invarianza escalar según Sexo y Grado de instrucción, invarianza configural según Lugar de aplicación, y no presenta invarianza según Edad y Estado Civil ($\Delta CFI > .01$). Se hallaron correlaciones positivas entre el PHQ-9 con el HADS y sus dimensiones ($r > .6$) y con el autoreporte de diagnóstico depresión ($r=.71$). Se identificó que el punto de corte más adecuado del PHQ-9 es ≥ 8 (Se=.87; Sp=.77), y se puede aplicar con un algoritmo de aplicación (Se=.70; Sp=.86) o la versión abreviada PHQ-2 ≥ 3 (Se=.70; Sp=.85). En conclusión, la evidencia de validez del PHQ-9 justifica su uso para el objetivo propuesto.

Palabra clave: Depresión, estudio de validación, pacientes oncológicos, propiedades psicométricas, tamizaje

ABSTRACT

Approximately 30% of peruvian cancer patients experience clinically significant depressive symptoms, so it is recommended to screen patients for depression in all phases of the disease and treatment. Therefore, the evidence of validity of the scores of the peruvian version of the Patient Health Questionnaire (PHQ-9) for depression screening was evaluated in a population of 500 adult cancer patients from a specialized hospital in metropolitan Lima. Evidence of validity was found based on the content of the questionnaire ($V > 0.8$). A simple two-factor internal structure that explains 47% of the variance was identified. It presented an adequate goodness of fit (CFI=.942; TLI=.92; AGFI=.946; SRMR=.044; RMSEA=.058). The PHQ-9 and its Affective/Cognitive dimension showed adequate reliability ($\omega_1=.76$; $\omega_2=.72$), but moderate in the Somatic dimension ($\omega=.69$). The instrument presents scalar measurement invariance when factoring participant's Sex and Level of Education, configural invariance with Place of application, and differentially measured depression when factoring Civil status and Age groups ($\Delta CFI > .01$). Positive correlations were found between the PHQ-9 and the HADS and its dimensions ($r > .6$), and with the self-report of depression ($r=.71$). The most suitable cut-off point for the PHQ-9 was ≥ 8 (Se=.87; Sp=.77), and it can be used with an application algorithm (Se=.70; Sp=.86), or the abbreviated version PHQ-2 ≥ 3 (Se=.70; Sp=.85). In conclusion, the evidence of validity of the PHQ-9 justifies its use for the proposed objective.

Keyword: Depression, validation study, cancer patients, psychometric properties, screening

INTRODUCCIÓN

La población oncológica presenta una elevada prevalencia de depresión. Por este motivo, las Guías de Práctica Clínicas (GCP) suelen recomendar el uso de instrumentos psicométricos para tamizaje de depresión en todas las fases del tratamiento oncológico; en particular, se recomienda el Cuestionario sobre la Salud del Paciente (PHQ-9). Sin embargo, cuando se emplea un instrumento psicométrico, uno debe de contar con evidencia de la validez de dicho instrumento para el objetivo propuesto en la población de interés.

Por consiguiente, la presente investigación tuvo como objetivo evaluar la medida en la que la evidencia de validez de las puntuaciones de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9) justifica su uso para tamizaje de depresión en una población de pacientes adultos con cáncer de un Hospital de atención especializada de Lima metropolitana. Participaron 451 participantes, y se utilizó adicionalmente el HADS.

La presente investigación se desarrolla en cinco capítulos. En el primer capítulo, se menciona la identificación del problema, su justificación e importancia, y se mencionan los objetivos del estudio. En el segundo apartado, se presenta la conceptualización de la depresión y de la validez, así como los antecedentes alrededor del tema. En el tercero, se describe la metodología usada, incluyendo el tipo y diseño del estudio, así como los instrumentos y procedimientos. En el cuarto y quinto apartado, se describen los resultados de los diferentes análisis estadísticos y su discusión respectiva.

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

1. IDENTIFICACIÓN DEL PROBLEMA

Aproximadamente 30% de los pacientes oncológicos en el Perú experimentan sintomatología depresiva clínicamente significativa (Berrospi-Reyna, Herencia-Souza, y Tarazona, 2017). Además, el malestar emocional que experimentan los pacientes oncológicos con depresión puede interferir con su capacidad para adaptarse y afrontar el cáncer, sus síntomas y su tratamiento (Holland et al., 2018). La evidencia sugiere que aquellos pacientes oncológicos con necesidades psicosociales no atendidas son más propensos a hacer uso de los servicios de emergencia, demandan más atención por parte del personal de salud, presentan menor adherencia al tratamiento, están más tiempo hospitalizados, y aumentan los costos de salud, entre otras consecuencias (Barber et al., 2015). Por estos motivos es importante realizar las intervenciones oportunas en los casos que amerite.

Sin embargo, existen investigaciones internacionales que señalan que menos de la mitad de los pacientes oncológicos con casos clínicos de depresión reciben la atención necesaria (Kohn et al., 2005; Martínez et al., 2017). En el Perú, el Instituto Nacional de Salud Mental (INSM) reporta que 75% a 85% de la población general que manifestó necesidad de atención en salud mental no recibieron ningún tipo de tratamiento (Instituto Nacional de Salud Mental, 2009, 2013). Esto se debe en parte a que los médicos no especializados en salud mental muestran conocimiento y prácticas inadecuadas en relación al diagnóstico y tratamiento de la depresión (Agreda-Carrillo, Galindo-Morales, Larios-Falcón, Arévalo-Flores, y Cruzado, 2017; Leonardo Olivera, Villalobos Aguinaga, y León

Jiménez, 2016). Hasta el momento no se hallaron investigaciones en el Perú sobre la cantidad de pacientes oncológicos con depresión que no reciben atención oportuna. No obstante, considerando los hallazgos en otros países y en el Perú, y las limitaciones que presentan los servicios de salud mental de los centros oncológicos peruanos en comparación con otros países, se puede suponer que en el Perú el porcentaje de personas con depresión que no reciben atención oportuna es similar o superior a lo encontrado en otros países.

Por estos motivos, en múltiples países se han desarrollado guías de práctica clínica (GCP) sobre el manejo de malestar emocional, ansiedad y depresión en pacientes oncológicos. Las GCPs coinciden en recomendar que la depresión sea evaluada en todas las fases de la enfermedad y tratamiento, y por lo menos, en la visita inicial e intervalos apropiados de relevancia clínica, empleando instrumentos de tamizaje breves y adecuadamente validados al contexto local.

En específico, la GCP pan-canadiense desarrollada por la Canadian Association of Psychosocial Oncology (CAPO), la GCP norteamericana desarrollada por la American Society of Clinical Oncology (ASCO), y la GCP norteamericana desarrollada por la National Comprehensive Cancer Network (NCCN) coinciden en recomendar el Cuestionario sobre la Salud del Paciente de dos ítems (PHQ-2) y de nueve ítems (PHQ-9) (Andersen et al., 2014; Holland et al., 2018; Howell et al., 2015).

En el contexto peruano, el Ministerio de Salud del Perú (MINSA) y la Organización Panamericana de Salud (OPS) desarrollaron la “Guía técnica de diagnóstico y tratamiento de depresión en adultos de 18 a 65 años para médicos no

psiquiatras”. Esta GCP informada por consenso deviene de la adaptación de GCP internacionales de alta calidad, y concuerda en recomendar el uso del PHQ-9 para diagnóstico de depresión y para la evaluación de la eficacia de los tratamientos antidepresivos (Ministerio de Salud y Organización Panamericana de Salud, 2013).

Sin embargo, en el Perú, no hay suficiente evidencia para apoyar la validez del PHQ-9 para el uso propuesto en esta investigación, es decir, para realizar tamizaje de depresión en población adulta con cáncer de un hospital público especializado de Lima metropolitana. Por ende, los profesionales e investigadores que deseen usar el PHQ-9 solo podrán emplear versiones no adaptadas, adaptadas parcialmente, o versiones adaptadas en otros países y en otras poblaciones, sin ninguna garantía científica de la confiabilidad ni validez de los resultados obtenidos. Este uso de instrumentos psicométricos en el ámbito hospitalario es sub-óptimo y no recomendable.

Considerando lo anteriormente mencionado, la pregunta principal que guiará esta investigación es: ¿En qué medida la evidencia de validez de las puntuaciones de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9) justifica su uso para tamizaje de depresión en pacientes adultos con cáncer de un Hospital público especializado de Lima metropolitana?

2. JUSTIFICACIÓN E IMPORTANCIA

En un plano práctico, la validación de este instrumento para tamizaje de depresión ofrecería múltiples beneficios.

En primer lugar, el PHQ-9 cuenta con un algoritmo diagnóstico de dos pasos: primero se aplican las primeras dos preguntas (PHQ-2), y luego, si la puntuación es superior a algún criterio (i.e. Puntuar dos o más en cualquiera de los primeros dos ítems), se continúa con el resto del instrumento.

Por lo tanto, emplear el PHQ-9 reduciría el tiempo para tamizaje de depresión lo cual permitiría evaluar más personas en menor tiempo. Además, instrumentos más cortos son más apropiados para evaluar pacientes hospitalizados o en quimioterapia, pues su situación de salud puede impedir una evaluación extensa.

En segundo lugar, la validación del PHQ-9 como instrumento de tamizaje de depresión sería de utilidad en la implementación de algoritmos de diagnóstico y tratamiento de depresión y malestar emocional de los pacientes oncológicos de acuerdo a las recomendaciones de diversas GCP internacionales y los lineamientos del MINSA (Andersen et al., 2014; Holland et al., 2018; Howell et al., 2015; Ministerio de Salud y Organización Panamericana de Salud, 2013).

En un plano metodológico, este estudio propone validar un instrumento para realizar tamizaje de depresión en población adulta peruana con cáncer en el contexto hospitalario.

En el Perú, Calderón et al. (2012) evaluaron el contenido de la versión española del PHQ-9, desarrollando la versión peruana del PHQ-9 empleada en esta investigación. Gil (2019) hizo un estudio cualitativo del contenido de la versión peruana del PHQ-9 con gestantes usuarios del primer nivel de atención de salud y obstetras. Zhong, Gelaye, Fann, Sanchez, and Williams (2014) evaluaron el uso de

la versión española del PHQ-9 para medir depresión en mujeres embarazadas. Adicionalmente, se han estudiado las propiedades psicométricas de la versión española del PHQ-9 en pacientes usuarios de atención primaria en Chile y Ecuador (Baader et al., 2012), pacientes hospitalizados en Argentina, y estudiantes universitario en Colombia (Baader et al., 2012; Cassiani-Miranda, Vargas-Hernández, Pérez-Anibal, Herazo-Bustos, y Hernández-Carrillo, 2017; Fabara, 2013; Montenegro, Kilstein, y Guillermo, 2017).

Sin embargo, existen dos razones por las cuales los resultados de estas investigaciones no brindan evidencia de validez suficiente como para justificar el uso propuesto en esta investigación.

En primer lugar, la versión española del PHQ-9 que fue estudiada en Perú y en los otros países presenta contenido que no es el más apropiado para el contexto peruano (Calderón et al., 2012; Gil, 2019), por lo que podría ser una fuente de sesgo. Por ese motivo, en esta investigación se empleará la versión peruana del PHQ-9 desarrollada por Calderón et al. (2012).

En segundo lugar, las propiedades psicométricas halladas en una población no pueden extrapolarse a otra población, a menos que las poblaciones sean suficientemente similares. Por esos motivos, siempre es ideal investigar la evidencia de validez de un instrumento en el contexto en el cual planea ser empleada para evitar el sesgo de extrapolación (Maxim, Niebo, y Utell, 2014).

Además, hay razones para esperar que el PHQ-9 presente propiedades psicométricas distintas en la población oncológica peruana en comparación a otras poblaciones peruanas y latinoamericanas. Es probable que haya una confluencia de

síntomas somáticos y psicológicos en los pacientes con cáncer que no se da en otras poblaciones (Maté, Hollenstein, y Gil, 2004). Por ejemplo, los pacientes oncológicos suelen presentar sintomatología depresiva relacionada al aumento de citoquinas pro-inflamatorias asociadas al cáncer (Breitbart et al., 2014; Sotelo, Musselman, y Nemeroff, 2014). Esto podría afectar la estructura interna del instrumento, las relaciones test-criterio, y otras propiedades psicométricas.

Por otra parte, algunos de los estudios antes mencionados emplean las técnicas de análisis factorial de manera inadecuada, empleando métodos de extracción no apropiados y/o no reportando la bondad de ajuste de sus modelos (Baader et al., 2012; Cassiani-Miranda et al., 2017). Y finalmente, solo una investigación estudió la invarianza de medida, ya sea a nivel de escala o funcionamiento diferencial de los ítems (DIF) (Zhong et al., 2014).

De esta manera, la presente investigación se propone evaluar la validez de las puntuaciones de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9) para tamizaje de depresión en pacientes adultos con cáncer de un Hospital público especializado de Lima metropolitana.

3. LIMITACIONES DE LA INVESTIGACIÓN

Esta investigación se limita a discutir la validez del PHQ-9 para el uso propuesto, es decir, como instrumento de tamizaje de depresión en pacientes adultos con cáncer de un Hospital público especializado de Lima metropolitana. No se sabe qué tan válidos sean otras interpretaciones de las puntuaciones del PHQ-9 (e.g. para medir la severidad de sintomatología depresiva empleando cuatro puntos de corte en una población de estudiantes universitarios). En ese sentido, es responsabilidad

de futuros investigadores realizar un argumento para la validez de algún uso propuesto del instrumento distinto del planteado por esta investigación.

Además, para evaluar la sensibilidad y la especificidad del PHQ-9, se emplearán las puntuaciones del HADS. Por lo tanto, los puntos de corte identificados en esta investigación deben ser usados con cautela. Se sugiere que futuras investigaciones evalúen la presencia de psicopatologías mediante entrevistas diagnósticas para poder evaluar la precisión diagnóstica del PHQ-9 de una manera más certera.

4. OBJETIVOS DE LA INVESTIGACIÓN

4.1. Objetivo general

Evaluar la medida en la que la evidencia de validez de las puntuaciones de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9) justifica su uso para tamizaje de depresión en una población de pacientes adultos con cáncer de un Hospital de atención especializada de Lima metropolitana.

4.2. Objetivos específicos

- Identificar la evidencia de validez basada en el contenido de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9) para su uso propuesto mediante el método de criterio de jueces.
- Identificar la evidencia de validez basada en la estructura interna de las puntuaciones de la versión peruana del Cuestionario modificado sobre la Salud del Paciente (PHQ-9) para su uso propuesto evaluando la dimensionalidad mediante Análisis Factorial Exploratorio y Confirmatorio.

- Identificar la evidencia de validez basada en la estructura interna de las puntuaciones de la versión peruana del Cuestionario modificado sobre la Salud del Paciente (PHQ-9) para su uso propuesto mediante análisis de la confiabilidad tau-equivalente alfa de Cronbach y confiabilidad congénica mediante el coeficiente omega de McDonald.
- Identificar la evidencia de validez basada en la estructura interna de las puntuaciones de la versión peruana del Cuestionario modificado sobre la Salud del Paciente (PHQ-9) para su uso propuesto evaluando la invarianza de medida entre los grupos mediante Análisis Factorial Confirmatorio Multi-Grupo (AFCMG).
- Identificar la evidencia convergente de validez de las puntuaciones de la versión peruana del Cuestionario modificado sobre la Salud del Paciente (PHQ-9) para su uso propuesto mediante la correlación del PHQ-9 con HADS y autoreporte de diagnóstico depresión.
- Corroborar el punto de corte más adecuado de la versión peruana del Cuestionario modificado sobre la Salud del Paciente (PHQ-9) para su uso propuesto.
- Contrastar el poder diagnóstico del método directo y del método de dos pasos de calificación de las puntuaciones de la versión peruana del Cuestionario modificado sobre la Salud del Paciente (PHQ-9).
- Contrastar el poder diagnóstico de la versión peruana del Cuestionario modificado sobre la Salud del Paciente (PHQ-9) y su versión abreviada (PHQ-2).

CAPÍTULO II: MARCO TEÓRICO

1. ASPECTOS CONCEPTUALES

1.1. Depresión

Kraepelin, quien sentó los cimientos de las clasificaciones psiquiátricas modernas hacia el final del siglo XIX, indicó que la “demencia melancólica-maníaca” es una entidad patológica explicada mediante un modelo médico y neurológico, con una etiología y patología orgánicas y específicas (Paykel, 2008). La concepción Kraepeliniana de los trastornos mentales sería un precursor del modelo biomédico moderno.

Posteriormente, Meyer a inicios del siglo XX, luego de reemplazar el término melancolía con el término “depresión”, indicaría que los trastornos mentales como la depresión no tienen una etiología orgánica ni específica, ni son la consecuencia de una patología neurológica (Meyer, 1910). Él concibió la depresión como una reacción afectiva a eventos vitales estresantes en los que influyen factores biológicos, psicológicos y sociales (Ban, 2014). La concepción Meyeriana de los trastornos mentales sería un precursor del modelo biopsicosocial moderno.

Por otra parte, la obra seminal “Duelo y Melancolía” escrita por Freud (1917) desarrollaría lo que vendría a ser la perspectiva psicoanalítica de la depresión, indicando que “las características distintivas de la melancolía [depresión] son el abatimiento, pérdida de interés en el mundo exterior, pérdida de la capacidad para amar, inhibición de la actividad, y una disminución del autoestima hasta el grado de expresar autoreproches y autodesprecio, y culmina en una

expectativa delirante de castigo” (Freud, 1917). De acuerdo a esta perspectiva, la melancolía estaría asociada a pérdidas reales o simbólicas de objetos de deseo, cuyos procesos de duelo tomaban lugar en el inconsciente, por lo que no eran completamente comprendidos o identificados. A diferencia de la concepción Kraepeliniana, la escuela psicoanalista consideraba la etiología de la depresión casi exclusivamente psicógena.

La escuela cognitiva planteó que los síntomas afectivos, motivacionales y comportamentales de la depresión se deben a pensamientos automáticos negativos que las personas tienen sobre sí mismas, el mundo, y el futuro (Beck, 2008). Estos pensamientos son automáticos porque son aceptados de manera acrítica y negativos porque distorsionan la información de tal manera que ignoran lo positivo y resaltan los estímulos negativos. Además, estos pensamientos se suscitan debido a creencias disfuncionales preexistentes que funcionan como un factor de diátesis, los cuales se activan a causa de eventos vitales estresantes (Beck, 2008). Si bien los síntomas de la depresión son característicamente afectivos, de acuerdo a la teoría cognitiva, el trastorno es mejor entendido como una distorsión del pensamiento, y a la luz de las investigaciones en neurobiología de la depresión, el enfoque cognitivo señala que la depresión se debe a una integración de factores biopsicosociales (Beck, 2008).

En la actualidad no existe consenso sobre la naturaleza de la depresión, es decir, si es un tipo de reacción afectiva (posición Meyeriana) o una entidad patológica (posición Kraepeliniana), o si su etiología es psicógena u orgánica, ni hay completo acuerdo sobre la naturaleza central de los síntomas (afectivos, cognitivos o somáticos), pero sí hay consenso sobre los síntomas que configuran la

depresión sobre los cuales se proponen las diferentes definiciones (Alford y Beck, 2009).

Tabla 1. Criterios diagnósticos de depresión según DSM V y CIE-10.

	DSM V	CIE 10
Síntomas Esenciales	1) Estado de ánimo depresivo la mayor parte del día, casi cada día	1) Estado de ánimo depresivo
	2) Disminución acusada del interés o de la capacidad para el placer en todas o casi todas las actividades, la mayor parte del día, casi cada día	2) Pérdida de la capacidad de interesarse y disfrutar de las cosas. 3) Disminución de la vitalidad que lleva a una reducción del nivel de actividad, cansancio exagerado y fatigabilidad.
Síntomas Secundarios	3) Pérdida importante de peso sin hacer régimen o aumento de peso, o pérdida o aumento del apetito	4) Disminución de la atención y concentración
	4) Insomnio o hipersomnia	5) Pérdida de la confianza en sí mismo y sentimientos de inferioridad
	5) Agitación o retraso psicomotor (observable por otros)	6) Ideas de culpa y de ser inútil
	6) Fatiga o pérdida de energía	7) Perspectiva sombría del futuro
	7) Sentimientos de inutilidad o de culpa excesivos o inapropiados	8) Pensamientos y actos suicidas o autoagresiones
	8) Disminución de la capacidad para pensar o concentrarse, o indecisión	9) Trastorno del sueño
	9) Pensamientos recurrentes de muerte, ideación suicida recurrente sin un plan específico o una tentativa de suicidio o un plan específico para suicidarse	10) Pérdida del apetito

Fuente: Organización Mundial de la Salud (2017) y American Psychiatric Association (1995)

El Manual Diagnóstico y Estadístico de los Trastornos Mentales (DSM) de la American Psychiatric Association (APA) define la depresión como un trastorno mental caracterizado por: “un estado de ánimo deprimido o una pérdida de interés

o placer en casi toda las actividades... cambios de apetito o peso, del sueño y de la actividad psicomotora; falta de energía; sentimientos de infravaloración o culpa; dificultad para pensar, concentrarse o tomar decisiones, y pensamientos recurrentes de muerte o ideación, planes o intentos suicidas... acompañado de un malestar clínico significativo o de deterioro social, laboral o de otras áreas importantes de la actividad del individuo” (American Psychiatric Association, 1995).

Similarmente, la Clasificación Internacional de Enfermedades (CIE) de la Organización Mundial de la Salud (OMS) define la depresión como un trastorno mental caracterizado por la presencia de tristeza, pérdida de interés o placer, sentimientos de culpa o falta de autoestima, trastornos del sueño o del apetito, sensación de cansancio y falta de concentración, que puede ser de larga duración y crónica o episódica y recurrente, perjudicando la capacidad para funcionar en los diferentes ámbitos de la vida y la capacidad para poder afrontar las dificultades de la vida, y en su forma más severa, puede llevar al suicidio (Organización Mundial de la Salud, 2017).

Paykel (2008) define el concepto moderno de depresión como un síndrome clínico con síntomas especificados en las clasificaciones diagnósticas (ver Tabla 1). Las versiones recientes de estos manuales de diagnóstico presentan un trastorno mental unitario que consiste de un síndrome central llamado “Episodio depresivo” en el CIE-10 (Organización Mundial de la Salud, 2000) y “Depresión mayor” en el DSM-V (American Psychiatric Association, 2014). Las definiciones del DSM y del CIE son similares, y se consideran por fines prácticos como equivalentes, aunque presenten algunas diferencias.

Estos manuales ofrecen descripciones de los trastornos mental y de sus criterios diagnósticos para que los profesionales de la salud mental puedan sobrellevar la subjetividad y la ambigüedad de los procesos diagnósticos, reducir la varianza de los criterios clínicos entre los profesionales de la salud mental, y facilitar un proceso diagnóstico más fiable (Maj, 2013).

En el caso del DSM-V, la depresión se diagnostica cuando una persona presenta un total de 5 o más de los síntomas de la depresión, de los cuales uno o más son síntomas esenciales (American Psychiatric Association, 2014). Los síntomas esenciales son dos: Estado de ánimo depresivo y anhedonia/abulia. Además, los síntomas deben durar la mayor parte del día, cada día, durante por lo menos dos semanas. Debe haber un malestar clínicamente significativo o un claro deterioro funcional del paciente con respecto a su estado previo. Los síntomas no se pueden atribuir a efectos fisiológicos del consumo de alguna sustancia u otra afectación médica, psicótica ni debe haber antecedentes de hipomanía o manía. Una mayor cantidad e intensidad de síntomas indica que el cuadro depresivo es más severo.

En el caso del CIE-10, la depresión se diagnostica cuando una persona presenta un total de 4 o más de los síntomas de la depresión, de los cuales dos o tres son síntomas esenciales, y dos o más son síntomas secundarios (Organización Mundial de la Salud, 2000). Los síntomas esenciales son tres: estado de ánimo depresivo, anhedonia/abulia y astenia. Similarmente, los síntomas deben durar por lo menos dos semanas, no deben haber antecedentes de episodios maníacos o hipomaniacos, y no debe ser atribuible al consumo de sustancias psicoactivas o

algún trastorno mental orgánico. Una mayor cantidad e intensidad de síntomas significa que el cuadro depresivo es más severo.

1.1.1 Depresión en pacientes oncológicos

En cuanto a la población peruana general, el Instituto Nacional de Salud Mental (2009) indica en el informe general del estudio epidemiológico de salud mental que en el 2002 se identificó una prevalencia de vida de depresión de 18.2% en la población de Lima y Callao. Similarmente, en el 2003 y 2005 se hallaron resultados similares en población urbana de la Sierra y Selva (desde 16% hasta 21.4%). Asimismo, en el 2008 se identificó en población adulta de la Sierra rural una prevalencia de vida de trastornos depresivos de 15.1% y una prevalencia actual de trastornos depresivos de 1,4%. El Instituto Nacional de Salud Mental (2013) en la última publicación del informe general del estudio epidemiológico de salud mental en Lima metropolitana y Callao se halló una prevalencia de vida y actual de depresión de 17.3% y 2.8% respectivamente. En otras palabras, la evidencia sugiere que 1 de cada 5 peruanos ha experimentado depresión en algún momento de su vida, y aproximadamente 2% de la población peruana padece actualmente de depresión.

En cuanto a pacientes oncológicos en otros países, investigaciones internacionales han encontrado prevalencias muy variables dependiendo de la metodología empleada.

Krebber et al. (2014) realizaron un meta-análisis incluyendo 211 estudios con un total de 82,426 pacientes oncológicos. En general, la prevalencia variaba dependiendo del método de evaluación, sea HADS-D \geq 8 (18%), HADS-D \geq 11 (7%), CES-D (24%), o entrevistas diagnósticas (14%). Empleado entrevistas

diagnósticas, las menores prevalencias medias se presentaron en pacientes con cáncer del tracto respiratorio 3% (IC_{95%}=2–6%) y hematológicos 8% (IC_{95%}=5–11%), y las mayores prevalencias se presentaron en pacientes con cáncer de cerebro 28% (IC_{95%}=20–38%) y de genitales femeninos 23% (IC_{95%}=15–33%), aunque estos resultados presentan una alta varianza debida al error muestral (I₂=0%). Las prevalencias obtenidas mediante entrevistas con menor error muestral (I₂≥75%) se ubicaban entre 11% a 13%. Empleando instrumentos autoaplicados, las menores prevalencias se presentaron en pacientes con cáncer de piel 7% (IC_{95%}=4–11%) y de genitales masculinos 10% (IC_{95%}=8–13%), y las mayores prevalencias se presentaron en pacientes con cáncer de huesos y tejidos blandos 33% (IC_{95%}=21–48%), de tracto digestivo 27% (IC_{95%}=18–37%), y de genitales femeninos 26% (IC_{95%}=18–35%).

Maass, Roorda, Berendsen, Verhaak, and de Bock (2015) realizaron una revisión sistemática de 17 estudios de sobrevivientes de cáncer con una población total de 12,499 personas. Se halló una prevalencia de depresión entre 9.4% y 66.1% con un promedio general de 39.9%, aunque en los estudios que emplearon la HADS-D no identificaron diferencias significativas en la prevalencia de depresión en mujeres sobrevivientes de cáncer y la población general. Finalmente, se encontró que la prevalencia de depresión no cambia en función al tiempo de diagnóstico.

Watts et al. (2014) realizaron un meta-análisis de 27 estudios comprendiendo un total de 4,494 pacientes con cáncer de próstata. Se halló una prevalencia de depresión de 17.27% (IC_{95%}=15.06–19.72%) antes del tratamiento,

14.7% (IC_{95%}=11.92–17.99%) durante el tratamiento, y 18.44% (IC_{95%}=15.18–24.31%) después del tratamiento.

Similarmente, Watts, Prescott, Mason, McLeod, and Lewith (2015) realizaron un meta-análisis de 24 estudios comprendiendo un total de 3,623 pacientes con cáncer de ovario. Se halló una prevalencia de depresión de 25.34% (IC_{95%}=22.79–28.07%) antes el tratamiento, 22.99% (IC_{95%}=19.85–26.46%) durante el tratamiento, 12.71% (IC_{95%}=10.14–15.79%) después del tratamiento.

Finalmente, Walker et al. (2013), realizaron una revisión sistemática de 15 estudios de prevalencia de depresión en pacientes oncológicos que usaron entrevistas diagnósticas y otros criterios para asegurar la calidad de los datos. El tamaño muestral de los estudios se encontraba entre 100 a 3,938 personas, con una media de n=399 y mediana de n=129. Se encontraron prevalencias de depresión desde 5% a 16% en pacientes ambulatorios, 4% a 14% en pacientes hospitalizados, y de 7% a 49% en pacientes en cuidados paliativos.

De acuerdo con lo hallado, la depresión en pacientes oncológicos es más común en pacientes mujeres que hombres, mayor en pacientes en cuidados paliativos, mayor en pacientes con cáncer de páncreas, y menor en pacientes durante su tratamiento. Adicionalmente, los estudios que usan pruebas psicométricas suelen hallar mayor prevalencia de depresión que estudios que usan entrevistas semi-estructuradas lo cual sugiere mayor sensibilidad y/o menor especificidad de los instrumentos psicométricos.

En cuanto a pacientes oncológicos peruanos, se ha identificado una prevalencia de depresión de 21.29% en una muestra de 385 pacientes mujeres con

cáncer de mama de un hospital público de Lima mediante la escala PHQ-9 (Berrospi-Reyna et al., 2017), aunque no se puede garantizar la fiabilidad ni precisión del instrumento por lo que la cifra es referencial. Vilela-Estrada (2019) realizó un estudio con una muestra de 500 participantes, y halló una prevalencia de depresión de 9.79% empleando el HADS. Además, estudios correlacionales anteriores con muestras pequeñas ($n < 50$) identificaron prevalencias de depresión alrededor del 30% (Cabrera-Gutierrez, Mendoza-Luna, y Obando-Castro, 2010; Tateishi, 2012).

A pesar de la falta de meta-análisis y revisiones sistemáticas en la población oncológica peruana, la información existente indica que la prevalencia de depresión en pacientes oncológicos es entre 3 a 10 veces mayor que la prevalencia de depresión en la población general.

Por otra parte, los pacientes con depresión suelen experimentar una serie de síntomas somáticos que son esenciales para realizar un diagnóstico, como pérdida del apetito y del peso, fatiga, o pérdida del interés sexual (American Psychiatric Association, 1995, 2014; Organización Mundial de la Salud, 2000). Sin embargo, estos indicadores carecen la misma utilidad diagnóstica en los pacientes oncológicos porque estos síntomas son comunes a las enfermedades neoplásicas (Maté et al., 2004).

Una de las razones por las que los pacientes con cáncer presentan sintomatología somática depresiva es el efecto biológico del tumor. La sintomatología depresiva elevada de los pacientes oncológicos no puede explicarse solamente debido a estrés psicosocial, sino, se asocia al rol de citoquinas pro-

inflamatorias que inducen síntomas pseudo-depresivos (Sotelo et al., 2014). Por ejemplo, Breitbart et al. (2014) identificaron una relación directa entre la presencia de citoquinas pro-inflamatorias y sintomatología depresiva ($r=.34$; $p<.01$) en pacientes con cáncer de páncreas. Esto indica que los pacientes con este tipo específico de cáncer son más propensos a manifestar sintomatología depresiva sea que cursen o no con un cuadro de depresión, y aquellos que cursan con un cuadro de depresión experimentan sintomatología depresiva somática más severa.

Finalmente, se entiende que el diagnóstico de depresión es de criterio clínico, es decir, los criterios diagnósticos presentes en las clasificaciones nosológicas son guías flexibles y no son rígidas, y se espera que el proceso diagnóstico sea realizado por un especialista en salud mental que juzgue la medida en la que los criterios se cumplen o no, evaluando la severidad de los síntomas, el grado en que generan malestar emocional y perjudican la funcionalidad, entre otros factores (American Psychiatric Association, 1995, 2014; Chochinov, Wilson, Enns, y Lander, 1994; Maj, 2013). Esto es particularmente importante considerando la mezcla de síntomas somáticos y afectivos en los pacientes con cáncer producto de enfermedades oncológicas, tratamientos quimioterapéuticos, reacciones emocionales normales, duelos, trastornos adaptativos, entre otros.

1.1.2 Instrumentos validados para tamizaje de depresión en población hospitalaria en el Perú

La primera versión del Inventario de Depresión de Beck (BDI-IA) se diseñó en base a observaciones clínicas sistemáticas de actitudes y síntomas de pacientes psiquiátricos deprimidos con el propósito de detectar y evaluar la gravedad de la

depresión en personas adultas (Beck, Ward, Mendelson, Mock, y Erbaugh, 1961). Es una escala de tipo auto-informe compuesto por 21 ítems de tipo Likert de cuatro puntos con un tiempo de aplicación de 5 a 10 minutos, y mide dos dimensiones (Somática-ejecutiva y Cognitivo-afectiva). El BDI-IA ha sido validado en el Perú por Vega-Dienstmaier, Coronado-Molina, and Mazzotti (2014) empleando una muestra de pruebas del año 1998 empleadas para diagnosticar depresión en pacientes hospitalizados de medicina general, determinando una adecuada confiabilidad de consistencia interna ($\alpha=.998$), una adecuada sensibilidad ($Se=87.5\%$) y alta especificidad ($Es=98.21\%$).

El Cuestionario de Auto-Reporte (SRQ-18) fue desarrollado para evaluar el malestar psicológico inespecífico de tal manera que identifica la severidad de la sintomatología depresiva/ansiosa sin necesariamente proveer un diagnóstico (Harding et al., 1980). El instrumento es un auto-informe que consta de 28 ítems con un formato de respuesta dicotómico (“Sí” o “No”), los primeros 18 ítems exploran sintomatología neurótica (i.e. depresiva y ansiosa), los ítems 19 al 22 exploran sintomatología psicótica, el 23 explora trastornos convulsivos y del 24 al 28 exploran problemas relacionados al consumo de alcohol. El SRQ-18 fue validado en el Perú para detectar pacientes con probables trastornos psiquiátricos por Véliz (1998). Empleando el coeficiente de concordancia (α), el índice Kappa (K) y la prueba Z, se determinó una alta confiabilidad inter-evaluadores ($\alpha=.88$, $K=.76$, $Z=28.25$, $p<.001$). Además, se halló evidencia de validez concurrente con el diagnóstico de pacientes psiquiátricos que acuden a atención ambulatoria del servicio de psiquiatría del mismo hospital ($\alpha=.83$, $K=.65$, $Z=20.96$, $p<.001$) y menor en el grupo control compuesto de personas supuestamente sanas ($\alpha=.62$,

$K=.62$, $Z=2.84$, $p<.001$), lo cual sugiere menor especificidad. También se determinó que la evidencia de validez del constructo teórico de la sub-escala de neurosis es adecuada empleando la prueba chi-cuadrada χ^2 ($p<.001$) para evaluar el acuerdo entre la presencia o ausencia de un síntoma de acuerdo al instrumento y la identificación como paciente psiquiátrico o supuestamente sano para cada ítem.

La Escala de Depresión del Centro Epidemiológico (CES-D) fue desarrollado con el objetivo de medir la sintomatología depresiva actual en grupos de la población general con énfasis en los síntomas afectivos (Radloff, 1977). El instrumento es una escala auto-informe compuesta por 20 ítems de tipo Likert que puntúan desde el 0 hasta el 3, que mide cuatro dimensiones latentes (afectos depresivos, afectos positivos, somática, e interpersonal). Ruiz-Grosso et al. (2012) realizaron la validación del test para medir depresión en población hospitalaria peruana, encontrando una confiabilidad adecuada ($\alpha=.93$; $\omega=.90$), correlaciones aceptables ítem-test ($r=.36-.83$), ítem-rest ($r=.29-.81$), y una covarianza promedio inter-ítem entre .50 y .45, y una elevada precisión diagnóstica (auROC=.83; IC95%=.76-.90). Adicionalmente, para detectar pacientes con depresión entre pacientes con otros trastornos psiquiátricos se recomendó un punto de corte ≥ 29 (Se=77.1%; Es=79.4%), y para distinguir pacientes con depresión de pacientes sin evidencia de trastornos psiquiátricos se recomendó un punto de corte ≥ 24 (Se=91.4%; Es=96.7%).

La Escala de Depresión de Zung (ZSDS) fue desarrollada con el propósito de evaluar la depresión como un trastorno psiquiátrico identificando su presencia o ausencia y su severidad (Zung, 1965). El instrumento consta de 20 ítems con un

formato de respuesta Likert que puntúa del 1 al 4 y mide una sola dimensión. La ZSDS fue validada en el Perú para medir depresión en el contexto hospitalario general por Ruiz-Grosso et al. (2012). Se halló una confiabilidad adecuada ($\alpha=.89$; $\omega=.75$), correlaciones aceptables ítem-test ($r=.18-.77$), ítem-rest ($r=.10-.73$), una covarianza promedio inter-ítem se ubicaba entre 0.26 y 0.30, y una elevada precisión diagnóstica (auROC=.84; IC95%=.76-.91). Para detectar pacientes con depresión entre pacientes con otros trastornos psiquiátricos se recomendó un punto de corte ≥ 47 (Se=85.7%; Es=71.4%), y para distinguir pacientes con depresión de pacientes sin evidencia de trastornos psiquiátricos se recomendó un punto de corte ≥ 45 (Se=91.8%; Es=91.6%).

Finalmente, la Escalada de Ansiedad y Depresión Hospitalaria (HADS) fue diseñada para evaluar los síntomas cognitivos y conductuales de la ansiedad y la depresión en ambientes hospitalarios (Zigmond y Snaith, 1983). Esta es una fortaleza del HADS por sobre las otras escalas, pues no depende de los síntomas somáticos como fatiga, cambio del peso y apetito, los cuales son comórbidos con una gran cantidad de enfermedades y malestares, y por lo tanto, una fuente de sesgo. El instrumento cuenta con 14 ítems, conformando dos sub-escalas, una de ansiedad y otra de depresión de 7 ítems cada uno. Los ítems se puntúan en escala Likert del 0 al 3, con un rango de puntuación total del 0 al 21. Un estudio realizado por Vilela-Estrada (2019) indagó la evidencia de validez del HADS para medir depresión y ansiedad en pacientes oncológicos, evidenciando una adecuada parsimonia y ajuste al modelo bidimensional (CIM/DF>2; AGFI=.929; CFI=.935; RMSEA=.053; ITL=.923; IFI=.936), con saturaciones moderadas y elevadas (.38-.69), una adecuada fiabilidad ($\omega_1=.79$; $\omega_2=.77$), y correlaciones estimadas fuertes entre el

HADS y el BDI (.89) y el BAI (.84) (Vilela-Estrada, 2019). Considerando que el HADS es el test usado por el personal de la institución donde se realiza el estudio como referente diagnóstico, y es el instrumento con mayor evidencia de validez para ser usado como instrumento diagnóstico de depresión en pacientes peruanos oncológicos, se puede considerar al HADS el gold standard para medir depresión.

1.2. Validez

Los *Standards for educational and psychological testing* (Standards) son una serie de guías para la evaluación psicométrica y educativa producidas por la *American Psychological Association* (APA), la *American Educational Research Association* (AERA) y el *National Council of Measurements in Education* (NCME).

Elosua (2003) identificó tres etapas en la historia de la teoría de la validez mediante el análisis de las ediciones de los Standards desde 1974 hasta 1999. Al momento de la elaboración de esta investigación, la última edición de los Standards fue publicada en el 2014 (American Educational Research Association, American Psychological Association, y National Council on Measurement Education, 2014).

Primero, la etapa operacional, dominada por pragmatismo y el concepto de validez externa, primando las correlaciones de la prueba con otras variables para determinar si un test es válido o no. Se identificaron tres tipos de validez: contenido, constructo y criterio, y la validez referida al criterio englobaría la validez predictiva y concurrente.

Luego, la etapa teórica, caracterizada por la notabilidad otorgada a la teoría psicológica que precede lógicamente a la prueba, por lo cual el análisis de validez de constructo se transformaría en el eje principal de los otros aspectos como la validez de contenido y de criterio.

Finalmente, la fase contextual, en la cual se amplían las acepciones anteriores y se integra el concepto de “uso propuesto” de la prueba en relación a la “interpretación propuesta” de las puntuaciones.

La última versión de los estándares señala que la validez es un constructo unitario definido como “...el grado en que la evidencia y la teoría apoyan la interpretación de las puntuaciones de un test”; en ese sentido, un test no es inherentemente válido o inválido porque no se valida el test en sí mismo, sino el uso que se le da a las interpretaciones de las puntuaciones del test (American Educational Research Association et al., 2014).

En ese sentido, el concepto moderno de validación consiste en acumular evidencia relevante para proveer un sustento científico que justifique una interpretación específica de las puntuaciones de un test para un propósito específico (American Educational Research Association et al., 2014). Para este propósito se señalan cinco fuentes de evidencia de validez: la evidencia de validez basada en el contenido del test, basada en los procesos de respuesta a los ítems, basada en la estructura interna del test, basada en la relación con otras variables, y basada en las consecuencias de la aplicación del test; cabe señalar que no todas las fuentes de validez deben ser exploradas para todos los instrumentos psicométricos, sino,

cuando es relevante (American Educational Research Association et al., 2014; Sireci y Padilla, 2014).

1.2.1 Evidencia de validez basada en el contenido del test

La evidencia de validez basada en el contenido del test, usualmente llamada “validez de contenido”, se define como el grado en el cual el contenido de un test es congruente con el propósito del test (American Educational Research Association et al., 2014; Sireci y Faulkner-Bond, 2014). En ese sentido, la evidencia de validez de contenido se relaciona con cuatro elementos de la calidad de un test: definición del dominio, representación del dominio, relevancia del dominio, y adecuación del proceso de desarrollo del test (Sireci y Faulkner-Bond, 2014).

La definición del dominio se refiere a la definición operacional concreta mediante la cual se mide el dominio o constructo teórico que el test pretende medir (Elosua, 2003; Sireci y Faulkner-Bond, 2014). Evaluar la definición del dominio del test involucra buscar un consenso sobre la relación entre la definición operacional que subyace a un test y el constructo que el test pretende medir tal como es comprendido por expertos en la disciplina en cuestión (Sireci y Faulkner-Bond, 2014), lo cual puede realizarse mediante revisión sistemática de la literatura pre-existente y/o mediante el juicio de jueces expertos.

La representación de dominio se refiere al grado en el cual un test representa y mide adecuadamente el dominio tal como es definido en las especificaciones del test, y la relevancia de dominio se refiere al grado en el cual cada ítem es relevante al constructo que el test pretende medir (Sireci y Faulkner-Bond, 2014). En ese sentido, se busca que un test contenga ítems que midan todos los aspectos

importantes del dominio que busca medir, y que no contenga ítems triviales o contenido irrelevante, pues este genera sesgo de medición (American Educational Research Association et al., 2014; Sireci y Faulkner-Bond, 2014).

Para evaluar la representación y relevancia, se pueden realizar pruebas pilotos o se realizan entrevistas a los examinados para indagar como percibían el contenido del test, y también se pueden emplear jueces expertos en la disciplina en cuestión con el propósito de determinar si los ítems son congruentes con la definición conceptual, y por lo tanto, si miden apropiadamente el constructo que el test pretende medir, y si consideran que el contenido de los ítems es relevante al constructo que el test pretende medir (American Educational Research Association et al., 2014). Se puede hacer una evaluación cualitativa, o se puede cuantificar pidiendo que los jueces puntúen en una escala Likert el grado en el cual consideran que un ítem mide el constructo que pretende medir y qué tan relevante es, y las puntuaciones resultantes se pueden evaluar empleando el índice V de Aiken o coeficientes de confiabilidad inter-jueces para evaluar el acuerdo entre los evaluadores (Elosua, 2003; Sireci y Faulkner-Bond, 2014).

Considerando \bar{X} la media de las calificaciones de los jueces, l la mínima calificación de la escala, y k el rango utilizado, la fórmula para obtener el coeficiente V de Aiken es la siguiente (Cordon, 2015):

$$V = \frac{\bar{X} - l}{k}$$

Finalmente, la adecuación del proceso de desarrollo del test se refiere al proceso usado al construir el test para asegurar que su contenido represente de

manera adecuada el constructo que el test pretende medir, y que no mida elementos irrelevantes, empleando medidas para asegurar la calidad de los procedimientos al desarrollar el test, y si existe una explicación adecuada para los formatos y procedimientos empleados (Sireci y Faulkner-Bond, 2014). Ejemplos de medidas para el control de la calidad son el uso de paneles de expertos para revisión del contenido y calidad de los ítems, o pruebas pilotos de los ítems para seleccionar aquellos que midan de manera más adecuada el constructo que se pretende medir y para retirar aquellos que sean más difíciles de entender.

Aparte de estos cuatro elementos, también se debe considerar la situación del test, pues es importante señalar el formato de los ítems, cómo se administra el test, y cómo se puntúa con el propósito de evitar fuentes de sesgo asociado a elementos irrelevantes al propósito del test (Elosua, 2003).

1.2.2 Evidencia de validez basada en el proceso de respuesta

La evidencia basada en el proceso de respuesta se define como la evidencia relativa al acuerdo entre el constructo que el test pretende medir y la naturaleza del desempeño o el proceso de respuesta en el cual se involucran los examinados (American Educational Research Association et al., 2014; Padilla y Benítez, 2014). La evidencia de validez basada en el proceso de respuesta se refiere al acuerdo entre los procesos cognitivos que deberían suscitarse frente a los ítems del test para poder medir adecuadamente el constructo que el test pretende medir, y las operaciones cognitivas que realmente toman lugar durante la resolución del test.

En ese sentido, de la misma manera que puede haber sesgo de medición debido a contenido irrelevante al constructo que el test pretende medir, también

puede ocurrir que un ítem suscite una variedad de respuestas diferentes a las previstas o que se encuentren formas de solucionar el test distintas a las previstas (American Educational Research Association et al., 2014; Padilla y Benítez, 2014).

Esta fuente de evidencia de validez debería ser estudiada si las interpretaciones de las puntuaciones de un test psicológico dependen de premisas acerca de los procesos psicológicos y operaciones cognitivas empleadas por los evaluados (American Educational Research Association et al., 2014). Los investigadores deberían evaluar la evidencia basada en los procesos de respuesta para justificar un uso particular de una prueba cuando el uso de la prueba es justificado a través de las siguientes proposiciones (Padilla y Benítez, 2014):

- El desempeño de los evaluados en los ítems del test refleja procesos psicológicos y/o operaciones cognitivas descritas en las especificaciones del test.
- El proceso de los jueces y observadores al evaluar el desempeño de los evaluados es consistente con el uso propuesto de las interpretaciones de las puntuaciones.
- Los grupos de evaluados separados por demografía, idioma u otras condiciones asociadas al uso propuesto del test, no eran diferentes en la naturaleza de su desempeño o en sus respuestas debido a fuentes de varianza por contenido irrelevante al constructo.

Para poder medir el proceso de respuesta, se pueden emplear métodos de entrevista a los evaluados sobre sus estrategias de respuesta, documentar el proceso de respuesta de los ítems, y el uso de criterio de jueces externos, entre otros

(American Educational Research Association et al., 2014). Además, se pueden catalogar los métodos en dos grupos (Padilla y Benítez, 2014): los que brindan acceso directo a los procesos psicológicos y operaciones cognitivas (e.g. pensar el alto, grupo focal, y entrevistas), y aquellos que ofrecen indicadores indirectos que requieren inferencia (e.g. eye tracking y tiempo de respuesta).

1.2.3 Evidencia de validez basada en la estructura interna de la prueba

La evidencia de validez basada en la estructura interna del test se define como el grado en el cual las relaciones entre los ítems y los componentes del test son congruentes con el constructo sobre el cual se basan las interpretaciones propuestas de las puntuaciones del test (American Educational Research Association et al., 2014). Según Rios and Wells (2014), hay tres aspectos básicos de la estructura interna: dimensionalidad, invarianza de medida, y confiabilidad.

En primer lugar, la dimensionalidad se refiere al grado en el cual la covarianza entre las puntuaciones de los ítems evidencia los constructos latentes que el test pretende medir y sobre los cuales se realizan las interpretaciones propuestas. Los métodos para evaluar la dimensionalidad de un test más usados son el Análisis Factorial Exploratorio (AFE) y el Análisis Factorial Confirmatorio (AFC), técnicas de análisis multivariante que pueden representarse mediante Modelos de Ecuaciones Estructurales (MES).

El AFE es un método que busca determinar el número de constructos latentes (o factores) que explican la varianza y covarianza entre un conjunto de variables observables (o ítems) para poder realizar interpretaciones y detectar la estructura interna no-observable de los datos, y suele ser usado cuando no se tiene

mucha información sobre la estructura de los datos debido a falta de investigación o cuando se realiza procesos de validación de modelos de medición (Matsunaga, 2010; Treiblmaier y Filzmoser, 2010). En el AFE, la estructura factorial no se decide a priori, sino, se extraen una cantidad de factores sobre la base de ciertos criterios, como el criterio Kaiser-Guttman ($\text{autovalor} > 1$), observando el gráfico de sedimentación de autovalores, u observando la varianza total acumulada explicada por los factores, y se calculan las cargas factoriales de todos factores con las variables observables (Costello y Osborne, 2005; Treiblmaier y Filzmoser, 2010).

El AFC es un método similar al AFE, sin embargo, el AFC especifica el modelo explícito determinando el número de factores latentes, la relación de cada ítem con su factor y otros parámetros. En ese sentido, el AFC busca evaluar si los datos se ajustan al modelo teórico establecido a priori (Brown, 2006; Matsunaga, 2010).

Acerca de los métodos de extracción, se recomienda la Factorización de ejes principales (PAF) o Mínimos cuadrados no ponderados (ULS) cuando los datos no son normales, y el método de extracción de Máxima Verosimilitud (ML) cuando los datos presentan normalidad multivariada, aunque es relativamente robusto cuando no se cumple el supuesto de normalidad multivariada (Costello y Osborne, 2005; Matsunaga, 2010; Treiblmaier y Filzmoser, 2010; Yang y Liang, 2013).

Por otra parte, los métodos de rotación tienen el propósito de simplificar y clarificar la estructura de los datos. Pueden ser de dos tipos: las rotaciones ortogonales, las cuales asumen que los factores latentes no están correlacionados, y las rotaciones oblicuas, que permiten que las correlaciones entre los factores sean

diferentes de cero. En general, se recomiendan las rotaciones oblicuas, como Promax (Brown, 2006; Matsunaga, 2010; Treiblmaier y Filzmoser, 2010).

Una vez que se determina que los datos se ajustan al modelo, se revisa las cargas factoriales de los ítems a los factores para corroborar la validez convergente y divergente. La validez convergente indica si los ítems que teóricamente miden un mismo constructo latente cargan alto en un mismo factor ($>.70$), y por lo tanto, miden adecuadamente el constructo que dicen medir, y la validez divergente indica si los ítems presentan cargas factoriales leves con factores distintos al constructo latente al cual teóricamente pertenecen ($<.30$), y por lo tanto, indica si un instrumento puede distinguir entre las diferentes dimensiones latentes de un constructo multidimensional (Brown, 2006).

En segundo lugar, la invarianza de medida señala qué tan “justas” son las mediciones de un test (Rios y Wells, 2014). En otras palabras, que no se evidencie un sesgo debido a deficiencias sistémicas en el test que conlleven interpretaciones diferenciales de las puntuaciones de un mismo test entre sub-grupos, pues lo deseado es que las mediciones entre los sub-grupos de una muestra sean equivalentes. Esto se puede realizar en dos niveles: A nivel de escala y a nivel de ítem (Elosua, 2003).

A nivel de ítem, se plantea la duda de si los ítems de la prueba están sesgados entre sub-grupos, en el sentido de que dos personas pertenecientes a sub-grupos diferentes en igualdad de condiciones, con un nivel idéntico del atributo o rasgo medido por el test, ejecutan de manera distinta un mismo ítem en virtud de pertenecer al grupo al el cual pertenecen, es decir, el ítem funciona diferencialmente

(DIF) (Hidalgo, Galindo, Inglés, Campoy, y Ortiz, 1999). Para evaluar el DIF, es decir, si se favorece a uno u otro grupo, se puede emplear el estadístico de Mantel-Haenszel (MH) si los ítems son dicotómicos (Elosua, 2003; Hidalgo et al., 1999).

A nivel de escala, se evalúa si hay invarianza de medida a nivel de los constructos propuestos del test, es decir, si las puntuaciones del test son equivalentes entre los sub-grupos de una población (Rios y Wells, 2014). Para evaluar la invarianza de medida a nivel de escala se puede emplear el AFCMG para evaluar los diferentes niveles de invarianza (Elosua, 2003; Rios y Wells, 2014; Xu y Tracey, 2017). Este procedimiento se realiza de manera sistemática determinando si cada modelo sucesivo, los cuales son menos laxos que el anterior, se diferencian o no significativamente del modelo base (Xu y Tracey, 2017). Primero, se evalúa la invarianza configural, la cual confirma si el patrón de los constructos latentes y sus respectivos ítems se mantiene entre los diferentes sub-grupos evaluando los diferentes índices de bondad de ajuste (Rios y Wells, 2014; Xu y Tracey, 2017). Luego, la invarianza métrica se confirma si los pesos factoriales de los ítems con sus respectivos factores son iguales entre los diferentes sub-grupos (Rios y Wells, 2014; Xu y Tracey, 2017). Si no se puede confirmar una invarianza métrica, entonces no tendría sentido comparar las medias de los constructos latentes entre sub-grupos porque los constructos no serían equivalentes (Xu y Tracey, 2017). La invarianza escalar indica que los interceptos de los ítems son equivalentes, pues de otra forma no se pueden realizar comparaciones directas entre las medias del os grupos (Rios y Wells, 2014; Xu y Tracey, 2017).

Por otra parte, la confiabilidad de consistencia interna se refiere al grado en el cual un test mide consistentemente aquello que indica medir (Tavakol y Dennick, 2011). El análisis de la confiabilidad de un test se realiza debido a que todo sistema de medición presenta un error inherente, y por lo tanto, es pertinente evaluar el grado en el que las puntuaciones de un test son debidas al error. En ese sentido, considerando que la puntuación observada es la suma de la puntuación verdadera y el error, la confiabilidad es operacionalmente definida como la proporción de la varianza de la puntuación verdadera y la varianza de la puntuación observada (Rios y Wells, 2014).

La confiabilidad de consistencia interna no es una medida de estructura interna ni estima la homogeneidad (unidimensionalidad) de un test. Las mediciones de consistencia interna presuponen una estructura unidimensional, mas no demuestran dicha estructura pues el coeficiente de confiabilidad de un test multidimensional no necesariamente es menor que el de un test unidimensional (Rios y Wells, 2014; Tavakol y Dennick, 2011).

Ahora bien, de acuerdo a la Teoría clásica de los test, la varianza observada de las puntuaciones se debe en parte a las diferencias verdaderas en el rasgo que está siendo medido y en parte debido al error, y debido a que las puntuaciones observadas son la suma entre dos variables latentes, hay un número infinito de formas en las que puede ser dividida la puntuación verdadera y el error, por lo que se deben hacer algunas suposiciones (Graham, 2006). La primera suposición es que las variables miden una misma variable latente verdadera, es decir, se presupone unidimensionalidad entre los ítems (Graham, 2006; Tavakol y Dennick, 2011).

Luego, hay otras suposiciones que dependen del modelo de medición que se emplea para estimar la confiabilidad de los datos.

El modelo paralelo es el modelo más restrictivo para definir la puntuación verdadera compuesta. Aparte de la suposición de unidimensionalidad, el modelo paralelo asume que todos los ítems son equivalentes, es decir, miden la misma variable, en la misma escala, con el mismo nivel de precisión y con la misma cantidad de error (Graham, 2006). En otras palabras, todos los ítems presentan la misma puntuación verdadera y el mismo error, y los ítems son equivalentes en contenido, proceso de respuesta y características estadísticas. Por ejemplo, en la Escala de Inteligencia Wechsler para adultos (WAIS) existen escalas que son intercambiables, es decir, son “paralelas”.

El modelo tau-equivalente es similar al modelo paralelo, pero es más laxo en sus supuestos. Al igual que el modelo paralelo, supone que los ítems individuales miden la misma variable latente (unidimensionalidad) en la misma escala con el mismo grado de precisión, pero con diferente error, y en ese sentido, se asume que toda la varianza única específica a un ítem se debe al error (Graham, 2006; Tavakol y Dennick, 2011). El modelo tau-equivalente supone que todos los ítems miden idénticamente la puntuación verdadera pero pueden presentar diferentes grados de error.

Es en base a las suposiciones del modelo tau-equivalente que se propone el uso del coeficiente α alfa de Cronbach. Una forma simple de expresar el coeficiente alfa de Cronbach es mediante la siguiente fórmula (Ventura-Leon y Caycho-Rodríguez, 2017):

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum S_i^2}{S_T^2}\right)$$

Considerando que K es la cantidad de ítems, $\sum S_i^2$ es la suma de la varianza de los ítems, y S_T^2 es la varianza total. Sin embargo, el coeficiente alfa de Cronbach presenta ciertas limitaciones. Primeramente, en la medida que la varianza de los ítems se aproxima a 0, el coeficiente alfa de Cronbach se acerca a 1, y por lo tanto, se asume que toda la varianza entre los ítems es debida al error aleatorio, o en otras palabras, se asume que las cargas factoriales de los ítems son iguales. Cuando las cargas factoriales de los ítems no son iguales, el coeficiente alfa de Cronbach subestima la confiabilidad del instrumento (Graham, 2006; Rios y Wells, 2014). Además, el coeficiente alfa de Cronbach es altamente sensible al número de ítems del test. Cuando el test tiene muchos ítems, el coeficiente alfa de Cronbach sobrestima la confiabilidad del instrumento (Tavakol y Dennick, 2011).

El modelo congénico es el más laxo y más general de los modelos usados para estimación de la confiabilidad (Graham, 2006). Este modelo asume que cada ítem individual mide la misma variable latente (i.e. unidimensionalidad), pero con posibles diferentes escalas, con posibles diferentes grados de precisión, y con posibles diferentes grados de error. En base al modelo congénico se plantea el coeficiente omega de McDonald, medido mediante la siguiente ecuación (Ventura-Leon y Caycho-Rodríguez, 2017):

$$\omega = \frac{(\sum_{i=1}^i \lambda_i)^2}{(\sum_{i=1}^i \lambda_i)^2 + (\sum_{i=1}^i 1 - \lambda_i^2)}$$

Considerando que λ_i es la carga factorial del ítem i . El coeficiente omega de McDonald es considerado superior debido a que no depende directamente del número de ítems, ni asume la tau-equivalencia o que la carga factorial de los ítems es equivalente (Ventura-Leon y Caycho-Rodríguez, 2017).

1.2.4 Evidencia de validez basada en la relación con otras variables

La evidencia de validez basada en la relación con otras variables, también llamada evidencia externa de validez, se refiere al grado en el que se relaciona la puntuación de un test con algún criterio que se espera pronostique el test, o con otros test que hipotéticamente midan el mismo constructo, constructo relacionados, o constructos diferentes (American Educational Research Association et al., 2014). Los resultados contribuirían a evaluar el grado en que las relaciones hipotéticas son consistentes con la interpretación propuesta (Elosua, 2003). En ese sentido, se consideran dos tipos de evidencia de validez: la evidencia convergente/discriminante y las relaciones test-criterio.

La evidencia convergente involucra la relación entre las puntuaciones de un test y otras medidas que miden el mismo constructo o un constructo similar, mientras que la evidencia discriminante involucra la relación entre las puntuaciones de un test y medidas que miden un constructo distinto (American Educational Research Association et al., 2014). La lógica detrás de la evidencia convergente y discriminante corresponde al uso de la matriz multirrasgo-multimétodo, en el cual la evidencia convergente corresponde a los valores monorrasgo-heterométrico, en otras palabras, el grado de relación entre distintos procedimientos (e.g. entrevista estructurada y test psicométrico) que miden el mismo constructo (e.g. Depresión),

y por otra parte, la evidencia discriminante corresponde a los valores heterorrasgomonómétodo, en otras palabras, el grado de relación entre distintas variables (e.g. Ansiedad y depresión) empleando el mismo método (e.g. Test psicométrico) (Elosua, 2003).

La relación test-criterio se refiere al grado en el cual las puntuaciones del test se relacionan con algunos criterios de acuerdo a la interpretación de las puntuaciones y la teoría (Elosua, 2003). Se llama evidencia concurrente de validez cuando las puntuaciones de un test se correlacionan con algún criterio al momento de la medición, y se llama evidencia predictiva de validez cuando las puntuaciones de un test predicen algún criterio después del momento de la medición. Por ejemplo, una prueba que evalúe bienestar psicológico debería relacionarse con la presencia o ausencia de trastornos mentales y desordenes psiquiátricos, o una prueba de cociente intelectual usada para selección de personal en el contexto organizacional debería predecir el criterio de desempeño laboral de los empleados contratados

1.2.5 Evidencia de validez basada en las consecuencias de la aplicación del test

La evidencia de validez basada en las consecuencias de la aplicación del test se refiere al grado en el cual las consecuencias deseables e indeseables implícitas del uso del test en un contexto específico son analizadas, y en base a ellas, se efectúa un juicio de valor sobre las consecuencias sociales de la aplicación del instrumento para evaluar la validez de su uso (American Educational Research Association et al., 2014; Lane, 2014; Padilla, Gomez, Hidalgo, y Muñiz, 2006).

Esta fuente de evidencia de validez debería ser investigada siempre que la validez de la interpretación deseada de las mediciones pueda resultar amenazada por evidencias de baja representación de constructo, es decir, si el contenido del instrumento no representa adecuadamente el constructo (Padilla et al., 2006). Además, esta fuente de validez debería ser estudiada cuando el uso de un instrumento está siendo validado dentro del contexto de un sistema o de un marco de práctica institucional (Lane, 2014). Por ejemplo, si se valida un instrumento para tamizaje de ansiedad en pacientes que acuden a atención primaria de salud como parte de la implementación de algoritmos de diagnóstico y tratamiento, la consecuencia implícita es que el número de personas con trastornos ansiosos que no reciben tratamiento debería reducirse, y por lo tanto, evaluar si dicha consecuencia se cumple contribuiría a la evidencia de validez del uso de dicho instrumento.

Para evaluar esta fuente de evidencia se recomienda detallar una teoría de acción que refleje un argumento interpretativo, un argumento de uso y un argumento de validez para los programas de evaluación (American Educational Research Association et al., 2014; Lane, 2014). Uno debe especificar qué es lo que la aplicación del test pretende lograr, qué argumentos uno puede proponer a favor y en contra del uso del instrumento, y qué otras consecuencias tiene el uso del instrumento en el sistema dentro del cual se quiere aplicar (Lane, 2014).

Cabe recalcar que todavía hay discusión y controversia acerca de la naturaleza de esta fuente de validez, cuándo debe ser estudiada, con qué métodos

debe ser evaluada, y si realmente contribuye al argumento de validez del test (Lane, 2014; Padilla et al., 2006).

2. INVESTIGACIONES EN TORNO AL PROBLEMA INVESTIGADO, NACIONALES E INTERNACIONALES

En el Perú, Calderón et al. (2012) realizaron un estudio de validación de la versión peruana del PHQ-9 para el diagnóstico de depresión en pacientes que acuden a atención primaria de salud. Mediante el juicio de un panel de expertos compuesto por veintitrés psiquiatras, tres psicólogos y una enfermera, se realizaron sugerencias y modificaciones que devendrían en la versión peruana del PHQ-9, y se concluyó señalando que el PHQ-9 es un instrumento válido para el diagnóstico de depresión, y que puede ser de fácil uso en el contexto del sistema de atención primaria de salud en el Perú (Calderón et al., 2012).

Asimismo, Gil (2019) realizó un estudio cualitativo del contenido del PHQ-9 mediante grupos focales con gestantes usuarias del primer nivel de atención de salud y obstetras proveedoras de salud. Señalaron que los ejemplos del ítem 7 no eran adecuados aunque igual era entendible, y los ítems 1, 5 y 8 eran difíciles de entender porque presentaban situaciones múltiples y/o opuestas en un mismo ítem, especialmente el ítem 8.

Además, en el Perú, Zhong et al. (2014) estudiaron la validez de la versión española del PHQ-9 empleando 1520 participantes mujeres embarazadas entre 18 y 49 años con menos de 16 semanas de gestación que acuden a las clínicas de cuidado pre-natal del INMP usando un análisis basado en el modelo TRI de Rasch. Mediante el análisis de componentes principales de los residuos se determinó que

los datos se ajustaban a un modelo unidimensional, se identificó independencia local evaluando las correlaciones inter-ítem ($r < .3$), se halló una adecuada consistencia interna ($PSI = 9.72$), y se determinó que colapsar las categorías “más de la mitad de los días” y “casi todos los días” en una sola categoría mejoraba el ajuste del modelo (Zhong et al., 2014). Finalmente, se halló que la invarianza de medida no fue estadísticamente significativa, pues dos ítems presentaron DIF de acuerdo al nivel educativo de los participantes, sin embargo, el tamaño del efecto del DIF fue ínfimo ($R^2 = .0040$; $R^2 = .0085$), al igual que el DIF de un ítem con respecto a la edad ($R^2 = .0079$) (Zhong et al., 2014).

En Chile, Baader et al. (2012) estudiaron la validez y utilidad del PHQ-9 para el diagnóstico de depresión empleando 1327 pacientes usuarios de atención primaria de salud. Mediante un análisis de componentes principales se determinó un modelo unidimensional que explica 47% de la varianza total, y una confiabilidad de consistencia interna aceptable ($\alpha = .835$), una correlación r de Pearson positiva y fuerte ($r > .75$) entre las puntuaciones del PHQ-9 y la escala de depresión de Hamilton (HDS) evidenciando validez convergente, y se encontró una alta superposición de los resultados categóricos del HDS, PHQ-9, y CIE-10 evidenciando una elevada precisión diagnóstica (Baader et al., 2012).

En Argentina, Montenegro et al. (2017) estudiaron la utilidad del PHQ-9 en el diagnóstico de trastorno depresivo en una sala de clínica médica empleando 152 pacientes. Se evidenció una concordancia elevada entre los pacientes diagnosticados con y sin depresión mediante el PHQ-9 y el HADS ($K = .67$; $IC95\% = .56-0.78$) lo cual evidencia validez convergente, una asociación

estadísticamente significativa entre los niveles de dependencia de terceros medido por el Índice de Barthel y la presencia de depresión medida por el PHQ-9, lo cual evidencia validez de criterio concurrente, y se determinó una adecuada precisión diagnóstica (Se=98.7%; Es=70.7%; VPN=98.1%) (Montenegro et al., 2017).

En Colombia, Cassiani-Miranda et al. (2017) se estudió la confiabilidad y dimensionalidad del PHQ-9 para la detección de síntomas de depresión en estudiantes de ciencias de la salud de Cartagena, empleando 550 participantes con una edad media de 20 años (DE = 2.59). Se realizó un AFE obteniendo dos factores que explicaban 42.8% de la varianza total, y se obtuvo una alta confiabilidad mediante el coeficiente alfa de Cronbach ($\alpha=.83$) y omega de McDonald ($\omega=.89$), determinando que la evidencia de validez basada en la estructura interna apoya una interpretación de las puntuaciones del test considerando dos dimensiones, sin embargo, el autor señala que las diferencias entre modelos unidimensionales y bidimensionales se pueden deber a la naturaleza del tipo de población usada (Cassiani-Miranda et al., 2017).

En Ecuador, Fabara (2013) estudió la validez del PHQ-9 y el GHQ-12 como instrumentos para tamizaje de depresión en el ámbito de atención primaria de salud. Empleando 189 participantes determinó que el PHQ-9 presentaba una adecuada precisión considerando un punto de corte ≥ 10 (Se=84.2%; Es=91.2%; LR+=9.53; LR-=.17) (Fabara, 2013).

3. DEFINICIONES CONCEPTUALES Y OPERACIONALES

3.1. Definición conceptual

APA define la depresión como un trastorno mental caracterizado por un estado de ánimo deprimido o una pérdida de interés o placer en casi todas las actividades, cambios de apetito o peso, del sueño y de la actividad psicomotora, falta de energía, sentimientos de infravaloración o culpa; dificultad para pensar, concentrarse o tomar decisiones, y pensamientos recurrentes de muerte o ideación, planes o intentos suicidas, acompañado de un malestar clínico significativo o de deterioro social, laboral o de otras áreas importantes de la actividad del individuo (American Psychiatric Association, 1995).

3.2. Definición operacional

La depresión se mide a través de la versión peruana del Cuestionario modificado sobre la Salud del Paciente (PHQ-9). Existen tres formas de definir la depresión.

En primer lugar, de acuerdo a la definición directa, la presencia de depresión se define por un puntaje total por encima de un punto de corte.

Segundo, se puede aplicar el instrumento mediante un algoritmo de dos pasos. Primero, se aplican los ítems 1 y 2. Si puntúa menos de dos en ambos ítems, entonces ya no se continúa con el resto del instrumento. Y si puntúa dos o más en el ítem 1 o el ítem 2, se continúa con el resto del instrumento y se suman los puntajes de todos los ítems. La depresión se define como un puntaje ≥ 2 en el ítem 1 o el

ítem 2 y un puntaje total por encima de un punto de corte. Para mayor claridad, este método se llamará PHQ-ALG en el resto de la investigación.

Tercero, al algoritmo diagnóstico, el Desorden Depresivo Mayor (DD) se define por una puntuación de dos o más (por lo menos “más de la mitad de los días”) en cinco o más ítems de los cuales por lo menos uno de los ítems es el 1 o el 2 (“Poco interés en hacer las cosas” o “Sentirse triste, desanimado, deprimido o sin esperanza”), y Otros Síndromes Depresivos (ODD) se define por una puntuación de dos o más en dos a cuatro ítems, y por lo menos uno de los ítems es el 1 o el 2.

Además, la depresión también se medirá usando la Escala de Ansiedad y Depresión Hospitalaria (HADS). Los 7 ítems impares miden la dimensión de Ansiedad, y los 7 ítems pares miden la dimensión de Depresión. La ansiedad y depresión se definen como puntuaciones mayores de 8 en sus respectivas subescalas.

Tabla 2. Operacionalización de la variable de estudio

Variable	Instrumento	Dimensiones	ítems	Escala de medición
Depresión	PHQ-9	Depresión	1, 2, 3, 4, 5, 6, 7, 8, 9	Cuantitativa de intervalo
	HADS	Ansiedad	1, 3, 5, 7, 9, 11, 13	Cuantitativa de intervalo
		Depresión	2, 4, 6, 8, 10, 12, 14	Cuantitativa de intervalo

CAPÍTULO III: METODOLOGÍA

1. TIPO DE INVESTIGACIÓN

La presente investigación es de tipo empírico con metodología cuantitativa e instrumental porque el estudio presenta datos empíricos originales producidos por el investigador y analiza las propiedades psicométricas de un instrumento de medida psicológico (Montero y León, 2007).

2. DISEÑO DE LA INVESTIGACIÓN

El estudio es una investigación de diseño no experimental retrospectivo transversal teniendo como objetivo evaluar la evidencia de validez de las puntuaciones de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9) para tamizaje de depresión en una población de pacientes oncológicos adultos de un Hospital de atención especializada de Lima metropolitana (Veiga de Cabo, Fuente-Díez, y Zimmermann-Verdejo, 2008).

Es una investigación no experimental ya que no se manipulan variables; retrospectiva porque el momento del estudio se establece una vez se ha hecho presente la depresión; y transversal porque la investigación fue realizada en un momento específico sin temporalidad base.

3. NATURALEZA DE LA MUESTRA

3.1. Participantes

En el presente estudio, los participantes pertenecieron a uno de dos grupos:

- Para validación de contenido, psicólogos expertos en pacientes oncológicos miembros del Equipo Funcional de Salud Mental Oncológica del Departamento de Especialidades Médicas del INEN
- Para el resto de los análisis, pacientes adultos con cáncer del INEN

3.2. Método de selección de los participantes

El muestreo es no probabilístico multietápico. Es multietápico porque se dio en dos etapas (i.e. primero los jueces y luego los pacientes), y no probabilístico dado que la selección de los jueces expertos fue intencional, seleccionando a todos aquellos que se tenían al alcance y contaban con la disponibilidad y disposición de participar; y dado que la selección de los participantes fue accidental y se realizó según la disponibilidad de los pacientes que acudían al servicio de quimioterapia y pacientes hospitalizados de las áreas de Cirugía, Adolescentes y UTM (Hernández, Fernández, y Baptista, 2010).

En cuanto al número participantes, los estudios instrumentales del PHQ-9 realizados en otros países latinoamericanos han incluido desde 189 hasta 1327 participantes (Baader et al., 2012; Cassiani-Miranda et al., 2017; Fabara, 2013; Montenegro et al., 2017). El último estudio instrumental en el INEN del cual se tiene conocimiento empleó 500 participantes de los cuales 25 (5%) no completaron correctamente sus cuestionarios (Vilela-Estrada, 2019). Por otra parte, la literatura recomienda seguir el criterio de por lo menos 20 participantes por cada ítem para un adecuado análisis factorial asumiendo que la data no es ideal (e.g. bajas comunalidades, carga factoriales cruzadas, etc.), lo cual se traduciría en esta investigación en un mínimo de 180 participantes; sin embargo, en la literatura

siempre se recomienda procurar tener la mayor cantidad de participantes posibles para este tipo de análisis (Costello y Osborne, 2005; Matsunaga, 2010; Treiblmaier y Filzmoser, 2010).

Por lo tanto, para la validación de contenido, se optó por incluir a los 8 psicólogos del Equipo Funcional de Salud Mental Oncológica del Departamento de Especialidades Médicas del INEN; y 500 pacientes adultos con cáncer del INEN para el resto del estudio.

3.3. Criterios de inclusión

- Pacientes adultos con cáncer que cuenten con historia clínica
- Pacientes mayores de edad.
- Pacientes hospitalizados en UTM, Adolescentes y Cirugía.
- Pacientes ambulatorios de Quimioterapia.

3.4. Criterios de exclusión

- Pacientes con diagnóstico de delirium, demencia, o trastornos neurológicos.
- Pacientes con alteraciones o deterioro cognitivo que impidiesen el desarrollo adecuado del cuestionario.
- Pacientes que se encuentran actualmente bajo tratamiento psicofarmacológico para depresión.
- Pacientes con alteraciones funcionales que dificulten la comprensión del instrumento.
- Pacientes que no comprenden el idioma español.

3.5. Criterios de eliminación

- No completar el PHQ-9.
- Marcar más de dos opciones en cualquier ítem del PHQ-9.
- Aquellos que tengan datos incompleto según cada uno de los procesos estadísticos.

En la Tabla 3 se puede apreciar un resumen de las características de los participantes incluidos en esta investigación. De los 500 participantes iniciales, 49 fueron eliminados por no completar correctamente el PHQ-9. Por lo tanto, se emplearon 451 participantes.

Tabla 3. Características de los participantes

		Frecuencia	%
PHQ-9 completo	Validos	451	100%
PHQ-9 y HADS completos	Validos	438	97.1%
	Incompleto	13	2.9%
Sexo	Hombre	218	48.3%
	Mujer	233	51.7%
Edad	18 a 35 años	186	41.2%
	36 a 50 años	116	25.7%
	51 a 65 años	118	26.2%
	66 años a más	29	6.4%
	Incompleto	2	0.4%
Lugar de aplicación	Hospitalización	256	56.8%
	Quimioterapia	190	42.1%
	Incompleto	5	1.1%
Estado Civil	Casado/Conviviente	222	49.2%
	Divorciado/Separado	44	9.8%
	Soltero	166	36.8%
	Viudo	18	4%
	Incompleto	1	0.2%
Situación laboral	Desempleado	309	68.5%
	Empleado	38	8.4%
	Independiente/Ama del hogar	103	22.8%
	Incompleto	1	0.2%

Grado de instrucción	Sin estudios	14	3.1%
	Primaria completa	61	13.5%
	Secundaria completa	215	47.7%
	Técnica completa	55	12.2%
	Superior completa	105	23%
	Incompleto	1	0.2%
Fase de Tratamiento	Pre-tratamiento	78	17.3%
	En tratamiento	296	65.6%
	Post-tratamiento	60	13.3%
	Incompleto	17	3.8%

Adicionalmente, algunos procedimientos estadísticos requirieron que, además del PHQ-9, otros datos estén debidamente completos. Por lo tanto, en estos casos se excluyeron aquellos que tenían datos incompletos. Los participantes usados para cada procedimiento estadístico serán los siguientes:

- Para el análisis de la dimensionalidad, se emplearon 451 participantes.
- Para análisis de la confiabilidad, se emplearon 451 participantes.
- Para el análisis de la invarianza de medida, se emplearon 381 a 451 participantes dependiendo de la variable de agrupación.
- Para el análisis de la correlación del PHQ-9 otras variables, se emplearon entre 145 a 429 participantes dependiendo de la variable.
- Para identificar el punto de corte del PHQ-9, PHQ-ALG y PHQ-2, y comparar sus poderes diagnóstico, se emplearon 438 participantes.

4. INSTRUMENTOS

En la presente investigación se emplearon dos instrumentos. El primero es la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9). El segundo es la Escala de Ansiedad y Depresión Hospitalaria (HADS).

4.1. Cuestionario sobre la Salud del Paciente (PHQ-9)

El Cuestionario sobre la Salud del Paciente (PHQ-9) fue originalmente desarrollado por Spitzer, Kroenke, and Williams (1999). Es una escala clínica que mide la depresión. Puede ser aplicado individual o colectivamente, a adultos mayores de 17 años, con un tiempo de aplicación aproximado de 5 minutos. Los materiales necesarios son el cuestionario, lápiz y papel. La última adaptación al contexto de aplicación de esta investigación, es decir, población hospitalaria, fue realizada por Calderón et al. (2012).

El PHQ-9 es un instrumento unidimensional que evalúa la depresión de acuerdo a los criterios del DSM-IV mediante 9 ítem con formato de respuesta Likert del 0 (“nunca”) al 3 (“casi todos los días”) que indica la severidad de la sintomatología depresiva, y provee una puntuación total cuantitativa de intervalo del 0 hasta 27 (Kroenke, Spitzer, y Williams, 2001; Kroenke y Spitzer, 2002).

Kroenke and Spitzer (2002) estudiaron la evidencia de validez del PHQ-9 empleando 6000 pacientes de ocho clínicas de atención primaria y siete clínicas gineco-obstétricas, hallando una elevada confiabilidad de consistencia interna ($\alpha_1=.89$; $\alpha_2=.86$) y de test-retest ($r=.84$), adecuada precisión diagnóstica empleando el punto de corte ≥ 10 ($Se=88\%$; $Es=88\%$; $LR=7.1$; $auROC=.95$), y una relación lineal directa entre las puntuaciones del PHQ-9 y las sub-escalas del Cuestionario de salud versión corta (SF-20) ($r=.33-.73$), días de discapacidad ($r=.39$), uso de servicios de salud ($r=.24$), y percepción de dificultad para realizar actividades y mantener relaciones debido a los síntomas ($r=.55$) (Kroenke y Spitzer, 2002).

Múltiples investigaciones coinciden en identificar una elevada sensibilidad y adecuada especificidad, resaltando un alto valor predictivo negativo, por lo que se recomienda su uso como instrumento de tamizaje de depresión en población hospitalaria (Inoue et al., 2012; Montenegro et al., 2017; Suzuki, Kumei, Ohhira, Nozu, y Okumura, 2015; Thombs, Ziegelstein, y Whooley, 2008).

Calderón et al. (2012) realizaron la adaptación del contenido del test a la realidad peruana mediante una reunión técnica de un panel de 27 especialistas en salud mental, y se concluyó que la versión peruana del PHQ-9 es un instrumento válido para diagnosticar de depresión en población peruana que acude a atención primaria de salud. Y en el contexto de pacientes mujeres embarazadas peruanas que acuden centros de cuidado pre-natal, se evidenció unidimensionalidad, invarianza de medida, independencia local y una adecuada consistencia interna si se colapsan las dos últimas categorías de respuesta (Zhong et al., 2014).

4.2. Escala de Ansiedad y Depresión Hospitalaria (HADS)

La Escala de Ansiedad y Depresión Hospitalaria (HADS) fue originalmente desarrollado por Zigmond y Snaith (1983). Es una escala clínica compuesta por dos sub-escalas, una que mide depresión y otra que mide ansiedad. Puede ser aplicado individual o colectivamente, a adultos mayores de 18 años, con un tiempo de aplicación aproximado de 5 minutos. Los materiales necesarios son el cuestionario, lápiz y papel. La última adaptación al contexto de aplicación de esta investigación fue realizada en el Instituto Nacional de Enfermedades Neoplásicas por Vilela-Estrada (2019).

La Escalada de ansiedad y depresión hospitalaria (HADS) es un instrumento de 14 ítems que forman dos sub-escalas, una de ansiedad y otra de depresión, de 7 ítems cada uno. Los ítems se puntúan en escala Likert del 0 al 3, con un rango de puntuación de cada sub escala del 0 al 21 (Zigmond y Snaith, 1983). Una mayor puntuación refleja mayor severidad de ansiedad y depresión, y además, un punto de corte de 8 puntos para indicar niveles clínicamente significativos de ansiedad y depresión.

Se indagó la evidencia de validez del HADS para medir depresión y ansiedad en pacientes oncológicos peruanos basándose en el contenido del test mediante la evaluación de jueces expertos en psicooncología ($V=.83-1$), un adecuado ajuste al modelo bifactorial (CIM/DF<2; AGFI=.929; CFI=.935; RMSEA=.053; ITL=.923; IFI=.936), saturaciones moderadas y elevadas (.41-.71), una adecuada fiabilidad ($\omega_1=.77$; $\omega_2=.75$), y correlaciones estimadas fuertes entre el HADS y el BDI (.89) y el BAI (.84) (Vilela-Estrada, 2019).

5. PROCEDIMIENTO

Primeramente, a los jueces expertos se les otorgó el protocolo de validación, los cuales fueron devueltos completados luego de una semana.

Luego, se armaron los 500 cuestionarios para los participantes, que incluían el consentimiento informado, la ficha sociodemográfica, el PHQ-9, y el HADS. Se rotó el lugar de aplicación (e.g. Salas de Quimioterapia ambulatoria, pisos de Cirugía, servicio de Adolescentes, y sala de UTM) según día de la semana. En cada espacio hay una lista con los nombres e historias clínicas de los pacientes. Se siguió

ese orden para visitar a los pacientes. Si se encontraban despiertos y en condiciones de responder el protocolo de cuestionario, se les preguntaba si desean participar en la investigación usando la hoja informativa, y a aquellos que aceptaron voluntariamente se les entregó el cuestionario para llenar de manera física. El llenado del cuestionario sociodemográfico se hizo en conjunto con el investigador, y luego, al participante se le dieron los cuestionarios con una tabla y un lapicero para que llene de manera individual; a aquellos que tuvieron dificultades para llenarlo por sí mismos, se les brindó asistencia. Al terminar, a todos los participantes se les brindó información del servicio de psicología del INEN, y se les extendió una invitación a sus charlas y talleres.

6. PLAN DE ANÁLISIS DE DATOS

En primer lugar, se emplearía Google Sheets para ordenar las calificaciones de los jueces expertos de los protocolos de validación, y para evaluar la evidencia de validez basada en el contenido estimando el coeficiente V de Aiken y sus intervalos de confianza IC (95%). Luego, se estimaría la confiabilidad inter-jueces de las puntuaciones empleando el coeficiente de correlación intra-clase bidireccional de efectos fijos, de múltiples medidas y acuerdo absoluto ICC(3,k).

Luego, se emplearía Google Sheets para organizar la base de datos de los participantes. Se removerían del análisis estadístico aquellos casos con datos incompletos en sus cuestionarios PHQ-9, según el criterio de eliminación. Después, se usaría JASP para el posterior tratamiento estadístico.

Se evaluaría la normalidad multivariada del PHQ-9 mediante los coeficientes de Mardia de curtosis y asimetría multivariada.

Después, se analizaría la evidencia de validez de estructura interna. Primero se realizaría un AFE empleando el método de extracción de máxima verosimilitud (ML) si los datos son normales, o sino, factorización de ejes principales (FEP), con una rotación oblicua promax ($Kappa=4$). Además, se evaluaría la adecuación de los datos mediante el test KMO ($KMO>.8$) y el test de esfericidad de Bartlett ($p<.05$). Se determinaría el número óptimo de factores considerando el gráfico de sedimentación de los autovalores.

Seguidamente, se realizaría el AFC evaluando el ajuste al modelo unidimensional y otros modelos dependiendo de los hallazgos del AFE y otros modelos teóricos. Se emplearía el método de extracción de máxima verosimilitud (ML) y se determinaría la bondad de ajuste a través de los indicadores descritos en el marco teórico.

Para evaluar la confiabilidad del instrumento, se emplearía JASP para calcular el coeficiente alfa de Cronbach, y se emplearía Google Sheets para calcular el coeficiente omega de McDonald empleando las cargas factoriales de los ítems con sus factores.

Se emplearía JASP para evaluar la invarianza de medida entre sub-grupos en función al sexo, edad, estado civil, lugar de aplicación, grado de instrucción y situación laboral, empleando un enfoque progresivo de AFCMG para evaluar si estas categorías sesgan los puntajes del instrumento.

Primero, se evaluaría la invarianza configural, comparando el modelo base identificado previamente mediante AFC contra el mismo modelo considerando la variable de agrupación. Si la bondad de ajuste no se reduce significativamente ($\Delta CFI \leq 0.01$), se aceptaría que el modelo cuenta con invarianza configural, y se continuaría con el siguiente paso.

El segundo paso sería evaluar la invarianza métrica, comparando el modelo agrupado creado en el paso anterior contra el mismo modelo, pero ahora restringiendo las cargas factoriales para que sean iguales entre los grupos. Similarmente, se evaluaría si la bondad de ajuste empeora significativamente, y si no, se aceptaría que el modelo cuenta con invarianza métrica, y se continuaría con el siguiente paso.

El tercer paso sería evaluar la invarianza escalar, comparando el modelo creado en el paso anterior contra el mismo modelo, pero ahora adicionalmente restringiendo los interceptos para que sean iguales entre los grupos. Igualmente, se evaluaría si la bondad de ajuste empeora, y si no, se aceptaría que el modelo cuenta con invarianza escalar, y se continuaría con el siguiente paso.

Finalmente, se evaluaría la invarianza estricta, comparando el modelo creado en el paso anterior contra el mismo modelo, pero ahora adicionalmente restringiendo los valores y las medias para que sean iguales entre los grupos. Igualmente, se evaluaría como cambia la bondad de ajuste, y si no empeorase significativamente, se diría que el instrumento es invariante.

Luego, para determinar la evidencia de validez basada en la relación con otras variables, se usaría el coeficiente r de correlación lineal de Pearson para

determinar la correlación entre la puntuación total del PHQ-9 y del HADS, y se compararían las diferencias de la puntuación del PHQ-9 entre grupos usando el sexo, la edad, el tipo de patología, y fase de tratamiento empleando t de Student y d de Cohen cuando se comparan dos grupos, y ANOVA y η^2 eta-cuadrado cuando se comparan tres o más.

Después, usando como test diagnóstico el HADS, se identificaría la precisión diagnóstica de los diferentes puntos de corte para determinar el más apropiado para el método directo, para el algoritmo de calificación (PHQ-ALG), y para la versión abreviada (PHQ-2). Se emplearía el PPV (Valor predictivo positivo), NPV (Valor predictivo Negativo), ACC (Precisión), TPR (Sensibilidad), TNR (Especificidad), FRN (Proporción de falsos negativos), FPR (Proporción de falsos positivos), LR+ (Razón de verosimilitud positiva), LR- (Razón de verosimilitud negativa), DOR (Diagnostic Odds Ratio), NNM (Numero necesitado para diagnóstico errado), NNMw (Numero necesitado para diagnóstico errado ponderado), d de la diagonal, J de Youden, y diagonal máxima.

7. CONSIDERACIONES ÉTICAS

7.1. Autonomía

Este principio reconoce que las personas, en calidad de su dignidad, merecen respeto de su voluntad. En el presente estudio, se brindó información clara y precisa acerca de los objetivos y procedimientos de la evaluación, y la información que sería usada, respetando la decisión del paciente en el caso de no desear participar.

7.2. Beneficencia

Este principio dicta que el propósito principal de la indagación científica es velar por el bienestar de los participantes, y que del estudio y sus resultados se obtengan beneficios. En el presente estudio, este principio se satisface a través de los resultados de las pruebas que favorecen la identificación oportuna de pacientes con depresión, ofreciendo tamizaje de depresión y ansiedad. Además, se brindó acompañamiento en hospitalización y consejería a todos los pacientes que lo solicitaron, independientemente de su aceptación participación en la investigación. Asimismo, todos los participantes fueron invitados a las charlas psicoeducativas y talleres brindados por el servicio de psicología, y se garantizó su acceso a un espacio para orientación, consejería y acompañamiento psicológico. Finalmente, aquellos que presentaron niveles clínicamente significativos de malestar emocional fueron atendidos por el personal del servicio de psicología e invitados a consulta externa.

7.3. No maleficiencia

Este principio dicta que uno no debe hacer daño de manera intencional, ni perjudicar de manera innecesaria al paciente. En el presente estudio, este principio se satisface garantizando que las preguntas de las pruebas no dañan. La evidencia recopilada en las Guías de práctica clínica indica que preguntar acerca de síntomas depresivos o ideación suicida no es nocivo para los pacientes y es más bien práctica recomendada si es que la institución tiene un servicio de salud mental. Sin embargo, se decidió no aplicar dichos instrumentos a pacientes en el servicio de Emergencias ni Cuidados Paliativos debido a su probable vulnerabilidad psicológica, y se

garantizó que aquellos pacientes que requirieron atención psicológica la recibieron de manera oportuna.

7.4. Justicia

Este principio dicta se debe ser imparcial y evitar discriminación en tanto a los riesgos y beneficios de la investigación. Este principio se satisface garantizando que todos aquellos que cumplieron los criterios de selección de participantes tuvieron la misma probabilidad de ser invitados al estudio. Además, todos los participantes, sea que acepten o no, fueron invitados a las charlas y presentaciones psicoeducativas del servicio de psicología. Y todos aquellos pacientes en los cuales se identificó un elevado malestar emocional tuvieron la misma oportunidad de acceder a un espacio para orientación, consejería y acompañamiento psicológico.

CAPÍTULO IV: RESULTADOS

En cuanto al primer objetivo específico, se evaluó la evidencia de validez basada en el contenido mediante el criterio de los 8 jueces expertos empleando el coeficiente V de Aiken, identificando sus intervalos de confianza, y midiendo la confiabilidad inter-jueces mediante el $ICC_{(3,k)}$.

Tabla 4. Criterio de Jueces del PHQ-9

	Ítems	V de Aiken	IC _(95%) de V de Aiken		ICC (3,k)
			Límite inferior	Límite superior	
Representación	1	1	.922	1	.18
	2	1	.922	1	
	3	.938	.828	.979	
	4	.938	.828	.979	
	5	.938	.828	.979	
	6	.938	.828	.979	
	7	.906	.787	.962	
	8	.875	.749	.943	
	9	1	.922	1	
Relevancia	1	1	.922	1	.35
	2	1	.922	1	
	3	.969	.871	.993	
	4	.938	.828	.979	
	5	.875	.749	.943	
	6	.938	.828	.979	
	7	.875	.749	.943	
	8	.938	.828	.979	
	9	1	.922	1	
Claridad	1	.969	.871	.993	.46
	2	1	.922	1	
	3	.969	.871	.993	
	4	1	.922	1	
	5	.875	.749	.943	
	6	.938	.828	.979	
	7	.969	.871	.993	
	8	.844	.712	.922	
	9	1	.922	1	

N=8

En la Tabla 4, se aprecia que, según el criterio de los ocho jueces expertos, todos los ítems presentaron contenido apropiado para el uso propuesto ($V_{Inf} > .7$). Además, las puntuaciones de la Claridad, Representación y Relevancia de los ítems evidencian una confiabilidad inter-jueces pobre ($ICC_{(3,k)} < .5$). Se debe de considerar que en todas las categorías todos los ítems fueron puntuados con 3 o más.

En cuanto al segundo objetivo específico, se evaluó la dimensionalidad de la versión peruana del Cuestionario sobre la Salud del Pacientes (PHQ-9) mediante un análisis factorial exploratorio (AFE) y análisis factorial confirmatorio (AFC), para lo cual se emplearon 451 casos con respuestas al PHQ-9 completas.

Tabla 5. Medidas de asimetría y curtosis multivariada del PHQ-9

	curtosis	z	p	asimetría	$\chi^2(gl)$	p
Multivariado	58.63	37.09	<.0001			
Coefficiente de Mardia	147.47	36.58	<.001	28.662	2154.424 (165)	<.001

N=451

En la Tabla 5 se presentan las medidas de normalidad multivariada del PHQ-9 mediante el coeficiente de Mardia y valor crítico. Se aprecia que las tres medidas indican que estadísticamente no se puede afirmar que los datos presentan una buena bondad de ajuste a una distribución normal multivariada ($p < .001$).

Además, previo a realizar el AFE, se evaluó la adecuación de la matriz de correlación de los ítems del PHQ-9 para realizar análisis factorial mediante el índice Kaiser-Meyer-Olkin (KMO) y el test de esfericidad de Bartlett.

Los resultados en la Tabla 6 indican que la matriz de correlación de los ítems del PHQ-9 es ideal para realizar análisis factorial. El índice KMO sugiere que la proporción de varianza explicada por correlaciones parciales es reducida y por lo

tanto la matriz es adecuada pues probablemente refleje una estructura simple (KMO=.827), y el test de esfericidad de Bartlett indica que las correlaciones entre los ítems del PHQ-9 son estadísticamente significativas ($p < .0001$).

Tabla 6. Indicadores de adecuación y esfericidad de la matriz de correlación del PHQ-9

		Valor
Índice de adecuación muestral	KMO	0.827
Test de esfericidad de Bartlett	χ^2	512.74
	gl	36.00
	p	<.0001

N=451

Luego, dado que no se cumple el supuesto de normalidad multivariada, se realizó un AFE empleando el método de extracción de factorización de ejes principales (FEP), y se empleó una rotación promax.

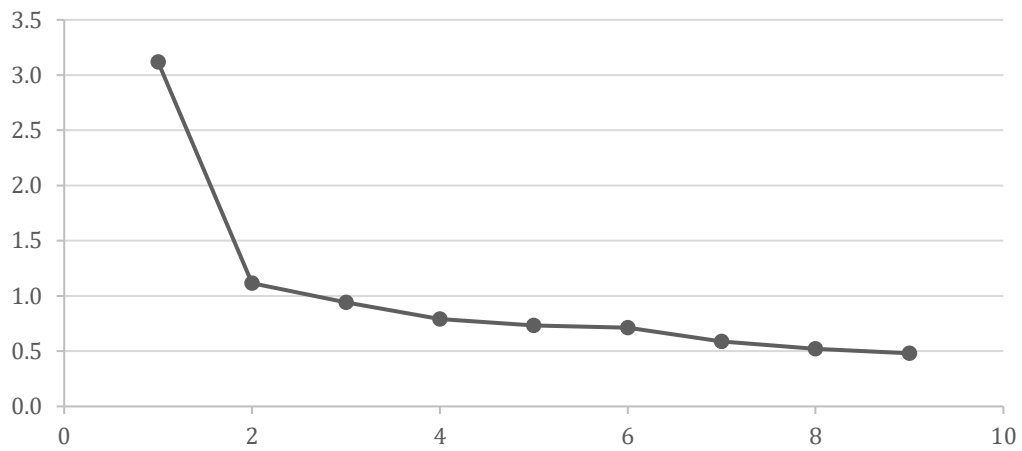
Tabla 7. Varianza explicada del PHQ-9 en pacientes oncológicos

Factor	Autovalor	% de varianza	Varianza acumulada
1	3.117	34.638	34.638
2	1.116	12.403	47.040
3	0.941	10.455	57.495
4	0.792	8.798	66.293
5	0.733	8.145	74.438
6	0.711	7.904	82.342
7	0.588	6.529	88.871
8	0.521	5.786	94.658
9	0.481	5.342	100.000

N=451

En la Tabla 7, se puede apreciar que el modelo unidimensional y el modelo bidimensional contarían con factores con autovalores superiores a 1, mientras que el resto de los factores presentan autovalores inferiores a 1. Sin embargo, cabe resaltar que un tercer factor se aproximaría a 1 ($F3=0.941$). Asimismo, un modelo unidimensional explica 34.6% de la varianza, y un modelo de dos factores explicaría 47% de la varianza.

Gráfico 1. Gráfico de sedimentación



En el Gráfico 1, se pueden apreciar los autovalores de los factores. Se aprecia que cualitativamente, el punto de inflexión de la curva se ubica en el segundo factor; y cuantitativamente, se reportan autovalores superiores a 1 hasta el segundo factor.

En la Tabla 8 se pueden apreciar las comunalidades de los ítems del PHQ-9 en un modelo de un factor (Modelo 1) y de dos factores (Modelo 2).

Tabla 8. Comunalidades de los ítems del PHQ-9 según AFE

	Comunalidades	
	Modelo 1	Modelo 2
Ítem 1: Anhedonia	.408	.395
Ítem 2: Distimia	.449	.466
Ítem 3: Sueño	.186	.287
Ítem 4: Cansancio	.27	.348
Ítem 5: Apetito	.154	.228
Ítem 6: Culpa	.377	.533
Ítem 7: Desconcentración	.17	.166
Ítem 8: Psicomotricidad	.212	.228
Ítem 9: Ideación suicida	.198	.282

Nota: Método de extracción Factorización de ejes principales (FEP).
N=451

Las comunalidades indican el grado en el cual cada ítem se correlaciona con el resto de los ítems. Se observa que los ítems 3, 4, 5, 7, 8 y 9 presentan comunalidades bajas (<.30). Por lo tanto, existe la posibilidad de que estos ítems no carguen significativamente con alguno de los factores y que hayan otros factores que deben de ser considerados dentro del modelo.

Tabla 9. Cargas factoriales del PHQ-9 obtenidas mediante AFE e índices de bondad de ajuste

	Modelo 1	Modelo 2	
	Factor 1	Factor 1	Factor 2
Ítem 1: Anhedonia	.639	.365	.324
Ítem 2: Distimia	.67	.576	-
Ítem 3: Sueño	.432	-	.589
Ítem 4: Cansancio	.52	-	.58
Ítem 5: Apetito	.392	-	.522
Ítem 6: Culpa	.614	.796	-
Ítem 7: Desconcentración	.412	.273	-
Ítem 8: Psicomotricidad	.461	-	.393
Ítem 9: Ideación suicida	.445	.602	-
χ^2	95.296	39.492	
gl	27	19	
p	< .001	.004	
RMSEA	.076	.05	
IC (90%)	0.059 - 0.092	0.027 - 0.071	
TLI	.867	.943	

Nota: Método de extracción Factorización de ejes principales (FEP), con rotación promax con normalización Kaiser. Se obviaron las cargas factoriales insignificantes ($\lambda < 0.2$)
N=451

En la Tabla 9 se pueden apreciar las cargas factoriales del modelo de dos factores (Modelo 2) y de un factor (Modelo 1). En el Modelo 1, el ítem 1 ($\lambda=.639$), el ítem 2 ($\lambda=.67$) y el ítem 6 ($\lambda=.614$) presentan cargas factoriales elevadas, mientras que el ítem 5 ($\lambda=.392$) evidencia una carga factorial disminuida. Además, la prueba estadística exacta chi-cuadrado χ^2 indicaría que los datos no se ajustan al modelo ($\chi^2=95.296$, $gl=27$, $p<.001$), la discrepancia promedio entre la matriz de covarianzas

hipotética y observada refleja un ajuste moderado (RMSEA=.076), y el índice de ajuste relativo TLI refleja un ajuste mediocre (TLI=.867).

En el Modelo 2, el primer factor explica los ítems 1 (Distimia), ítem 2 (Anhedonía), ítem 6 (Culpa), ítem 7 (Desconcentración), e ítem 9 (Ideación suicida), siendo el ítem 6 ($\lambda=.796$) y el ítem 9 ($\lambda=.602$) aquellos con carga factorial más elevadas, y el ítem 1 ($\lambda=.365$) y el ítem 7 ($\lambda=.273$) aquellos con cargas factoriales más reducidas. Un segundo factor explica los ítems 1 (Distimia), ítem 3 (Sueño), ítem 4 (Cansancio), ítem 5 (Apetito), e ítem 8 (Psicomotricidad), siendo el ítem 1 ($\lambda=.324$) y el ítem 8 ($\lambda=.393$) los ítems con carga factorial mediocre y ningún ítem evidencia una carga factorial elevada ($\lambda>.6$). Además, la prueba estadística exacta chi-cuadrado χ^2 evidencian que los datos no se ajustan al modelo ($\chi^2=39.492$, $gl=19$, $p=.004$), la discrepancia promedio entre la matriz hipotética y observada de covarianzas refleja un ajuste excelente (RMSEA=.05), y se no se rechaza la hipótesis nula de ajuste-cerrado ni la hipótesis nula de ajuste-pobre (IC_{90%} de RMSEA=.027-.071), y el índice de ajuste relativo TLI refleja un ajuste adecuado (TLI=.943).

En base a los dos modelos del AFE, se propusieron cuatro modelos para el AFC, de tal forma de poder confirmar la dimensionalidad del PHQ-9. Se propuso el Modelo 1 (M1), que es unifactorial, y tres modelos equivalentes al Modelo 2, que es bidimensional.

El Modelo 2a (M2a), que consta de dos factores correlacionados de primer orden, con el primer factor explicando la varianza de los ítems 1, 2, 6, 7 y 9, y el otro factor explicando los ítems del 3, 4, 5 y 8.

El Modelo 2b (M2b), que es un modelo jerárquico con dos factores de primer orden que tienen la misma estructura que el M2a, y un factor de segundo orden.

Y el Modelo 2c (M2c), que consta de dos factores de primer orden no correlacionados con el primer factor explicando la varianza de todos los ítems y el segundo factor explicando la varianza de los ítems 3, 4, 5 y 8.

Gráfico 2. Modelo 1 (M1)

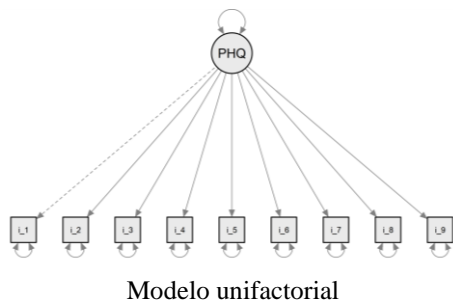


Gráfico 3. Modelo 2a (M2a)

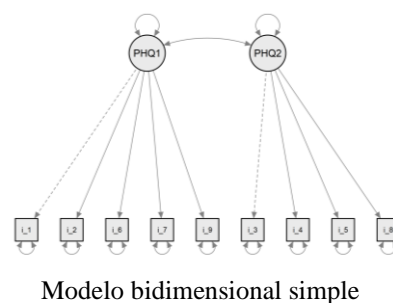


Gráfico 4. Modelo 2b (M2b)

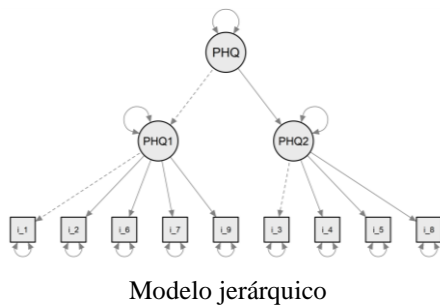


Gráfico 5. Modelo 2c (M2c)

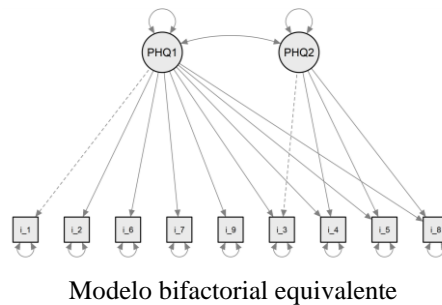


Tabla 10. Índices de bondad de ajuste de los diferentes modelos según AFC

Modelos	χ^2	gl	p	CFI	TLI	AGFI	SRMR	RMSEA A	RMSEA IC(90%)
M1	95.843	27	< 0.001	.901	.867	.919	.053	.075	.059 - .092
M2a	65.994	26	< 0.001	.942	.92	.946	.044	.058	.041 - .076
M2b	65.994	25	< 0.001	.941	.915	.944	.044	.06	.043 - .078
M2c	60.834	22	< 0.001	.944	.908	.94	.041	.063	.044 - .082

Nota: χ^2 : Chi-cuadrado, gl: Grados de libertad, CFI: índice de ajuste comparativo, TLI: índice Tucker-Lewis, AGFI: índice de bondad de ajuste ajustado, SRMR: Raíz Cuadrada del Residuo Cuadrático Medio Estandarizado, RMSEA: Raíz Cuadrada del Error Cuadrático Medio de la Aproximación, RMSEA IC(90%): Intervalo de confianza de 90% de la RMSEA
N=451

En la Tabla 10 se pueden apreciar los índices de bondad de ajuste de los modelos propuestos obtenidos empleado AFC mediante SEM con método de extracción de máxima verosimilitud para poder evaluar la bondad de ajuste del modelo a los datos.

En todos los modelos, la prueba estadística exacta chi-cuadrado χ^2 indica que el modelo no se ajusta a los datos ($p < .001$). Similarmente, en el Modelo 1, los indicadores de ajuste relativo reflejan que el modelo es inadecuado (CFI=.901; TLI=.867), pero los indicadores de ajuste absoluto evidencian un ajuste moderado (AGFI=.919; SRMR=.053; RMSEA=.075). Por otra parte, salvo por el indicador TLI, el Modelo 2a evidencia un ajuste moderado (CFI=.942; TLI=.92; AGFI=.946; SRMR=.044; RMSEA=.058), al igual que el Modelo 2b (CFI=.941; TLI=.92; AGFI=.944; SRMR=.044; RMSEA=.06) y el Modelo 2c (CFI=.944; TLI=.908; AGFI=.94; SRMR=.041; RMSEA=.063). Se aprecia que los modelos de dos factores se ajustan más a los datos que el modelo unifactorial. Además, entre los modelos de dos factores, el Modelo 2a y Modelo 2b son prácticamente idénticos, y presentan el mayor que el Modelo 2c.

Tabla 11. Cargas factoriales del PHQ-9 obtenidas mediante AFC

	Modelo 1	Modelo 2a		Modelo 2b		Modelo 2c	
		F1	F2	F1	F2	F1	F2
Ítem 1: Anhedonia	0.643	0.636		0.636		0.638	
Ítem 2: Distimia	0.678	0.698		0.698		0.697	
Ítem 3: Sueño	0.419		0.517		0.517	0.336	0.464
Ítem 4: Cansancio	0.516		0.595		0.595	0.452	0.339
Ítem 5: Apetito	0.375		0.449		0.449	0.303	0.347
Ítem 6: Culpa	0.630	0.666		0.666		0.665	
Ítem 7: Desconcentración	0.404	0.400		0.400		0.401	
Ítem 8: Psicomotricidad	0.444		0.488		0.488	0.394	0.242
Ítem 9: Ideación suicida	0.457	0.481		0.481		0.482	

N=451

En la Tabla 11 se pueden apreciar las cargas factoriales de los ítems con sus respectivas dimensiones. Se aprecia que las cargas factoriales del Modelo 2a y del Modelo 2b son idénticas y todas significativas ($\lambda > 0.40$). Además, se observa que los ítems 1, 2 y 6 presentan cargas factoriales moderadas en todos los modelos de manera consistente. Asimismo, salvo con excepción del ítem 8 en el Modelo 2c, todas cargas factoriales son superiores a 0.30; pero por otra parte, también son inferiores a 0.70.

En cuanto al tercer objetivo específico, se midió la confiabilidad de consistencia interna de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9), su dimensión Cognitiva/Afectiva (F1) y su dimensión Somático (F2), empleando el coeficiente α alfa de Cronbach de confiabilidad tau-equivalente, y el coeficiente de confiabilidad congénica ω omega de McDonald, para lo cual se emplearon 451 participantes.

Tabla 12. Confiabilidad del PHQ-9

Dimensión	α	ω
PHQ-9 (Total)	.75	.76
Cognitiva/Afectiva (F1)	.70	.72
Somática (F2)	.60	.60

N=451

En la Tabla 12 se puede apreciar que en el Modelo 1 presenta una confiabilidad de consistencia interna adecuada ($\alpha = .75$, $\omega = .76$). Modelo 2, el primer factor presenta una confiabilidad de consistencia interna adecuada ($\alpha_1 = .70$, $\omega_1 = .72$), y el segundo factor evidencia una confiabilidad de consistencia interna pobre ($\alpha_2 = .60$, $\omega_2 = .60$).

En cuanto al cuarto objetivo específico, se evaluó la invarianza de medida entre diferentes grupos según sexo (n=451), edad (n=451), estado civil (n=432), lugar de aplicación (n=446) y grado de instrucción (n=381), empleando AFCMG.

Tabla 13. Resumen de índices de bondad de ajuste del Modelo 2a según sexo, edad, estado civil, lugar de aplicación y grado de instrucción.

	N	Tipo de inv.	χ^2	gl	p	CFI	Δ CFI	RMSEA		SRMR
								Est.	IC 90%	
Sexo	451	Base	65.99	26	<.001	.942		.058	.041 - .076	.044
		Config.	95.30	52	<.001	.936	.006	.061	.041 - .080	.053
		Métrica	116.75	61	<.001	.918	.018	.064	.046 - .081	.095
		Escalar	120.81	68	<.001	.922	.004	.059	.041 - .076	.087
		Estricto	170.07	77	<.001	.863	.059	.073	.058 - .088	.113
Edad	451	Base	65.99	26	<.001	.942		.058	.041 - .076	.044
		Config.	188.85	104	<.001	.887	.055	.085	.066 - .104	.067
Estado Civil*	432	Base	68.13	26	<.001	.935		.061	.044 - .079	.046
		Config.	140.08	78	<.001	.907	.028	.074	.054 - .094	.061
Lugar de aplicación	446	Base	66.63	26	<.001	.939		.059	.042 - .077	.045
		Config.	99.50	52	<.001	.933	.006	.064	.045 - .083	.054
		Métrica	128.13	61	<.001	.908	.025	.069	.052 - .086	.083
Grado de instruc.**	381	Base	70.85	26	<.001	.927		.067	.049 - .086	.05
		Config.	130.96	78	<.001	.912	.015	.073	.051 - .095	.067
		Métrica	148.55	96	<.001	.913	.001	.066	.044 - .086	.096
		Escalar	160.34	110	.001	.917	.004	.06	.038 - .079	.09
		Estricto	168.17	128	<.001	.889	.028	.064	.045 - .082	.104

*Se excluyó el grupo "Viudo"

**Se excluyó el grupo "Sin estudios" y "Técnico completo"

En la tabla 13 se puede observar que considerando la variable sexo, el modelo cuenta con invarianza escalar, pues la diferencia entre los modelos es insignificante (Δ CFI \leq .01), con la excepción del modelo estricto, el cual es significativamente peor que el modelo escalar (Δ CFI=.059). En cuanto a las variables edad y estado civil, no se puede afirmar que el modelo cuente con invarianza configural (Δ CFI₁ =.055; Δ CFI₂ =.028). En cuanto a la variable lugar de aplicación, se puede afirmar que el modelo cuenta con invarianza configural pues

la diferencia entre el modelo base y el configural es insignificante ($\Delta\text{CFI} \leq .01$), pero el modelo métrico si presenta peor bondad de ajuste que el modelo configural ($\Delta\text{CFI} = .025$). En cuanto al grado de instrucción, el modelo cuenta con invarianza escalar. La diferencia entre los modelos es insignificante ($\Delta\text{CFI} \leq .01$), con la excepción del modelo estricto, el cual es considerablemente peor que el modelo escalar ($\Delta\text{CFI} = .028$).

En cuanto al quinto objetivo específico, se identificó la correlación entre la puntuación total de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9), la dimensión Cognitiva/Afectiva (PHQ-9 F1), la dimensión Somática (PHQ-9 F2), la puntuación del PHQ-9 con el método de aplicación de dos pasos (PHQ-ALG), la puntuación de la versión abreviada del Cuestionario sobre la Salud del Paciente (PHQ-2), el índice de severidad de síntomas del PHQ-9 (SEV), la puntuación de la Escala de Ansiedad y Depresión Hospitalaria (HADS) y su dimensión de Depresión (HADD) y de Ansiedad (HADA), y el autoreporte de diagnóstico de depresión (DEP).

El autoreporte de diagnóstico de diagnóstico de depresión hace referencia a si la persona reporta haber cursado con un cuadro depresivo en los últimos dos meses. Esta información se recolecto cuando se completaba la ficha sociodemográfica y se indagaban los antecedentes psicopatológicos. Solo se emplearon 144 participantes para identificar las correlaciones del autoreporte de diagnóstico depresión con las otras variables pues se excluyen los casos con datos incompletos o personas que indican no estas seguras. Para el resto de correlaciones, se emplearon 429 participantes.

Tabla 14. Resumen de índices de correlaciones entre la puntuación del PHQ, HADS y Diagnósticos

		Correlaciones*								
	SEV	PHQ-2	PHQ-9 F1	PHQ-9 F2	PHQ-9	PHQ-ALG	HADA	HADD	HADS	DEP
SEV		.469	.505	.424	.550	.479	.423	.411	.475	.414
PHQ-2	.467		.874	.430	.727	.992	.468	.514	.553	.633
PHQ-9 F1	.517	.881		.445	.810	.872	.539	.565	.625	.615
PHQ-9 F2	.466	.492	.498		.869	.457	.469	.449	.535	.523
PHQ-9	.569	.798	.872	.859		.740	.593	.594	.682	.598
PHQ-ALG	.504	.896	.857	.609	.850		.472	.515	.556	.644
HADA	.471	.505	.595	.480	.622	.502		.518	.871	.445
HADD	.455	.590	.627	.476	.639	.576	.556		.856	.553
HADS	.525	.621	.693	.542	.715	.611	.880	.884		.533
DEP	.442	.728	.712	.592	.706	.794	.436	.596	.563	

Nota: Sev: Índice de severidad de los síntomas del PHQ-9, r: Coeficiente de correlación de Pearson, rpb: Coeficiente de correlación biserial puntual, ρ : Coeficiente de correlación de Spearman, DEP: Autoreporte de diagnóstico de depresión. La diagonal inferior muestra el coeficiente de correlación de Pearson (r), y diagonal superior muestra el coeficiente de correlación de Spearman (ρ).

*N=429 participantes para las correlaciones entre el PHQ-9 y el HADS, y N=144 participantes para las correlaciones de DEP con otras variables.

En la Tabla 14 se pueden observar que el PHQ-2 tiene correlaciones reducidas con el indicador de severidad y con el HADA ($\rho < .50$), moderadas con el HADD, HADS y autoreporte de diagnóstico de depresión ($\rho_1 = .51$; $\rho_2 = .55$; $\rho_3 = .63$). Similarmente, el PHQ-ALG presenta una correlación reducida con el indicador de severidad ($\rho = .48$), reducida con el HADA ($\rho < .50$), moderada con el HADD y el HADS ($\rho_1 = .52$; $\rho_2 = .56$), y con el autoreporte de diagnóstico de depresión ($\rho = .64$). En otras palabras, el PHQ-2 y el PHQ-ALG presenta mayor correlación con la escala de depresión que con la de ansiedad.

La dimensión Cognitiva/Afectiva (F1) del PHQ-9 presenta una correlación moderada con el indicador de severidad ($\rho = .51$), con el HADA, HADD y HADS ($\rho_1 = .54$, $\rho_2 = .57$, $\rho_3 = .63$), y con el autoreporte de diagnóstico de depresión ($\rho = .62$). La dimensión Somática (F2) del PHQ-9 tiene una correlación reducida con el indicador de severidad, el HADA y el HADD ($\rho_1 < .50$; $\rho_2 = .47$; $\rho_3 = .45$), moderadas con el HADS ($\rho = .54$), y moderadas con el autoreporte de diagnóstico de depresión

($\rho=.52$). Asimismo, el PHQ-9 presenta una correlación moderada con el indicador de severidad ($\rho=.55$), moderada con el HADA, HADD y HADS ($\rho_1=.59$; $\rho_2=.59$; $\rho_1=.68$), y moderada con el autoreporte de diagnóstico de depresión ($\rho=.59$). Es decir, tanto el PHQ-9 como sus dimensiones tienen correlaciones muy similares con la escala de depresión que con la de ansiedad.

Por otra parte, en la Tabla 15 se pueden observar los resultados de las comparaciones de media según Sexo (N=451), Diagnóstico (N=343), Grado de instrucción (N=450), Situación laboral (N=450), Fase de Tratamiento (N=434) y Edad (N=451).

Tabla 15. Resumen de comparación de medias del PHQ-2, PHQ-9 y sus dimensiones, y PHQ-ALG según Sexo, Diagnóstico, Grado de instrucción, Situación laboral, Fase de Tratamiento y Edad

		Sexo	Diagnóstico	Grado de instrucción	Situación laboral	Fase de Tratamiento	Edad
F1 PHQ-9	Test	t=-3.6 (gl=432)	F=2200	F=1.41	F=1.53	F=5.47	F=.95
	Sig.	<.001	0.006	0.22	0.21	0.005	0.42
	Tamaño del efecto	0.34	0.095	0.016	0.01	0.025	0.006
F2 PHQ-9	Test	t=-0.7 (gl=448.4)	F=.89	F=1.67	F=.13	F=.034	F=.8
	Sig.	0.504	0.581	0.14	0.94	0.966	0.5
	Tamaño del efecto	0.06	0.041	0.018	<.001	<.001	0.005
PHQ-9	Test	t=-2.5 (gl=440.5)	F=1.67	F=1.8	F=.56	F=2.191	F=.19
	Sig.	0.013	0.057	0.112	0.63	0.113	0.9
	Tamaño del efecto	0.24	0.075	0.02	0.004	0.01	0.001
PHQ-ALG	Test	t=-3.2 (gl=427.2)	F=2.23	F=2.57	F=.62	F=5.161	F=1.07
	Sig.	0.001	0.006	0.026	0.61	0.006	0.36
	Tamaño del efecto	0.3	0.097	0.028	0.004	0.023	0.007
PHQ-2	Test	t=-3.4 (gl=438.6)	F=2379	F=1.34	F=.88	F=4.869	F=.64
	Sig.	0.001	0.003	0.025	0.45	0.008	0.59
	Tamaño del efecto	0.32	0.103	0.015	0.006	0.022	0.004

Nota: Para Tamaño del Efecto se empleó la d de Cohen para comparar las medias de dos grupos (Sexo), y η^2 para el resto

En cuanto a la dimensión Cognitiva/Afectiva (F1) del PHQ-9, se hallaron diferencias de medias estadísticamente significativas ($p < .05$) según sexo, diagnósticos oncológicos, fase de tratamiento y edad, pero solo la diferencia según sexo presenta un tamaño del efecto moderado ($d > .30$) y el resto eran prácticamente insignificantes.

En cuanto a la dimensión Somática (F2) del PHQ-9, las diferencias de media no fueron estadísticamente significativas ($p > .05$) y presentaron tamaños del efecto reducidos ($d < .1$). En cuanto al PHQ-9, se halló que la diferencia de medias según sexo fue estadísticamente significativa ($p < .05$) pero el tamaño de efecto fue reducido ($d < .3$). El resto de las diferencias de medias no fueron estadísticamente significativas y presentaron un tamaño del efecto irrelevante.

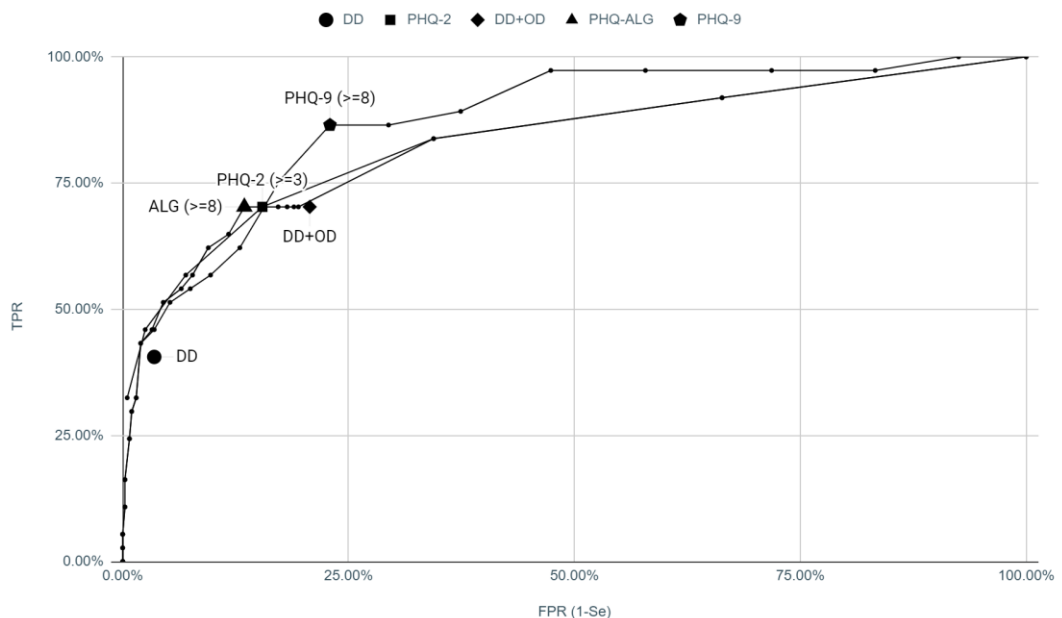
En cuanto al PHQ-ALG y el PHQ-2, en ambos se hallaron diferencias de medias estadísticamente significativas según sexo ($p < .01$), diagnóstico ($p < .01$), grado de instrucción ($p < .05$), y fase de tratamiento ($p < .01$), pero solo la diferencia de medias según sexo presentó un tamaño del efecto moderado ($d = .3$) y el resto eran prácticamente irrelevantes ($d < .1$).

En cuanto al sexto objetivo específico, se identificó el mejor punto de corte de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9), la puntuación del PHQ-9 con el método de aplicación de dos pasos (PHQ-ALG), la puntuación de la versión abreviada del Cuestionario sobre la Salud del Paciente (PHQ-2). Se usó el HADS como el *gold standard*, y se identificaron los indicadores de precisión diagnóstica para cada uno de los puntos de corte. Se asumió una

prevalencia de depresión de 9.79%. Para este objetivo se emplearon 438 participantes sin datos incompletos en el PHQ-9 y HADS.

Se midió la precisión diagnóstica de todos los puntos de corte posibles del PHQ-9, PHQ-ALG y PHQ-2, y dos métodos diagnósticos recomendados por los autores originales (EXT y COL). Para ver el detalle de los indicadores de precisión diagnóstica de cada prueba, revisar el Anexo 10.

Gráfico 6. ROC de las diferentes formas de calificación del PHQ-9



En el gráfico 6 se pueden apreciar las curvas del PHQ-2, PHQ-9 y el PHQ-ALG. Además, se añadió el método algorítmico de los autores originales del PHQ-9, el cual diagnosticaba “Desorden Depresivo Mayor”, “Otros Síndromes Depresivos” o “Ningún desorden”. Como este método no emplea puntos de corte, se usó como referente el diagnóstico que arrojaban, ya sea DD (Diagnóstico de Desorden Depresivo Mayor) y OD+DD (Diagnóstico de Desorden Depresivo

Mayor u Otros Síndrome Depresivos), y se compararon contra el *Gold Standard* para observar la precisión diagnóstica de estos métodos como puntos (no como curvas).

Se puede apreciar que las curvas del PHQ-2, PHQ-9 y PHQ-ALG se comportan de manera muy similar. En cuanto al PHQ-9 y el PHQ-ALG, el punto de corte ≥ 8 presenta el mejor poder diagnóstico. En cuanto al PHQ-2, el punto de corte ≥ 3 presenta el mejor poder diagnóstico. Más aún, el PHQ-9 presenta mayor sensibilidad que el resto de los métodos, y el PHQ-2 presenta la menor ratio de falsos positivos, es decir, mayor especificidad.

Tabla 16. Resumen de indicadores de rendimiento del instrumento

Indicadores	DD	OD+DD	PHQ-2 (≥ 3)	PHQ-ALG (≥ 8)	PHQ-9 (≥ 8)
PPV	.517	.239	.295	.325	.258
NPV	.967	.967	.969	.969	.984
ACC	.760	.785	.833	.852	.779
TPR (Se)	.405	.703	.703	.703	.865
FNR	.297	.297	.297	.297	.135
TNR (Sp)	.793	.793	.845	.865	.771
FPR	.035	.207	.155	.135	.229
AUC	-	-	.885	.871	.889
LR+	11.61	3.39	4.55	5.22	3.77
LR-	0.37	0.37	0.35	0.34	0.18
LR+ concreto	45.97%	17.14%	23.25%	26.37%	19.25%
LR- concreto	-5.88%	-5.88%	-6.12%	-6.20%	-7.93%
DOR	30.97	9.06	12.92	15.19	21.5
NNM	4.08	4.63	5.93	6.64	4.54
NNMw	2.09	3.01	3.51	3.74	3.66
Distancia a la diagonal	0.58	0.74	0.76	0.77	0.81
J de Youden	0.20	0.55	0.55	0.57	0.64
Diagonal Máxima	0.32	0.56	0.59	0.61	0.67

Nota: PPV: Valor predictivo positivo, NPV: Valor predictivo negativo, ACC: Precisión, TPR: Proporción de verdaderos positivos (Sensibilidad), FNR: Proporción de falsos negativos, TNR: Proporción de verdaderos negativos (Especificidad), FPR: Proporción de falsos positivos, LR: Razón de verosimilitud (Likelihood Ratio), DOR: Razón de cociente de probabilidades (Diagnostic Odds Ratio), NNM: Numero necesitado para diagnóstico errado (Number needed to misdiagnose), NNMw: NNM ponderado.

En la Tabla 16 se pueden apreciar los indicadores del rendimiento de las diferentes formas de calificación del instrumento. Se aprecia que todos los métodos presentan un alto NPV (>95%). Esto significa que si puntúan por debajo del punto de corte, hay por lo menos un 95% de probabilidades de que realmente no tengan depresión.

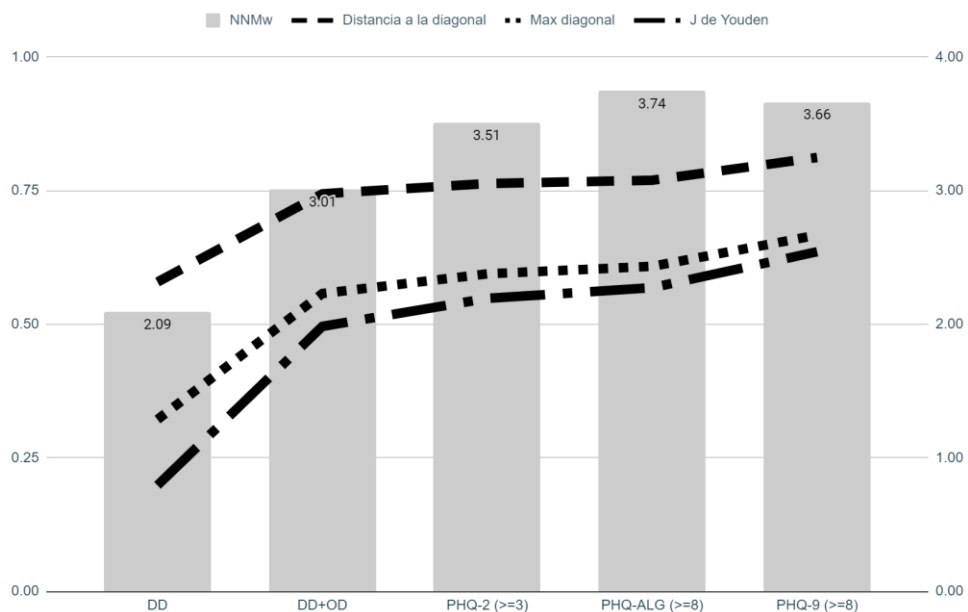
Con respecto a la sensibilidad, el método algorítmico DD presenta una sensibilidad mediocre (TPR=.41), el método algorítmico OD+DD, el PHQ-2 y el PHQ-ALG presentan una sensibilidad más adecuada (TPR=.703), y el PHQ-9 presenta la mayor sensibilidad (TPR=.865).

Con respecto a la especificidad, el PHQ-9 presenta la menor especificidad aunque igual es apropiada (TNR=.771), el método algorítmico DD y ODD presentan una sensibilidad elevada (TNR=.793), y la mayor especificidad corresponden al PHQ-2 (TNR=.845) y al PHQ-ALG (TNR=.865).

Asimismo, el método DD presenta el mayor LR+ de 11.61, lo cual sugiere que si una persona es diagnosticada con “Desorden Depresivo” a través del método algorítmico es 11.61 veces más probable de tener depresión, o puesto de otra forma, la probabilidad de que tenga depresión aumenta en un 45.97%. Todos los métodos presentan un LR- reducido, es decir, si puntúan por debajo del respectivo punto de corte, la probabilidad de tener depresión se reduce de 9.79% a 3.67% en el caso del PHQ-2, y hasta 1.79% en el caso de PHQ-9. Asimismo, el PHQ-ALG presenta la mayor precisión (ACC=.852) y el mayor NNM (NNM=6.64), es decir, que de 100 personas, 85 serán correctamente catalogadas como verdaderos positivos o verdaderos negativos, y considerando la prevalencia previa se necesitan evaluar a

6.64 personas para diagnosticar a una de manera errada (ya sea falso positivo o falso negativo). Sin embargo, cuando se pondera el NNM, considerando que para una prueba de tamizaje es preferible tener falsos positivos que falsos negativos, se encuentra que el PHQ-2, PHQ-9 y PHQ-ALG presentan un NNMw similar (NNMw=3.64-3.66).

Gráfico 7. Comparación de indicadores de precisión diagnóstica de las diferentes formas de calificación del PHQ-9



En el gráfico 7 se puede comparar visualmente el poder diagnóstico de las diferentes formas de calificación del PHQ. En cuanto al método DD, se aprecia que tiene un rendimiento bajo (NNMw=2.09; Distancia de la diagonal=.58; J=.20; MaxD=.32), al igual que el OD+DD (NNMw=3.01; Distancia de la diagonal=.74; J=.50; MaxD=.56). El PHQ-2, PHQ-9 y PHQ-ALG, las tres presentan un NNMw similar (>3.5), sin embargo, los otros indicadores reflejan que el PHQ-9 presenta un mejor rendimiento que el PHQ-ALG y el PHQ-2.

CAPÍTULO V: DISCUSIÓN

Con respecto al primer objetivo específico sobre la validez basada en el contenido evaluada mediante el criterio de jueces, se evidenció mediante los valores V de Aiken que los reactivos de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9) medían de manera clara y relevante la depresión.

Todos los reactivos obtuvieron valores superiores a 0.8, con un límite inferior superior a 0.7 en función a la representatividad, claridad y relevancia del reactivo. Los resultados son congruentes con lo reportado por Calderón et al. (2012), quienes concluyeron mediante criterio de jueces expertos que el instrumento era apropiado para medir la depresión en el contexto hospitalario peruano.

Ahora bien, el índice de confiabilidad inter-jueces (ICC) refleja un bajo acuerdo entre los jueces sobre la puntuación de los ítems ($ICC < 0.5$), con la mayor confiabilidad correspondiendo a la categoría Claridad ($ICC = 0.46$) y la menor al criterio Representación ($ICC = 0.18$). Sin embargo, a pesar de la baja confiabilidad de las puntuaciones, se debe resaltar que todos los ítems puntuaron por encima de 3 en una escala del 1 al 5. Es decir, hubo un bajo acuerdo entre los jueces sobre si los ítems eran moderados, buenos o excelentes para medir de manera clara y relevante la depresión, pero ninguno indicó que fuesen inadecuados.

Por estas razones, la evidencia basada en el contenido del PHQ-9 sugiere que las puntuaciones del instrumento son válidas para realizar tamizaje de depresión.

Con respecto al segundo objetivo específico sobre la validez basada en la dimensionalidad de la estructura interna, primero se realizó un análisis factorial exploratorio. Luego, mediante un análisis factorial confirmatorio, se contrastaron modelos equivalentes y se confirmó la estructura hallada.

En el análisis factorial exploratorio, considerando que los datos no presentaban normalidad multivariada, se empleó el método de extracción factorización de ejes principales y se extrajeron dos modelos.

Primero, se extrajo el Modelo 1 especificando la extracción de un solo factor que explica 37% de la varianza. La razón detrás de dicha especificación es que el PHQ-9 es fue originalmente diseñado como un instrumento unidimensional, y su estructura unidimensional ha sido observada en múltiples investigaciones con poblaciones latinoamericanas; por ejemplo, Baader et al. (2012), Merz, Malcarne, Roesch, Riley, and Sadler (2011) o Saldivia et al. (2019). Si bien el Modelo 1 es fácilmente interpretable, los índices de bondad de ajuste evidencian un ajuste pobre a los datos.

Seguidamente, extrayendo factores con autovalores superiores a 1, y usando el punto de inflexión del gráfico de sedimentación como guía, se extrajo el Modelo 2 el cual presenta una estructura bidimensional que explica 47% de la varianza. En el Modelo 2 se encontró una dimensión Somática y una dimensión Cognitiva/Afectiva. Estas dos dimensiones presentaban ítems con cargas factoriales definidas (ver Tabla 8), excepto por el ítem 1 el cual cargaba con ambas. Con respecto a los índices de bondad de ajuste, se aprecia que el Modelo 2 presenta un ajuste moderado a los datos, es decir, un mejor ajuste a los datos que el Modelo 1.

Cabe notar que las cargas factoriales de los ítems 1 (Anhedonia), ítems 7 (Concentración) y ítems 8 (Psicomotriz) fueron reducidas ($\lambda < 0.40$). Además, estos ítems presentaron bajas comunalidades. Esto podría indicar que los ítems 1, 7 y 8 están midiendo otros factores diferentes a los síntomas somáticos y cognitivos/afectivos de la depresión. Por ejemplo, puede ser que los problemas de concentración, anhedonia y las alteraciones psicomotrices reportadas por los participantes sean debidas a su situación de hospitalización o en reacción al tratamiento quimioterapéutico u estén midiendo otras variables extrañas diferentes a la depresión independientemente una de otra. Esto explicaría porque las comunalidades y las cargas factoriales son reducidas.

Sin embargo, considerando las recomendaciones de Mundfrom, Shaw, and Ke (2005) sobre la cantidad ideal de participantes para realizar análisis factorial en condiciones de baja comunalidad, y las recomendaciones de Hair, Black, Babin, and Anderson (2014) sobre el umbral mínimo de significancia práctica de las cargas factoriales dependiendo del tamaño muestral, se puede argumentar que salvo por el ítem 7 (Concentración), todas las cargas factoriales reflejan una estructura latente confiable.

De hecho, la estructura bidimensional encontrada es similar a lo reportado por otras investigaciones, coincidiendo en que los ítems 1, 2, 6 y 9 comprenden una dimensión Cognitiva/Afectiva, los ítems 3, 4 y 5 comprenden una dimensión somática, y cierta ambigüedad por parte de los ítems del 7 y 8 (Cassiani-Miranda et al., 2017; Doi, Ito, Takebayashi, Muramatsu, y Horikoshi, 2018).

Más aún, Doi et al. (2018) notan que los estudios sobre la estructura factorial del PHQ-9 que emplean muestras clínicas (e.g. personas con trastornos depresivos, trastornos de ansiedad, trastorno obsesivo-compulsivo, entre otros) suelen presentar una estructura unifactorial, mientras que los estudios que emplean poblaciones de personas con problemas de salud físicos suelen identificar modelos bidimensionales, con una dimensión somática y otra dimensión Cognitiva/Afectiva. También encontraron en su investigación psicométrica que el PHQ-9 presentaba una estructura bifactorial, es decir, de dos dimensiones (Cognitiva/Afectiva y Somática) y un factor general.

Para este estudio no se contaron con los programas estadísticos para poder evaluar este modelo, y se recomienda para futuras investigaciones considerar una estructura bifactorial.

Los resultados del análisis factorial exploratorio fueron posteriormente reconfirmados mediante análisis factorial confirmatorio. Se planteó una primera dimensión Cognitiva/Afectiva comprendida por los ítems 1 (Anhedonia), 2 (Distimia), 6 (Fracaso), 7 (Concentración), y 9 (Ideación suicida); y una segunda dimensión Somática comprendida por los ítems 3 (Sueño), 4 (Cansancio), 5 (Apetito) y 8 (Psicomotriz).

Se optó por incluir el ítem 1 en la primera dimensión por dos razones. Primero, el ítem 1 trata sobre el “interés” y el “placer”, por lo cual es más congruente ponerlo en la dimensión Cognitiva/Afectiva. Segundo, todos los estudios sobre la dimensionalidad del PHQ-9 citados en esta investigación ponen al ítem 1 y 2 en la misma dimensión.

Luego, mediante el análisis factorial confirmatorio, se indagó si el modelo bidimensional simple (Modelo 2a) era el más apropiado al contrastarlo de nuevo con el modelo unidimensional y con otros modelos bidimensionales equivalentes.

Para este procedimiento se empleó el método de máxima verosimilitud por dos razones. Primero, porque permite evaluar más índices de bondad de ajuste que otros métodos de extracción. Segundo, porque se considera robusto cuando los datos no reflejan una distribución normal multivariada, salvo casos extremos, aunque se deben emplear puntos de corte ligeramente más estrictos para determinar una bondad de ajuste adecuada (Yang y Liang, 2013). Por lo tanto, en cuanto a los indicadores de bondad de ajuste, se emplearon los puntos de corte recomendaciones por varios autores (Brown, 2006; Byrne, 2013; Kline, 2010; Yang y Liang, 2013). Además, no se presentaron diferencias significativas entre el AFC usando el método de máxima verosimilitud y el AFE usando el método de factorización de ejes principales.

Ahora bien, en todos los modelos, las pruebas chi-cuadrado indicaron una pobre bondad de ajuste. Sin embargo, este indicador no es comúnmente recomendado en investigaciones con muchos participantes debido a ser susceptible a la cantidad de casos empleados. En esta investigación se emplearon 451 casos para el análisis factorial confirmatorio, por lo tanto, es razonable esperar que el test chi-cuadrado presente resultados estadísticamente significativos. En cuanto al resto de indicadores, ningún modelo evidenció un ajuste excelente o perfecto.

El modelo unidimensional reflejó resultados mixtos. El RMSEA y el SRMR reflejan un ajuste adecuado, sin embargo, el resto de los indicadores demuestran una bondad de ajuste mediocre.

Por otra parte, el modelo bidimensional simple (M2a) reflejó un ajuste entre moderado a elevado (CFI=.942; TLI=.92; AGFI=.946; SRMR=.044; RMSEA=.058), y es el modelo con el mejor ajuste. Sin embargo, es importante indicar que el modelo bidimensional jerárquico (M2b) obtuvo resultados muy similares al M2a.

Por estos motivos, se afirma que la evidencia basada en la estructura interna evaluada mediante análisis factorial exploratorio y confirmatorio apoya una estructura factorial simple de dos dimensiones en acorde con lo teóricamente esperado. Por lo tanto, la evidencia sugiere que las puntuaciones del instrumento son válidas para realizar tamizaje de depresión, midiendo una dimensión Cognitiva/Afectiva y una dimensión Somática.

Además, la gran similitud entre el modelo bidimensional simple y el modelo bidimensional jerárquico evidencia que es válido emplear la suma de los puntajes totales del instrumento para medir la depresión en general. Asimismo, sería ideal para futuras investigaciones evaluar la bondad de ajuste de un modelo bifactorial.

Con respecto al tercer objetivo específico sobre la validez basada en estructura interna mediante el análisis de la confiabilidad, se emplearon los coeficientes de confiabilidad alfa de Cronbach y omega de McDonald.

El coeficiente alfa de Cronbach indica que las dos dimensiones del Modelo 2 presentan una confiabilidad moderada en la dimensión Cognitiva/Afectiva ($\alpha_1=.70$) y una confiabilidad reducida en la dimensión Somática ($\alpha_2<.60$). Sin embargo, se sabe que el coeficiente alfa de Cronbach suele subestimar la confiabilidad de un instrumento cuando hay pocos ítems, cuando no se cumple la presuposición de tau-equivalencia, o cuando los ítems no miden un mismo constructo (Tavakol y Dennick, 2011). Asimismo, el alfa de Cronbach se sesga cuando las varianzas de los errores están correlacionadas, cuando las cargas factoriales son menores a .70, cuando las diferencias entre las cargas factoriales son mayores a .20 (Deng y Chan, 2017).

En ese sentido, la dimensión Somática y la dimensión Cognitiva/Afectiva cuentan con pocos ítems, cuatro y cinco ítems respectivamente. Además, mediante el análisis factorial se observa que los ítems presentan cargas factoriales por debajo de .70 y con diferencias mayores a .20 (e.g. $\lambda_1=.70$; $\lambda_7=.40$). Por lo tanto, era de esperar que el alfa de Cronbach probablemente sesgue la confiabilidad del instrumento.

Es por estos motivos que se empleó el coeficiente omega de McDonald. Siguiendo los puntos de corte recomendados por Ventura-Leon and Caycho-Rodríguez (2017), se determinó que la dimensión Cognitiva/Afectiva presenta una confiabilidad aceptable ($\omega_1=.72$) y la dimensión Somática presenta una confiabilidad pobre ($\omega_2=.60$).

Estos resultados son similares a lo hallado mediante el alfa de Cronbach ($\alpha_1=.70$; $\alpha_2=.59$). Esto puede deberse a la situación de salud y tratamientos de los

participantes, los cuales podrían sesgar la confiabilidad del PHQ-9 para medir los síntomas somáticos de la depresión. Por ejemplo, el coeficiente omega de McDonald de la dimensión Somática cambia de un valor pobre ($\omega_2=.59$) a un valor casi aceptable ($\omega_2=.694$) cuando se excluyen de la muestra a los participantes que reciben tratamiento quimioterapéutico. Esto probablemente sugiere que la medición de la dimensión somática de la depresión presenta un elevado error el cual probablemente se deba a la confluencia de los síntomas y tratamientos oncológicos, entre otras variables desconocidas.

Por otra parte, el alfa de Cronbach y el omega de McDonald reflejan una confiabilidad adecuada cuando se considera el modelo unidimensional ($\alpha=.70$, $\omega=.72$). Esto es importante porque el instrumento debe de interpretarse empleando la puntuación total de tal manera que provea información relevante para los profesionales de salud. El puntaje de cada dimensión es un dato interesante, pero la mayor utilidad vendrá de la puntuación total del instrumento.

Por estos motivos, se afirma que la evidencia basada en la estructura interna evaluada mediante análisis de la confiabilidad apoya que el PHQ-9 es confiable para medir los síntomas cognitivos y afectivos de la depresión, así como la depresión en general. Sin embargo, las puntuaciones de los ítems correspondientes a la dimensión somática deben ser interpretados con cautela.

Con respecto al cuarto objetivo específico, específico sobre la evidencia de validez basada en la estructura interna mediante la evaluación de la invarianza de medida entre los grupos, se realizó un Análisis Factorial Confirmatorio Multi-

Grupo (AFCMG) siguiendo los pasos, recomendaciones y puntos de corte mencionados por (Xu y Tracey, 2017).

Se pudo evaluar la invarianza de medida entre grupos según sexo, edad, estado civil, lugar de aplicación y grado de instrucción. Se tuvieron que eliminar algunos sub-grupos para poder realizar el análisis debido a que el tamaño de estos grupos era muy reducido como para realizar el AFCMG, como por ejemplo el conjunto de personas viudas (N=18).

Por otra parte, no se pudo realizar el análisis según fase de tratamiento pues cuando se agrupaba según esta variable la matriz de varianzas de la data no era positiva-definida. Esto puede suceder por varias razones. Por ejemplo, puede ser que el tamaño de la muestra haya sido muy reducido cuando se agrupaba la data según estas variables. Otra posibilidad es que el modelo simplemente no tenía sentido cuando se agrupaba según estas variables pues en algunos de los sub-grupos el modelo era bidimensional y en otros unidimensional. Se debe de considerar que la “fase de tratamiento” (pre-tratamiento, en tratamiento y post-tratamiento) no son categorías mutuamente exclusivas ni bien definidas. Por ejemplo, un paciente puede estar en tratamiento quimioterapéutico pero en pre-cirugía, o puede estar en post-tratamiento pero esperando una sesión de quimioterapia para prevenir una recaída.

En cuanto a la variable sexo, se aprecia que el PHQ-9 presenta invarianza escalar. Esto quiere decir que las respuestas de los hombres y las mujeres ante el PHQ-9 se ajustan por igual al modelo bidimensional, con las mismas cargas factoriales y el mismo intercepto. Por lo tanto, se puede afirmar que el PHQ-9 no presenta mayor sesgo en función al sexo.

En cuanto a la variable edad y estado civil, se aprecia que el PHQ-9 no presenta invarianza de medida, pues en estos grupos el PHQ-9 es interpretado diferencialmente. En otras palabras, el modelo bidimensional que se especificó en el AFC no es el más apropiado para todos los grupos. Según el estado civil y la edad de la persona, el instrumento medirá la depresión de manera distinta, ya sea midiéndola de manera unidimensional, o quizás bidimensional pero con los ítems relacionándose de otra forma y midiendo otros constructos latentes.

Esto tiene sentido pues es de esperarse que la depresión se manifieste de maneras diferentes en pacientes oncológicos jóvenes en comparación a pacientes oncológicos adultos o adultos mayores. Por ejemplo, en el Anexo 11 se puede apreciar el análisis factorial exploratorio según grupo de edad. En los pacientes jóvenes se identifica fácilmente una dimensión Cognitiva/Afectiva compuesta por la Distimia (ítem 1), Anhedonia (ítem 2) y Culpa (ítem 6), y una dimensión somática compuesta Sueño (ítem 3), Cansancio (ítem 4), y Apetito (ítem 5). Pero en los pacientes entre 36 y 50 años se observa que la Distimia (ítem 1) se agrupa junto con el Sueño (ítem 3) y el Cansancio (ítem 4) como síntomas somáticos, y la otra dimensión más afectiva se compone más bien por la pérdida de interés en las actividades cotidianas (ítem 2), la Culpa (ítem 6) y la ideación suicida (ítem 9). La distimia esta siendo interpretada de manera diferente según la edad del participante, y por lo tanto, es de esperarse que no haya invarianza configural.

Por otra parte, en cuanto a la invarianza de medida según el estado civil, la experiencia de depresión de un paciente oncológico soltero no es necesariamente igual a la de un paciente oncológico casado/conviviente. Por ejemplo, la

responsabilidad por el sostén económico de la familia o la maternidad pueden ser elementos que influyan en cómo se interpretan los ítems. Además, se debe de considerar que los tamaños muestrales son reducidos. Si bien 432 participantes es usualmente suficiente para un análisis factorial, se vuelve insuficiente cuando se divide en 3 grupos ($N_1=222$; $N_2=44$; $N_3=166$). Quizás con mayor muestra en estos grupos, o mayor homogeneidad en los tamaños muestrales entre los grupos, se evidenciaría una mayor bondad de ajuste del modelo a los resultados.

En cuanto a la variable lugar de aplicación, se aprecia que el PHQ-9 presenta invarianza configural. Esto quiere decir que el PHQ-9 mide la misma dimensión Cognitiva/Afectiva y la misma dimensión somática en pacientes hospitalizados y en pacientes en quimioterapia ambulatoria. Sin embargo, Las cargas factoriales no son iguales entre los grupos, es decir, el grado en el cual estos ítems miden estos constructos varía según grupos. Por ejemplo, el ítem 7 (Atención/concentración) presentaba una adecuada carga factorial con la dimensión Cognitiva/Afectiva en los pacientes hospitalizados ($\lambda=.45$), pero casi inexistente en los pacientes en quimioterapia ($\lambda=.18$). Similarmente, el ítem 5 (Apetito) era un buen indicador de los síntomas somáticos de la depresión en pacientes hospitalizados ($\lambda=.53$) pero no en los ambulatorios de quimioterapia ($\lambda=.27$). Esto se puede deber a que en los pacientes ambulatorios que acuden al servicio de quimioterapia, si bien la desconcentración y la falta de apetito puede ser un indicador de síntoma somáticos de depresión, también pueden deberse al tratamiento quimioterapéutico mismo. En cambio, la desconcentración y la falta de apetito son mejores predictores de sintomatología depresiva en pacientes hospitalizados. Por lo tanto, se puede afirmar que el PHQ-9 mide los mismos constructos en pacientes hospitalizados y

ambulatorios de quimioterapia, pero no con la misma carga factorial, por lo que puede haber un sesgo.

En cuanto a la variable Grado de instrucción, se aprecia que el PHQ-9 presenta invarianza escalar. Esto quiere decir que las respuestas de los participantes con educación primaria completa, secundaria completa, y superior completa se ajustan por igual al modelo bidimensional, con las mismas cargas factoriales y el mismo intercepto. Por lo tanto, se puede afirmar que el PHQ-9 no presenta mayor sesgo en función al grado de instrucción en los grupos mencionados. No se pudo investigar si el PHQ-9 era invariante considerando las personas sin estudios primarios o con estudios técnicos pues no se contaba con suficientes participantes como para realizar el AFCMG.

Por lo tanto, se puede afirmar que el PHQ-9 no presenta sesgo en función del sexo ni del grado de instrucción de la persona. Sin embargo, puede que los ítems no miden su constructo latente con la misma fuerza dependiendo de si el paciente esta hospitalizado o en quimioterapia ambulatoria. Asimismo, el instrumento no es invariante al considerar la edad y el estado civil de las personas.

Por lo tanto, la evidencia indica que el instrumento es apropiado realizar tamizaje de depresión en pacientes oncológicos; pero, se deben de interpretar con cuidado las diferencias de medias según lugar de aplicación, y en general si se compara según estado civil y edad.

Con respecto al quinto objetivo específico de identificar la evidencia convergente de validez, se identificaron las correlaciones entre del PHQ-9 con la

Escala de Ansiedad y Depresión Hospitalaria (HADS) y el autoreporte de diagnóstico de depresión.

Lo esperado es que un instrumento que mide un constructo presente una mayor correlación con otro instrumento que mide el mismo constructo que con un instrumento que mide otro constructo.

Cuando se observan los coeficientes de correlación de Pearson entre el PHQ-9 y sus dimensiones con el HADA y HADD, podemos observar que las correlaciones con el HADD son marginalmente mayores que las correlaciones con el HADA. La diferencia es aún menos pronunciada cuando se emplea el coeficiente ρ de Spearman, el cual es más apropiado considerando la naturaleza ordinal de las variables. En otras palabras, el PHQ-9 y sus dimensiones se correlacionan con instrumentos que miden depresión y ansiedad por igual. Esto puede darse por diversas razones.

Primero, los pacientes oncológicos están cursando momentos de elevado estrés, y es común que experimenten malestar emocional como parte de su proceso de adaptación a la enfermedad y al tratamiento (Holland et al., 2018). Este malestar emocional es más ambiguo que un marcado cuadro de depresión o de ansiedad (Martínez et al., 2017). Por lo tanto, existe la posibilidad de que el PHQ-9 no esté midiendo la depresión exclusivamente, sino el “malestar emocional”. Y es por eso que presenta correlaciones casi idénticas con instrumentos que miden depresión e instrumentos que miden ansiedad. Después de todo, el PHQ-9 también es recomendado como un instrumento de tamizaje de malestar emocional (Holland et al., 2018).

Segundo, puede ser que la presencia de sintomatología producto del cáncer y los tratamientos oncológicos influya en cómo estos instrumentos miden la depresión y la ansiedad, y por lo tanto aumente artificialmente las correlaciones entre estos constructos. Por ejemplo, si observamos las correlaciones del PHQ-2 el cual solo mide la Distimia (ítem 1) y la Anhedonia (ítem 2), y no considera los síntomas somáticos, podemos ver que se correlaciona más con el HADD que con el HADA. Similarmente, cuando se recalculan los coeficientes segmentando según lugar de aplicación, se aprecian que los pacientes que acudían a quimioterapia ambulatoria presentaban una mayor validez convergente, es decir, el PHQ-9 y sus dimensiones se correlacionaban más con el HADD que con el HADS. Y esto probablemente se debe a que los pacientes que realizan quimioterapia ambulatoria no suelen experimentar elevados niveles de malestar físico mientras que realizan quimioterapia. Usualmente el malestar físico lo experimentan a los dos o tres días, y si hay mucho riesgo de que experimenten neutropenia, fiebre o náuseas muy intensas, se les hospitaliza.

Por otra parte, se evidencia una clara validez convergente cuando se correlacionan las puntuaciones del PHQ-9 y sus dimensiones con el autoreporte de diagnóstico de depresión. Es decir, la relación entre indicar haber cursado o estar cursando con un cuadro depresivo fue mayor con el PHQ-9 y sus dimensiones que con el HADA.

Por estos motivos, se afirma que la evidencia convergente de validez apoya que el PHQ-9 es confiable para tamizaje de depresión en aquellos que pacientes que

no experimenten malestar físico significativo, sin embargo, ha de interpretarse con cuidado en pacientes hospitalizados.

Con respecto al sexto objetivo específico de identificar los puntos de corte más apropiados del PHQ-9, se emplearon una serie de indicadores para poder identificar el mejor punto de corte para nuestro objetivo propuesto.

Se encontró que para el PHQ-9 el mejor punto de corte es ≥ 8 . Este presentó una sensibilidad aceptable (TPR=.87) pero una especificidad baja (TNR=.77). También se aprecia un elevado valor predictivo negativo (NPV=.97) lo cual es ideal para realizar tamizaje, pero un bajo valor predictivo positivo (PPV=.26). Y en general, el PHQ-9 demostró un rendimiento adecuado (AUC=.90).

Esto es similar a lo reportado por otras investigaciones. Por ejemplo, Hartung et al. (2017) estudiaron la precisión diagnóstica del PHQ-9 empleando una muestra de más de tres mil pacientes oncológicos alemanes. Se identificó que el mejor punto de corte era ≥ 7 (TPR=.83; TNR=.61; PPV=.10; NPV=.99) y que en general la prueba tenía un AUC de 0.80. La precisión del PHQ-9 en esta investigación fue superior a lo hallado por Hartung et al. (2017), sin embargo, ellos mencionan en su discusión que anteriores investigaciones sobre el PHQ-9 en pacientes oncológicos se ha encontrado una mejor precisión diagnóstica, brindado como ejemplo un mayor AUC ($>.90$).

Similarmente, nuestros resultados coinciden dentro de los intervalos de confianza de los resultados de Levis et al. (2017). Ellos realizaron un meta-análisis sobre los puntos de corte del PHQ-9 empleando 13 estudios con más de 4500 pacientes. El punto de corte ≥ 8 presentaba resultados similares a los hallados en

esta investigación, con una sensibilidad de 0.93 (IC_{95%}=.85-.97) y una especificidad de 0.78 (IC_{95%}=.74-.82).

En Latinoamérica, Saldivia et al. (2019) encontraron resultados similares en pacientes chilenos de un centro de atención primaria de salud con el punto de corte de ≥ 7 (TPR=.80; TNR=.77; PPV=.30; NPV=.91; LR+=3.52; LR-=0.26), y Fabara (2013) en pacientes de atención primaria de Ecuador con el punto de corte ≥ 10 (TPR=.84; TNR=.91; PPV=.51; NPV=.98; LR+=9.54; LR-=0.17).

Cabe destacar que la mayoría de estas investigaciones escogen el punto de corte simplemente maximizando la especificidad y la sensibilidad. Si bien esto es muchas veces suficiente para escoger un punto de corte adecuado, no siempre es lo más recomendable. Esto se debe a que esto asume que la especificidad es igual de importante que la sensibilidad. Sin embargo, esto no es necesariamente verdad dependiendo del contexto.

Por ejemplo, si uno maximiza la precisión o el NNM del PHQ-9, obtiene como punto de corte ideal ≥ 15 . Con este punto de corte 93.38% de los casos son correctamente categorizados como verdaderos positivos o verdaderos negativos, y se necesitan 13.6 casos para erróneamente diagnosticar a una persona. Cualquier otro punto de corte es más propenso a diagnosticar erróneamente que ≥ 15 . Sin embargo, esto asume que un falso negativo es igual a un falso positivo, lo cual simplemente no es cierto.

Maxim et al. (2014) discuten las implicancias de falsos negativos y falsos positivos en diferentes tipos de pruebas de tamizaje en el ámbito médico, por ejemplo, un falso positivo genera estrés innecesario en la persona. Hartung et al.

(2017) también discuten las implicancias de escoger un punto de corte con alta especificidad a costas de una baja sensibilidad, indicando que una mayor tasa de falsos positivos implicaría dedicar recursos limitados a la atención de personas que no tienen depresión, lo cual se traduce en sobrecostos o sobrecarga del personal de salud.

Sin embargo, es necesario entender que una prueba de tamizaje no tiene el objetivo de diagnosticar una condición. Esta debe ser usada como un primer filtro dentro de un flujo de atención. Con ello en mente, para una prueba de tamizaje de depresión en pacientes oncológicos es preferible contar con falsos positivos que falsos negativos. De todas maneras, si una persona tiene un falso positivo, no cambia el hecho que es un paciente oncológico que esta cursando una situación difícil. Inclusive si no tiene depresión, si puntúa por encima del punto de corte, igual sería evidencia de malestar psicológico de algún tipo. En ese sentido, estas personas se beneficiarían de la atención psicológica inclusive si no tienen depresión. Por otra parte, si una persona tiene un falso negativo, no continuaría dentro del flujo de atención. Se trataría de un paciente oncológico con depresión que no recibe tratamiento oportuno. El riesgo puede ser desde una menor adherencia al tratamiento hasta conducta suicida.

Por lo tanto, en esta investigación se empleó el índice NNMw (Numero necesitado para diagnóstico errado ponderado) desarrollado por Habibzadeh, Habibzadeh, and Yadollahie (2016). Se pondero el peso relativo de un falso negativo y un falso positivo empleando un criterio de 5 a 1, y en consecuencia indica que el punto de corte más apropiado para el objetivo propuesto es ≥ 8 (NNMw=3.7).

Se decidió que el costo de un falso negativo era cinco veces mayor al de un falso positivo a través de la opinión de los jueces expertos.

Con respecto al séptimo objetivo específico de contrastar el poder diagnóstico del método directo (PHQ-9) y del método de dos pasos (PHQ-ALG), se emplearon múltiples indicadores del rendimiento del instrumento.

El método directo implica la suma simple de las puntuaciones. El método de dos pasos se realiza de la siguiente manera: Si puntúa ≥ 2 en el ítem 1 o en el ítem 2, se continua con el instrumento y se suman los puntajes obtenidos, si no, no se continua con el instrumento y se suman los puntajes de los ítems 1 y 2.

Una observación superficial del gráfico 7 evidencia que el PHQ-ALG y el PHQ-9 tienen un poder diagnóstico similar, e inclusive emplean el mismo punto de corte. Sin embargo, el PHQ-ALG presenta una menor sensibilidad (TPR=.70) que el PHQ-9 (TPR=.86). Por otra parte, el PHQ-ALG presenta una mayor especificidad (TNR=.86) que el PHQ-9 (TNR=.77). Otra forma de verlo es que el PHQ-ALG arroja más falsos negativos, mientras que el PHQ-9 más falsos positivos.

Esto probablemente se deba a que el resto de reactivos del PHQ-9 hacen referencia a síntomas que también pueden aparecer debido a los tratamientos oncológicos de los pacientes. Por lo tanto, es lógico que la aplicación de todos los ítems del PHQ-9 arroje más falsos positivos. El PHQ-ALG primero evalúa la presencia de anhedonia y de distimia antes de continuar con el resto de los ítems, por lo tanto, es lógico que se evidencie una mayor especificidad, pero menor sensibilidad.

Por lo tanto, el PHQ-9 y el PHQ-ALG presentan una adecuada precisión diagnóstica; sin embargo, considerando que el objetivo de esta investigación es validar la interpretación de las puntuaciones para tamizaje de depresión, se favorece la aplicación de todos los ítems del PHQ-9 por sobre la aplicación en dos pasos.

Con respecto al octavo objetivo específico de contrastar el poder diagnóstico de la versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9) y su versión abreviada (PHQ-2), se emplearon múltiples indicadores del rendimiento del instrumento.

El PHQ-2 con un punto de corte de ≥ 3 presenta una sensibilidad (TPR=.703) y especificidad (TNR=.845) similar al PHQ-ALG. Esto tiene sentido ya que ambos instrumentos dan prioridad a los ítems 1 y 2. Por lo tanto, la evidencia sugiere que es recomendable aplicar todos los ítems del PHQ-9 para realizar tamizaje de depresión, sin embargo, en caso se desee contar con una evaluación rápida, se puede emplear el PHQ-2 o PHQ-ALG, pero se debe tener en mente la mayor tasa de falsos negativos.

Finalmente, en cuanto al objetivo general, la evidencia de validez de las puntuaciones del PHQ-9, se deben de considerar las evidencias de validez halladas y evaluar si proveen justificación suficiente para usar el instrumento para realizar tamizaje de depresión en pacientes oncológicos del INEN.

Mediante el criterio de jueces expertos, el contenido del instrumento provee una adecuada evidencia de su validez para realizar tamizaje de depresión en pacientes oncológicos.

Similarmente, la estructura interna del instrumento hallada en esta investigación refleja un modelo bidimensional que es similar a lo teóricamente esperado y hallado en otras investigaciones. Por otra parte, la confiabilidad de la puntuación total y de la dimensión Cognitiva/Afectiva fue apropiada, pero la confiabilidad de la dimensión somática fue limitada, lo cual sugiere que mide con cierto grado de error los síntomas somáticos de la depresión. Y el instrumento demostró invarianza de medida según sexo y grado de instrucción, pero demuestra una estructura interna diferente según edad y estado civil, e interceptos diferentes según lugar de aplicación. Por lo tanto, la estructura interna del PHQ-9 provee evidencia mixta de su validez para realizar tamizaje de depresión. Se puede afirmar que justifica su uso, pero que la dimensión somática ha de interpretarse con cautela, y se debe de considerar que puede haber cierto grado de sesgo en la medición.

Y las correlaciones del PHQ-9 y sus dimensiones con el HADS y sus dimensiones sugieren cierta ambigüedad en cuanto al constructo objetivo. Puede que sea más próximo a “malestar emocional” que a “depresión”. Sin embargo, se puede argumentar que, dado que el instrumento tiene como objetivo realizar tamizaje, los resultados no serían tan perjudiciales para la validez del instrumento, como sí lo sería si el objetivo fuese realizar diagnósticos. En ese sentido, los instrumentos que se emplean para medir el malestar emocional también sirven para realizar tamizaje de depresión. Por lo tanto, este caso específico, la ambigüedad del constructo no sería una desventaja significativa. Además, el propósito de un tamizaje es identificar a los pacientes que requerirían apoyo psicológico, y posterior evaluación. Esto es algo que el PHQ-9 podría realizar inclusive si mide malestar emocional en general. Por otra parte, se evidenció validez convergente cuando se

consideraban solo los pacientes hospitalizados. Por lo tanto, la evidencia de validez convergente apoya el uso del PHQ-9 para realizar tamizaje de depresión, pero se debe de tener en mente que también podría estar midiendo el malestar emocional producto de la situación de estrés del paciente.

Y el poder diagnóstico del PHQ-9 sugiere que es un instrumento apropiado para realizar tamizaje, con elevada sensibilidad y especificidad. Sin embargo, se debe de considerar que al usar el HADS como referente diagnóstico, estos puntos de corte y los indicadores de poder diagnóstico son referenciales.

En conclusión, en cuanto al objetivo general, al evaluar la evidencia de validez de las puntuaciones del PHQ-9, se puede afirmar que el instrumento es válido para realizar tamizaje de depresión en pacientes oncológicos del INEN; sin embargo, los resultados han de interpretarse con cautela, y no debe de confundirse con un instrumento diagnóstico.

CAPÍTULO VI: CONCLUSIÓN

Tras el análisis de los datos se pudo llegar a las siguientes conclusiones sobre la evidencia de validez de las puntuaciones de la versión peruana del cuestionario sobre la salud del paciente (PHQ-9) para tamizaje de depresión en pacientes adultos con cáncer de un hospital público especializado de lima metropolitana:

Con respecto a la evidencia de validez basada en el contenido:

- Mediante el criterio de jueces, todos los ítems son considerados apropiados para medir la depresión. Sin embargo, no hay acuerdo en cuán apropiados son, es decir, si son moderadamente apropiados, buenos o excelentes.

Con respecto a la evidencia de validez basada en la estructura interna:

- Mediante el Análisis Factorial Exploratorio y Confirmatorio se evidencia que las puntuaciones del PHQ-9 presentan una adecuada bondad de ajuste a un modelo bidimensional simple. La primera dimensión mediría los síntomas cognitivos y afectivos, y la segunda dimensión mediría los síntomas somáticos. Asimismo, la evidencia demuestra que también hay una adecuada bondad de ajuste a un modelo bidimensional jerárquico, por lo que es válido interpretar la puntuación total del PHQ-9.
- Mediante análisis de la confiabilidad tau-equivalente alfa de Cronbach y confiabilidad congénica mediante el coeficiente

omega de McDonald se evidencia que la primera dimensión es confiable para medir los síntomas cognitivos y afectivos de la depresión; sin embargo, la segunda dimensión presenta resultados sub-óptimos, por lo que debe ser interpretada con cautela, especialmente en pacientes que estén cursando con elevado malestar somático debido a la enfermedad o tratamiento oncológico. Asimismo, el puntaje total del PHQ-9 es confiable para medir la sintomatología depresiva.

- Mediante el análisis de la invarianza de medida entre los grupos se evidenció que el instrumento no presenta sesgo según el Sexo y el Grado de instrucción. Pero puede haber cierto sesgo según la Edad, Estado civil y lugar de aplicación. El instrumento igual es apropiado para realizar tamizaje de depresión en la población oncológica; sin embargo, se debe proceder con cautela si es que se desea comparar las puntuaciones entre grupos.

Con respecto a la evidencia convergente de validez:

- Las correlaciones entre el PHQ-9 y sus dimensiones con el HADS y sus dimensiones, y con el autoreporte de diagnóstico depresión, brindan evidencia de que el PHQ-9 es válido para realizar tamizaje de depresión; sin embargo, las puntuaciones han de interpretarse con cautela cuando los pacientes experimentan síntomas somáticos, pues el instrumento podría medir el malestar emocional en general.

Con respecto a los puntos de corte y el poder diagnóstico del PHQ-9 y sus otros métodos de calificación:

- Se evidencia que, para realizar tamizaje de depresión en pacientes adultos con cáncer, el punto de corte más adecuado del PHQ-9 es ≥ 8 .
- Se evidencia que el PHQ-9 puede ser interpretado ya sea sumando directamente las puntuaciones, o empleando un método de aplicación de dos pasos (Primero, solo se aplican los primeros dos ítems. Si puntúa 2 o más en cualquiera de ellos, se continua con el resto del instrumento). Ambos métodos emplean el mismo punto de corte de ≥ 8 y presentan un poder diagnóstico adecuado y similar. Sin embargo, el método directo presenta una mayor sensibilidad, por lo que es más apropiado para realizar tamizaje de depresión.
- Se evidencia que el mejor punto de corte del PHQ-2 es ≥ 3 demostrando un adecuado poder diagnóstico. Sin embargo, el PHQ-9 presenta una mayor sensibilidad, por lo que es un mejor instrumento para tamizaje de depresión que su versión abreviada.

RECOMENDACIONES

- Se sugiere evaluar la bondad de ajuste del PHQ-9 a un modelo bifactorial.
- Se sugiere analizar las características de la invarianza de medida del PHQ-9 en la población oncológica empleando regresiones logísticas ordinales u otros métodos que permitan identificar cuan diferencialmente están midiendo los items la depresión.
- Se sugiere emplear entrevistas diagnosticas como el criterio *gold standard* para identificar depresión
- La implementación del PHQ-9 en algoritmos de tamizaje, diagnóstico y tratamiento de pacientes oncológicos.
- Capacitar al personal de salud no-especializado en salud mental en el uso del PHQ-9 como una guía útil y sencilla para la identificación de personas con depresión probable, para que puedan ser referidas a psicología, o si lo amerita, realizar interconsultas.
- Emplear instrumentos cuyos puntajes cuenten con evidencia de validez en la población con la cual se desea trabajar.
- Se sugiere continuar los estudios psicométricos de los instrumentos para tamizaje de depresión en población oncológica de tal forma de evidenciar la validez, o la ausencia de validez, de los instrumentos.
- Se recomienda el desarrollo de nuevos instrumentos breves de tamizaje de depresión adecuados para población oncológica, la cual tiene características particulares que podrían sesgar los puntajes de instrumentos pensados para la población general o clínica.

REFERENCIAS BIBLIOGRÁFICAS

- Agreda-Carrillo, E. R., Galindo-Morales, M. L., Larios-Falcón, K. E., Arévalo-Flores, J. M., y Cruzado, L. (2017). Conocimientos, actitudes y prácticas sobre depresión mayor en médicos residentes de un hospital de tercer nivel de atención en Lima, Perú. *Revista de neuro-psiquiatria*, 80(2), 94.
doi:10.20453/rmp.v80i2.3090
- Alford, B. A., y Beck, A. T. (2009). *Depression: Causes and Treatment* (2nd ed.). Filadelfia: University of Pennsylvania Press.
- American Educational Research Association, American Psychological Association, y National Council on Measurement Education. (2014). *Standards for Educational and Psychological Testing*. Washington DC: American Educational Research Association.
- American Psychiatric Association. (1995). Trastornos del estado de ánimo. In P. Pichot, J. J. López-Ibor, y M. Valdés (eds.), T. de Flores i Formenti, J. Toro Trallero, J. Masana Ronquillo, J. Treserra Torres, y C. Udina Abelló (trans.), *Manual Diagnóstico y Estadístico de los Trastornos Mentales DSM-IV-TR* (pp. 323–400). Barcelona: Masson.
- American Psychiatric Association. (2014). *Guía de consulta de los criterios diagnósticos del DSM-5*. American Psychiatric Pub.
- Andersen, B. L., DeRubeis, R. J., Berman, B. S., Gruman, J., Champion, V. L., Massie, M. J., ... American Society of Clinical Oncology. (2014). Screening, assessment, and care of anxiety and depressive symptoms in adults with cancer: an American Society of Clinical Oncology guideline

adaptation. *Journal of Clinical Oncology*, 32(15), 1605–1619.

doi:10.1200/JCO.2013.52.4611

Baader, T., Molina, J. L., Venezian, S., Rojas, C., Farías, R., Fierro-Freixenet, C., ... Mundt, C. (2012). Validación y utilidad de la encuesta PHQ-9 (Patient Health Questionnaire) en el diagnóstico de depresión en pacientes usuarios de atención primaria en Chile. *Revista chilena de neuro-psiquiatría*, 50(1), 10–22. doi:10.4067/S0717-92272012000100002

Ban, T. A. (2014). *From Melancholia To Depression: A History Of Diagnosis And Treatment*. International Network for the History of Neuropsychopharmacology. Obtenido de:
http://inhn.org/fileadmin/previews_new/From_Melancholia_to_Depression_March_6_2014.pdf

Barber, B., Dergousoff, J., Nesbitt, M., Mitchell, N., Harris, J., O'Connell, D., ... Seikaly, H. (2015). Depression as a predictor of postoperative functional performance status (PFPS) and treatment adherence in head and neck cancer patients: a prospective study. *Journal of Otolaryngology - Head & Neck Surgery*, 44, 38. doi:10.1186/s40463-015-0092-4

Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *The American Journal of Psychiatry*, 165(8), 969–977. doi:10.1176/appi.ajp.2008.08050721

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., y Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571. doi:10.1001/archpsyc.1961.01710120031004

- Berrospi-Reyna, S., Herencia-Souza, M., y Tarazona, A. S. (2017). Prevalencia y factores asociados a la sintomatología depresiva en mujeres con cáncer de mama en un hospital público de Lima, Perú. *Acta médica peruana*, 34(2), 95–100. Obtenido de: <http://www.fondoeditorial.cmp.org.pe/revistas/index.php/AMP/article/view/316>
- Breitbart, W., Rosenfeld, B., Tobias, K., Pessin, H., Ku, G. Y., Yuan, J., y Wolchok, J. (2014). Depression, cytokines, and pancreatic cancer. *Psycho-Oncology*, 23(3), 339–345. doi:10.1002/pon.3422
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press. Obtenido de: <http://www.kharazmi-statistics.ir/Uploads/Public/book/Methodology%20in%20the%20Social%20Sciences.pdf>
- Byrne, B. M. (2013). Chapter 3: Testing for the factorial validity of a theoretical construct (First-order CFA model). In *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming* (2nd ed.). New York: Routledge. Obtenido de: https://books.google.com.pe/books?id=1dHbAAAAQBAJ&source=gbs_navlinks_s
- Cabrera-Gutierrez, L., Mendoza-Luna, E., y Obando-Castro, P. (2010). Factores asociados a los niveles de depresión situacional en pacientes oncológicas. *Revista de enfermería herediana*, 3(2), 96–103. Obtenido de: <http://bases.bireme.br/cgi->

bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src=google&base=LIL
ACS&lang=p&nextAction=lnk&exprSearch=595449&indexSearch=ID

- Calderón, M., Gálvez-Buccollini, J. A., Cueva, G., Ordoñez, C., Bromley, C., y Fiestas, F. (2012). Validación de la versión peruana del PHQ-9 para el diagnóstico de depresión. *Revista peruana de medicina experimental y salud pública*, 29(4), 578–579. doi:10.1590/S1726-46342012000400027
- Cassiani-Miranda, C. A., Vargas-Hernández, M. C., Pérez-Anibal, E., Herazo-Bustos, M. I., y Hernández-Carrillo, M. (2017). Reliability and dimensionality of PHQ-9 in screening depression symptoms among health science students in Cartagena, 2014. *Biomedica: Revista Del Instituto Nacional de Salud*, 37(0), 112–120. doi:10.7705/biomedica.v37i0.3221
- Chochinov, H. M., Wilson, K. G., Enns, M., y Lander, S. (1994). Prevalence of depression in the terminally ill: effects of diagnostic criteria and symptom threshold judgments. *The American Journal of Psychiatry*, 151(4), 537–540. doi:10.1176/ajp.151.4.537
- Cordon, J. (2015). Analizando la V de Aiken Usando el Método Score con Hojas de Cálculo. Obtenido de:
https://www.researchgate.net/publication/277555967_Analizando_la_V_de_Aiken_Usando_el_Metodo_Score_con_Hojas_de_Calculo?enrichId=rgreq-a25f959526102275b25de89a5b0d30d6-XXX&enrichSource=Y292ZXJQYWdlOzI3NzU1NTk2NztBUzoyMzU1NjY5NjkxMjY5MTJAMTQzMzE3NDk1MDQ2NQ%3D%3D&el=1_x_3&_esc=publicationCoverPdf

- Costello, A. B., y Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment Research & Evaluation*, 10(7), 1–9. Obtenido de: <https://pareonline.net/pdf/v10n7.pdf>
- Deng, L., y Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educational and psychological measurement*, 77(2), 185–203. doi:10.1177/0013164416658325
- Doi, S., Ito, M., Takebayashi, Y., Muramatsu, K., y Horikoshi, M. (2018). Factorial validity and invariance of the Patient Health Questionnaire (PHQ)-9 among clinical and non-clinical populations. *Plos One*, 13(7), e0199235. doi:10.1371/journal.pone.0199235
- Elosua, P. O. (2003). Sobre la validez de los tests. *Psicothema*, 15(2), 315–321.
- Fabara, E. E. (2013). *Valoración del PHQ-9 y GHQ-12 como herramientas de tamizaje de la depresión en el ámbito de atención primaria en el Ecuador experiencia del uso de versiones electrónicas y auto administradas en la Clínica Universitaria USFQ sede San Rafael* (Undergraduate thesis). Universidad San Francisco de Quito. Obtenido de: <http://repositorio.usfq.edu.ec/handle/23000/2794>
- Freud, S. (1917). Mourning and Melancholia. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XIV (1914-1916): On the History of the Psycho-Analytic Movement, Papers on Metapsychology and Other Works* (pp. 243–258). London: The Hogarth press.

Gil, M. D. (2019). *Validación semántica de las preguntas del PHQ-9 (Patient Health Questionnaire-9) como instrumento para el despistaje de depresión en gestantes* (Undergraduate thesis). Obtenido de:

[http://bibvirtual.upch.edu.pe:1701/primo-explore/fulldisplay?docid=upch_aleph000046289&context=L&vid=UPCH&lang=es_PE&search_scope=UPCH&adaptor=Local%20Search%20Engine&isFrbr=true&tab=all_tab&query=any,contains,Validaci%C3%B3n%20sem%C3%A1ntica%20de%20las%20preguntas%20del%20PHQ-9%20\(Patient%20Health%20Questionnaire-9\)%20como%20instrumento%20para%20el%20despistaje%20de%20depr esi%C3%B3n%20en%20gestantes&sortby=date&facet=frbrgroupid,inclue,45606605&offset=0](http://bibvirtual.upch.edu.pe:1701/primo-explore/fulldisplay?docid=upch_aleph000046289&context=L&vid=UPCH&lang=es_PE&search_scope=UPCH&adaptor=Local%20Search%20Engine&isFrbr=true&tab=all_tab&query=any,contains,Validaci%C3%B3n%20sem%C3%A1ntica%20de%20las%20preguntas%20del%20PHQ-9%20(Patient%20Health%20Questionnaire-9)%20como%20instrumento%20para%20el%20despistaje%20de%20depr esi%C3%B3n%20en%20gestantes&sortby=date&facet=frbrgroupid,inclue,45606605&offset=0)

Graham, J. M. (2006). Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability. *Educational and psychological measurement*, 66(6), 930–944. doi:10.1177/0013164406288165

Habibzadeh, F., Habibzadeh, P., y Yadollahie, M. (2016). On determining the most appropriate test cut-off value: the case of tests with continuous results. *Biochemia medica*, 26(3), 297–307. doi:10.11613/BM.2016.034

Hair, J., Black, W., Babin, B., y Anderson, R. (2014). Exploratory Factor Analysis. In *Multivariate Data Analysis: Pearson New International Edition* (7th ed., p. 89). Edinburgh Gate: Pearson.

Harding, T. W., de Arango, M. V., Baltazar, J., Climent, C. E., Ibrahim, H. H., Ladrido-Ignacio, L., ... Wig, N. N. (1980). Mental disorders in primary health care: a study of their frequency and diagnosis in four developing

countries. *Psychological Medicine*, 10(2), 231–241.

doi:10.1017/s0033291700043993

Hartung, T. J., Friedrich, M., Johansen, C., Wittchen, H.-U., Faller, H., Koch, U.,

... Mehnert, A. (2017). The Hospital Anxiety and Depression Scale

(HADS) and the 9-item Patient Health Questionnaire (PHQ-9) as

screening instruments for depression in patients with cancer. *Cancer*,

123(21), 4236–4243. doi:10.1002/cncr.30846

Hernández, R., Fernández, C., y Baptista, P. (2010). *Metodología de la*

investigación (5ta ed.). México: McGraw-Hill.

Hidalgo, M. D., Galindo, F., Inglés, C. J., Campoy, G., y Ortiz, B. (1999). A

Study of Differential Item Functioning in a Social Skills Scale for

Adolescents. *Anales de Psicología*, 15(2), 331–343. Obtenido de:

<http://revistas.um.es/analesps/article/view/30271>

Holland, J. C., Deshields, T. L., Andersen, B., Braun, I., Breitbart, W. S., Brewer,

B. W., ... McMillian, N. (2018). *Distress management*. National

Comprehensive Cancer Network. Obtenido de:

https://www.nccn.org/professionals/physician_gls/pdf/distress.pdf

Howell, D., Keshavarz, H., Esplen, M. J., Hack, T., Hamel, M., Howes, J., ... Ali,

M. U. (2015). *Pan-Canadian Practice Guideline: Screening, Assessment*

and Management of Psychosocial Distress, Major Depression and Anxiety

in Adults with Cancer. Canadian Association of Psychosocial Oncology.

Obtenido de: [https://www.capo.ca/wp-](https://www.capo.ca/wp-content/uploads/2010/10/Distress_guideline_CAPO_201507311.pdf)

[content/uploads/2010/10/Distress_guideline_CAPO_201507311.pdf](https://www.capo.ca/wp-content/uploads/2010/10/Distress_guideline_CAPO_201507311.pdf)

- Inoue, T., Tanaka, T., Nakagawa, S., Nakato, Y., Kameyama, R., Boku, S., ...
Koyama, T. (2012). Utility and limitations of PHQ-9 in a clinic
specializing in psychiatric care. *BMC Psychiatry*, 12, 73.
doi:10.1186/1471-244X-12-73
- Instituto Nacional de Salud Mental. (2009). Estudio Epidemiológico de Salud
Mental en la Sierra Rural 2008. Informe General . *Anales de Salud Mental*,
25(1-2). Obtenido de:
[http://www.insm.gob.pe/investigacion/archivos/estudios/2008-ASM-
EESM-SR/files/res/downloads/book.pdf](http://www.insm.gob.pe/investigacion/archivos/estudios/2008-ASM-EESM-SR/files/res/downloads/book.pdf)
- Instituto Nacional de Salud Mental. (2013). Estudio Epidemiológico de Salud
Mental en Lima metropolitana y Callao - Replicación 2012. Informe
General. *Anales de Salud Mental*, 29(Supl. 1). Obtenido de:
[http://www.insm.gob.pe/investigacion/archivos/estudios/2012 ASM -
EESM -LM.pdf](http://www.insm.gob.pe/investigacion/archivos/estudios/2012 ASM - EESM -LM.pdf)
- Kline, R. B. (2010). *Principles And Practice Of Structural Equation Modeling*,
Third Edition (methodology In The Social Sciences) (3rd ed., p. 427). New
York: The Guilford Press.
- Kohn, R., Levav, I., Almeida, J. M. C. de, Vicente, B., Andrade, L., Caraveo-
Anduaga, J. J., ... Saraceno, B. (2005). Los trastornos mentales en
América Latina y el Caribe: asunto prioritario para la salud pública.
*Revista panamericana de salud publica = Pan American journal of public
health*, 18(4-5), 229–240. doi:10.1590/S1020-49892005000900002
- Krebber, A. M. H., Buffart, L. M., Kleijn, G., Riepma, I. C., de Bree, R.,
Leemans, C. R., ... Verdonck-de Leeuw, I. M. (2014). Prevalence of

depression in cancer patients: a meta-analysis of diagnostic interviews and self-report instruments. *Psycho-Oncology*, 23(2), 121–130.

doi:10.1002/pon.3409

Kroenke, K., y Spitzer, R. L. (2002). The PHQ-9: A New Depression Diagnostic and Severity Measure. *Psychiatric annals*, 32(9), 509–515.

doi:10.3928/0048-5713-20020901-06

Kroenke, K., Spitzer, R. L., y Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613.

Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema*, 26(1), 127–135. doi:10.7334/psicothema2013.258

Leonardo Olivera, W. A., Villalobos Aguinaga, M. A., y León Jiménez, F. E. (2016). Nivel de conocimientos sobre depresión mayor en médicos de atención primaria en Lambayeque, Perú, 2014. *Revista de neuro-psiquiatria*, 79(1), 23. doi:10.20453/rnp.v79i1.2765

Levis, B., Benedetti, A., Levis, A. W., Ioannidis, J. P. A., Shrier, I., Cuijpers, P., ... Thombs, B. D. (2017). Selective Cutoff Reporting in Studies of Diagnostic Test Accuracy: A Comparison of Conventional and Individual-Patient-Data Meta-Analyses of the Patient Health Questionnaire-9 Depression Screening Tool. *American Journal of Epidemiology*, 185(10), 954–964. doi:10.1093/aje/kww191

Maass, S. W. M. C., Roorda, C., Berendsen, A. J., Verhaak, P. F. M., y de Bock, G. H. (2015). The prevalence of long-term symptoms of depression and

- anxiety after breast cancer treatment: A systematic review. *Maturitas*, 82(1), 100–108. doi:10.1016/j.maturitas.2015.04.010
- Maj, M. (2013). Clinical judgment" and the DSM-5 diagnosis of major depression. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, 12(2), 89–91. doi:10.1002/wps.20049
- Martínez, P., Andreu, Y., Galdón, M. J., Romero, R., García-Conde, A., y Llobart Fuertes, P. (2017). Distrés emocional y problemas asociados en población oncológica adulta. *Psicooncología*, 14(2-3), 217–228. doi:10.5209/PSIC.57081
- Maté, J., Hollenstein, M. F., y Gil, F. L. (2004). Insomnio, ansiedad y depresión en el paciente oncológico. *Psicooncología*, 1(2), 211–230. Obtenido de: <http://revistas.ucm.es/index.php/PSIC/article/view/17109>
- Matsunaga, M. (2010). How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *International journal of psychological research*, 3(1), 97–110. Obtenido de: <http://www.redalyc.org/articulo.oa?id=299023509007>
- Maxim, L. D., Niebo, R., y Utell, M. J. (2014). Screening tests: a review with examples. *Inhalation Toxicology*, 26(13), 811–828. doi:10.3109/08958378.2014.955932
- Merz, E. L., Malcarne, V. L., Roesch, S. C., Riley, N., y Sadler, G. R. (2011). A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among English- and Spanish-speaking Latinas. *Cultural Diversity & Ethnic Minority Psychology*, 17(3), 309–316. doi:10.1037/a0023883

- Meyer, A. (1910). The nature and conception of dementia praecox. *The Journal of Abnormal Psychology*, 5(5), 274–285. doi:10.1037/h0073298
- Ministerio de Salud, y Organización Panamericana de Salud. (2013). *Guía técnica de diagnóstico y tratamiento de depresión en adultos de 18 a 65 años para médicos no psiquiatras* (Guía técnica) (pp. 1–30). Ministerio de Salud. Obtenido de:
http://www.portal.ins.gob.pe/es/component/rsfiles/preview?path=CNSP%20FUNAGESP%20Documentos+Tecnicos%202013%207_Guia+Tecnica+de+diagnostico+y+tratamiento+de+depresion.pdf
- Montenegro, Kilstein, M. B., y Guillermo, J. (2017). Utilidad del PHQ-9 en el diagnóstico de trastorno depresivo en una sala de Clínica Médica. *IntraMed Journal*, 6(1), 1–12. Obtenido de:
http://journal.intramed.net/index.php/Intramed_Journal/article/view/560
- Montero, I., y León, O. G. (2007). A guide for naming research studies in Psychology. *International journal of clinical and health psychology : IJCHP*, 7(3), 847–862. Obtenido de:
<http://www.redalyc.org/pdf/337/33770318.pdf>
- Mundfrom, D. J., Shaw, D. G., y Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5(2), 159–168. doi:10.1207/s15327574ijt0502_4
- Organización Mundial de la Salud. (2000). *Guía de bolsillo de la clasificación CIE-10; clasificación de los trastornos mentales y del comportamiento : con glosario y criterios diagnósticos de investigación : CIE-10: CDI-10*. (P. G. Parajuá y M. M. López, trans.). Madrid: Médica Panamericana.

Obtenido de:

http://apps.who.int/iris/bitstream/handle/10665/42326/8479034920_spa.pdf;jsessionid=3E80F88097F95733EF30C7DA36F080BE.pdf

Organización Mundial de la Salud. (2017). *Depression and Other Common'*

Mental Disorders: Global Health Estimates. Geneva: WHO.

Padilla, J. L., y Benítez, I. (2014). Validity evidence based on response processes.

Psicothema, 26(1), 136–144. doi:10.7334/psicothema2013.259

Padilla, J. L., Gomez, J., Hidalgo, M. D., y Muñiz, J. (2006). La evaluación de las

consecuencias del uso de los tests en la teoría de la validez. *Psicothema*,

18(2), 307–312. Obtenido de: <http://www.psicothema.com/pdf/3215.pdf>

Paykel, E. S. (2008). Basic concepts of depression. *Dialogues in Clinical*

Neuroscience, 10(3), 279–289.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for

research in the general population. *Applied psychological measurement*,

1(3), 385–401. doi:10.1177/014662167700100306

Rios, J., y Wells, C. (2014). Validity evidence based on internal structure.

Psicothema, 26(1), 108–116. doi:10.7334/psicothema2013.260

Ruiz-Grosso, P., Loret de Mola, C., Vega-Dienstmaier, J. M., Arevalo, J. M.,

Chavez, K., Vilela, A., ... Huapaya, J. (2012). Validation of the Spanish

Center for Epidemiological Studies Depression and Zung Self-Rating

Depression Scales: a comparative validation study. *Plos One*, 7(10),

e45413. doi:10.1371/journal.pone.0045413

Saldivia, S., Aslan, J., Cova, F., Vicente, B., Inostroza, C., y Rincón, P. (2019).

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

Revista Medica de Chile, 147(1), 53–60. doi:10.4067/S0034-98872019000100053

Sireci, S., y Faulkner-Bond, M. (2014). Validity evidence based on test content.

Psicothema, 26(1), 100–107. doi:10.7334/psicothema2013.256

Sireci, S., y Padilla, J.-L. (2014). Validating assessments: Introduction to the Special Section. *Psicothema*, 26(1), 97–99.

doi:10.7334/psicothema2013.255

Sotelo, J. L., Musselman, D., y Nemeroff, C. (2014). The biology of depression in cancer and the relationship between depression and cancer progression.

International Review of Psychiatry, 26(1), 16–30.

doi:10.3109/09540261.2013.875891

Spitzer, R. L., Kroenke, K., y Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *The Journal of the American Medical Association*, 282(18), 1737–1744.

doi:10.1001/jama.282.18.1737

Suzuki, K., Kumei, S., Ohhira, M., Nozu, T., y Okumura, T. (2015). Screening for major depressive disorder with the Patient Health Questionnaire (PHQ-9 and PHQ-2) in an outpatient clinic staffed by primary care physicians in Japan: a case control study. *Plos One*, 10(3), e0119147.

doi:10.1371/journal.pone.0119147

Tateishi, V. G. (2012, January 24). *Estilos de humor y sintomatología depresiva en pacientes con cáncer de mama* (Undergraduate thesis). Obtenido de:

<http://tesis.pucp.edu.pe/repositorio/handle/123456789/1182>

- Tavakol, M., y Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, 2, 53–55.
doi:10.5116/ijme.4dfb.8dfd
- Thombs, B. D., Ziegelstein, R. C., y Whooley, M. A. (2008). Optimizing detection of major depression among patients with coronary artery disease using the patient health questionnaire: data from the heart and soul study. *Journal of General Internal Medicine*, 23(12), 2014–2017.
doi:10.1007/s11606-008-0802-y
- Treiblmaier, H., y Filzmoser, P. (2010). Exploratory factor analysis revisited: How robust methods support the detection of hidden multivariate data structures in IS research. *Information & Management*, 47(4), 197–207.
doi:10.1016/j.im.2010.02.002
- Vega-Dienstmaier, J., Coronado-Molina, Ó., y Mazzotti, G. (2014). Validez de una versión en español del Inventario de Depresión de Beck en pacientes hospitalizados de medicina general. *Revista de neuro-psiquiatria*, 77(2), 95. doi:10.20453/rnp.v77i2.1151
- Veiga de Cabo, J., Fuente-Díez, E. de la, y Zimmermann-Verdejo, M. (2008). Modelos de estudios en investigación aplicada: conceptos y criterios para el diseño. *Medicina y Seguridad del Trabajo*, 54(210), 81–88. Obtenido de: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0465-546X2008000100011&lng=es&tlng=es.
- Véliz, J. (1998). Validez y confiabilidad del cuestionario de autoreporte en personal de una institución militarizada. *Anales de salud mental*, 14(An.

salud ment.), 85–102. Obtenido de:

http://repebis.upch.edu.pe/articulos/ansm/v14n1_2/a6.pdf

Ventura-Leon, J., y Caycho-Rodríguez, T. (2017). El coeficiente Omega: un método alternativo para la estimación de la confiabilidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 15(1), 625–627.

Obtenido de:

<http://www.redalyc.org/jatsRepo/773/77349627039/index.html>

Vilela-Estrada, A. L. (2019). *Validación de Escala de Ansiedad y Depresión Hospitalaria (HADS) en pacientes adultos oncológicos en un hospital público especializado de Lima metropolitana* (Undergraduate thesis). Universidad Cesar Vallejo.

Walker, J., Holm Hansen, C., Martin, P., Sawhney, A., Thekkumpurath, P., Beale, C., ... Sharpe, M. (2013). Prevalence of depression in adults with cancer: a systematic review. *Annals of Oncology*, 24(4), 895–900.

doi:10.1093/annonc/mds575

Watts, S., Leydon, G., Birch, B., Prescott, P., Lai, L., Eardley, S., y Lewith, G. (2014). Depression and anxiety in prostate cancer: a systematic review and meta-analysis of prevalence rates. *BMJ Open*, 4(3), e003901.

doi:10.1136/bmjopen-2013-003901



Watts, S., Prescott, P., Mason, J., McLeod, N., y Lewith, G. (2015). Depression and anxiety in ovarian cancer: a systematic review and meta-analysis of prevalence rates. *BMJ Open*, 5(11), e007618. doi:10.1136/bmjopen-2015-

007618

- Xu, H., y Tracey, T. J. G. (2017). Use of multi-group confirmatory factor analysis in examining measurement invariance in counseling psychology research. *The European Journal of Counselling Psychology*, 6(1), 75–82.
doi:10.5964/ejcop.v6i1.120
- Yang, Y., y Liang, X. (2013). Confirmatory factor analysis under violations of distributional and structural assumptions. *International Journal of Quantitative Research in Education*, 1(1), 61.
doi:10.1504/IJQRE.2013.055642
- Zhong, Q., Gelaye, B., Fann, J. R., Sanchez, S. E., y Williams, M. A. (2014). Cross-cultural validity of the Spanish version of PHQ-9 among pregnant Peruvian women: a Rasch item response theory analysis. *Journal of Affective Disorders*, 158, 148–153. doi:10.1016/j.jad.2014.02.012
- Zigmond, A. S., y Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, 67(6), 361–370. doi:10.1111/j.1600-0447.1983.tb09716.x
- Zung, W. W. (1965). A SELF-RATING DEPRESSION SCALE. *Archives of General Psychiatry*, 12, 63–70.
doi:10.1001/archpsyc.1965.01720310065008

ANEXOS

Anexo 1. Carta de aprobación del Comité Revisor de Protocolos de Investigación del Instituto Nacional de Enfermedades Neoplásicas (INEN)

	PERÚ	Sector Salud	Instituto Nacional de Enfermedades Neoplásicas	
---	-------------	--------------	--	---

"Año de la Lucha contra la Corrupción y la Impunidad"

Lima, 16 de enero de 2019

CARTA N° 011-2019-CRPI-DI-DICON/INEN





Sr.
DE LA FLOR VALDIVIA MARTIN ALONSO
Investigador Principal
Presente.-


De nuestra consideración:

Es grato dirigirme a usted para saludarle cordialmente e informarle que el Comité Revisor de Protocolos de Investigación del INEN, han revisado y aprobado el trabajo de Investigación Titulado: **"VALIDACIÓN DE LA VERSIÓN PERUANA DEL CUESTIONARIO SOBRE LA SALUD DEL PACIENTE (PHQ-9) PARA TAMIZAJE DE DEPRESIÓN EN PACIENTES ADULTOS CON CÁNCER DE UN HOSPITAL PÚBLICO ESPECIALIZADO DE LIMA METROPOLITANA"**. INEN 19-07.

De acuerdo con las normas deberá presentar un informe por correo electrónico a los 6 o 12 meses sobre los avances de dicho proyecto, así como las conclusiones del mismo a esta Oficina.

Atentamente,

 M.C. Vásquez Chávez Jule Presidente del Comité Revisor de Protocolos de Investigación del INEN	 M.C. Marga López Contreras Miembro del Comité Revisor de Protocolos de Investigación del INEN
 M.C. Alberto Lachos Dávila Miembro del Comité Revisor de Protocolos de Investigación del INEN	 M.C. Maria Eugenia Guillén Zeballos Miembro del Comité Revisor de Protocolos de Investigación del INEN


Mag. Carlos Barrionuevo Cornejo
Miembro Suplente del Comité Revisor de
Protocolos de Investigación del INEN

INSTITUTO NACIONAL DE ENFERMEDADES NEOPLÁSICAS
Av. Angamos Este 2520, Lima - 34 Tel.: 201-6500 Fax: 620-4991 Web: www.inen.edu.pe e-mail: postmaster@inen.edu.pe

Anexo 2. Carta de aprobación del Comité Institucional del Comité de Ética en Investigación (CIEI) del Instituto Nacional de Enfermedades Neoplásicas (INEN)



COMITÉ INSTITUCIONAL DE ÉTICA EN INVESTIGACIÓN
DECENIO DE LA IGUALDAD DE OPORTUNIDADES PARA MUJERES Y HOMBRES
"AÑO DE LA LUCHA CONTRA LA CORRUPCIÓN Y LA IMPUNIDAD"

Lima, 05 de abril de 2019

CARTA N° 131 -2019-CIEI/INEN

Señor
MARTIN DE LA FLOR VALDIVIA
Investigador Principal

Presente.-

REFERENCIA: PROTOCOLO: "VALIDACIÓN DE LA VERSIÓN PERUANA DEL CUESTIONARIO SOBRE LA SALUD DEL PACIENTE (PHQ-9) PARA TAMIZAJE DE DEPRESIÓN EN PACIENTES ADULTOS CON CÁNCER DE UN HOSPITAL PÚBLICO ESPECIALIZADO DE LIMA METROPOLITANA ". INEN 19-07

ASUNTO: REVISIÓN Y REVALUACIÓN

Mediante el presente, tengo a bien dirigirme a usted para informarle que el Comité Institucional de Ética en Investigación (CIEI) del INEN, *revisó, reevaluó y aprobó* la siguiente documentación remitida del protocolo en referencia:

- *Proyecto de Investigación*
- *Consentimiento Informado*

Así mismo deberán remitir el informe de avance semestral y final.

Sin otro particular, quedo de Ud.

Atentamente,

Dr. ARISTIDES JUVENAL SANCHEZ LINON
Presidente
Comité Institucional de Ética en Investigación
Instituto Nacional de Enfermedades Neoplásicas

c.c.: Archivo
L.C

Anexo 5. Cuestionario sociodemográfico

CUESTIONARIO SOCIO DEMOGRÁFICO

Fecha de aplicación de protocolo: ___/___/___ Fecha de nacimiento:
___/___/___

Hospitalización () Consulta externa () Quimioterapia
ambulatoria ()

INFORMACIÓN PERSONAL

- Edad:
- Sexo: () Masculino () Femenino
- Estado civil: () Casado/Conviviente () Divorciado/Separado () Soltero () Viudo
- Lugar de origen: () Perú () Otro especificar _____
- Lugar/ Distrito de residencia actual:
- Departamento de procedencia:
- Grado de instrucción: () Sin estudios () Primaria () Secundaria () Técnico () Superior
- Situación laboral actual: () No labora () Empleado () Independiente/Ama de casa

INFORMACIÓN MÉDICA

- Diagnóstico oncológico:
- Otros diagnósticos médicos:
- Antecedentes psicopatológicos:
- Recidiva de diagnóstico: Si () No ()
- Tiempo de diagnóstico oncológico:
- Estadio clínico: () I () II () III () IV () No aplica
- Tipo de tratamiento actual: () QT () RT () Cirugía () Hormonal () Inmunoterapia () Ninguno () Otros:

- Tipo de tratamiento previo: () QT () RT () Cirugía () Hormonal () Inmunoterapia () Ninguno () Otros:

- Tuvo contacto con algún psicólogo en INEN: () Si () No

Anexo 6. Versión peruana del Cuestionario sobre la Salud del Paciente (PHQ-9)

PHQ-9

En las últimas 2 semanas, ¿con qué frecuencia ha sentido molestia por cualquiera de los siguientes problemas? Marque su respuesta con una "X".

	Para nada	Varios días (1 a 6 días)	Más de la mitad de los días (7 a 11 días)	Casi todos los días (12 días a más)
1. Poco interés o placer en hacer las cosas	0	1	2	3
2. Sentirse triste, desanimado, deprimido o sin esperanza	0	1	2	3
3. Problemas para dormir o mantenerse dormido, o dormir demasiado	0	1	2	3
4. Sentirse cansado o con poca energía	0	1	2	3
5. Poco apetito o comer en exceso	0	1	2	3
6. Sentirse mal acerca de sí mismo. Sentir que es un fracasado, que se ha fallado a sí mismo o a su familia	0	1	2	3
7. Dificultad para poner atención o concentrarse en las cosas, tales como leer o ver televisión	0	1	2	3
8. ¿Moverse o hablar tan despacio que otras personas lo han notado? O lo contrario - estar tan inquieto o intranquilo que se ha estado moviendo mucho más de lo normal	0	1	2	3
9. Pensamiento de que sería mejor estar muerto o que quisiera hacerse daño de alguna forma	0	1	2	3

Si marcó algún problema, ¿cuánto le han dificultado estos problemas realizar su trabajo, encargarse de las cosas en la casa, o llevarse bien con otras personas? Marque su respuesta con una "X"

No ha sido difícil	Algo difícil	Muy difícil	Extremadamente difícil
--------------------	--------------	-------------	------------------------

Anexo 7. Escala de Ansiedad y Depresión Hospitalaria (HADS)

HADS

Este cuestionario se ha construido para saber cómo te sientes. Lee cada una de las frases y marca la respuesta que más se ajusta a cómo te sentiste durante los últimos siete (7) días.

1. Me siento tenso o nervioso			
Todos los días	Muchas veces	A veces	Nunca
2. Todavía disfruto las cosas como antes			
Muchísimo	No mucho	Un poco	Nada
3. Tengo una sensación de miedo, como si algo malo me va a suceder			
Definitivamente, y es muy fuerte	Sí, pero no es muy fuerte	Un poco, pero no me preocupa	Nada
4. Puedo reírme y ver el lado divertido de las cosas			
Al igual que siempre lo hice	No tanto ahora	Casi nunca	Nunca
5. Tengo mi mente llena de preocupaciones			
La mayoría de veces	Con bastante frecuencia	A veces, aunque no muy a menudo	Sólo en ocasiones
6. Suelo sentirme alegre			
Nunca	No muy a menudo	A veces	Casi siempre
7. Puedo estar sentado tranquilamente y sentirme relajado			
Siempre	Por lo general	No muy a menudo	Nunca
8. Me siento como si estuviera más lento para pensar y/o realizar mis actividades diarias			
Por lo general, en todo momento	Muy a menudo	A veces	Nunca
9. Tengo una sensación extraña, como hormigueo en el estómago			
Nunca	En ciertas ocasiones	Con bastante frecuencia	Muy a menudo
10. He perdido interés por mi aspecto físico y/o cuidado personal			
Totalmente	No me preocupa tanto como debería	Podría tener un poco más de cuidado	Me preocupo igual que siempre
11. Me siento tan inquieto que no puedo dejar de moverme			
Mucho	Bastante	No mucho	Nada
12. Veo las cosas con optimismo			
Igual que siempre	Menos de lo que se acostumbra	Muchos menos de lo que acostumbraba	Nada
13. Tengo sentimientos repentinos de pánico			
Muy frecuentemente	Bastante a menudo	No muy a menudo	Nada
14. Puedo disfrutar de un buen libro, la radio, un programa de televisión o alguna actividad que antes solía hacer.			
A menudo	A veces	No muy a menudo	Rara vez

Anexo 8. Documentos para validar el PHQ-9 mediante de juicio de expertos

CARTA DE PRESENTACIÓN

Señor(a) (ita):

Presente

Asunto: VALIDACIÓN DE INSTRUMENTOS A TRAVÉS DE JUICIO DE EXPERTOS.

Me es muy grato comunicarnos con usted para expresarle mi saludo y estima personal. Asimismo, hacer de su conocimiento que siendo Bachiller en Psicología de la Universidad Peruana Cayetano Heredia, se requiere validar el instrumento con el cual se recogerá la información necesaria para poder desarrollar la investigación con la cual se optará el título profesional de Licenciado en Psicología.

El título del proyecto de investigación es: *“Validación de la versión peruana del Cuestionario sobre la salud del paciente (PHQ-9) para tamizaje de depresión en pacientes adultos con cáncer de un hospital público especializado de Lima metropolitana”* y siendo imprescindible contar con la aprobación de personal especializado para poder aplicar el instrumento en mención, se ha considerado conveniente recurrir a usted, ante su connotada experiencia en temas de salud e investigación psicológica.

El expediente de validación, que le hacemos llegar contiene:

- Carta de presentación.
- Definiciones conceptuales de la variable.
- Matriz de operacionalización de la variable.
- Certificado de validez de contenido del instrumento.

Expresándole nuestros sentimientos de respeto y consideración nos despedimos de usted, no sin antes agradecerle por la atención que dispense a la presente.

Atentamente,

Firma
Martín Alonso de la Flor Valdivia
D.N.I: ██████████

DEFINICIÓN CONCEPTUAL DE LAS VARIABLES Y DIMENSIONES

VARIABLE UNIDIMENSIONAL DEPRESIÓN:

La depresión es un trastorno mental caracterizado por un estado de ánimo deprimido o una pérdida de interés o placer en casi todas las actividades, cambios de apetito o peso, del sueño y de la actividad psicomotora; falta de energía; sentimientos de infravaloración o culpa; dificultad para pensar, concentrarse o tomar decisiones, y pensamientos recurrentes de muerte o ideación, planes o intentos suicidas, acompañado de un malestar clínico significativo o de deterioro social, laboral o de otras áreas importantes de la actividad del individuo (APA, 1995, p.326).

MATRIZ DE OPERACIONALIZACIÓN

Dimensiones	Indicadores	ítems	Niveles o rangos
Depresión	Indican la severidad de la sintomatología depresiva	1, 2, 3, 4, 5, 6, 7, 8, 9	0 – 9: Sin depresión 10 – 27: Depresión

CUESTIONARIO SOBRE LA SALUD DEL PACIENTE – 9 (PHQ-9)

En las últimas 2 semanas, ¿con qué frecuencia ha sentido molestia por cualquiera de los siguientes problemas? Marque su respuesta con una "X".

	Para nada	Varios días (1 a 6 días)	Más de la mitad de los días (7 a 11 días)	Casi todos los días (12 días a más)
1. Poco interés o placer en hacer las cosas	0	1	2	3
2. Sentirse triste, desanimado, deprimido o sin esperanza	0	1	2	3
3. Problemas para dormir o mantenerse dormido, o dormir demasiado	0	1	2	3
4. Sentirse cansado o con poca energía	0	1	2	3
5. Poco apetito o comer en exceso	0	1	2	3
6. Sentirse mal acerca de sí mismo - o sentir que es un fracasado, o que se ha fallado a sí mismo o a su familia	0	1	2	3
7. Dificultad para poner atención o concentrarse en las cosas, tales como leer o ver televisión	0	1	2	3
8. ¿Moverse o hablar tan despacio que otras personas lo pueden haber notado? O lo contrario - estar tan inquieto o intranquilo que se ha estado moviendo mucho más de lo normal	0	1	2	3
9. Pensamiento de que sería mejor estar muerto o que quisiera hacerse daño de alguna forma	0	1	2	3

Si marcó algún problema, ¿cuánto le han dificultado estos problemas realizar su trabajo, encargarse de las cosas en la casa, o llevarse bien con otras personas? Marque su respuesta con una "X"

No ha sido difícil	Algo difícil	Muy difícil	Extremadamente difícil
--------------------	--------------	-------------	------------------------

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO

Por favor, puntúe la Representación, Relevancia y Claridad del contenido de los ítems del 1 al 5. A mayor puntuación, mayor representación, relevancia u claridad de los ítems.

- Representación: El grado en el cual cada ítem representa y mide la depresión.
- Relevancia: El grado en el cual cada ítem es relevante a la medición de la depresión.
- Claridad: El grado en el cual cada ítem es claro y fácil de entender

	Representación	Relevancia	Claridad
1. Poco interés o placer en hacer las cosas	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
2. Sentirse triste, desanimado, deprimido o sin esperanza	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
3. Problemas para dormir o mantenerse dormido, o dormir demasiado	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
4. Sentirse cansado o con poca energía	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
5. Poco apetito o comer en exceso	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
6. Sentirse mal acerca de sí mismo - o sentir que es un fracasado o que se ha fallado a sí mismo o a su familia	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
7. Dificultad para poner atención o concentrarse en las cosas, tales como leer o ver televisión	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
8. ¿Moverse o hablar tan despacio que otras personas lo pueden haber notado? O lo contrario - estar tan inquieto o intranquilo que se ha estado moviendo mucho más de los normal	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
9. Pensamiento de que sería mejor estar muerto o que quisiera hacerse daño de alguna forma	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5

Observaciones:

Opinión de aplicabilidad: Aplicable [] Aplicable después de corregir [] No aplicable []

Apellidos y nombres del juez validador:

DNI:.....

.....de del 2020

Anexo 9. Resumen de comparación de medias del PHQ-2, PHQ-9 y sus dimensiones, y PHQ-ALG

		N	Media (DE)				
			PHQ-2	PHQ-9 F1	PHQ-9 F2	PHQ-9	PHQ-ALG
Sexo	Hombre	233	1.2 (1.3)	2.9 (2.9)	3 (2.5)	5.2 (4.0)	3.9 (5.6)
	Mujer	218	1.7 (1.6)	2.1 (2.2)	3.1 (2.6)	6.2 (4.9)	2.5 (4.1)
Diagnóstico	CA MAMA	63	1.63 (1.71)	2.92 (2.97)	3.22 (2.5)	6.19 (4.8)	3.89 (5.45)
	LLA	51	1.16 (1.02)	2.08 (1.89)	2.96 (2.14)	5.14 (3.52)	2.14 (3.77)
	CA ESTOMAGO	24	0.83 (0.92)	2 (2.15)	2.88 (2.33)	4.88 (3.75)	1.88 (3.81)
	CA COLON	22	1.19 (1.33)	2.18 (2.48)	2.5 (2.52)	4.76 (4.23)	1.76 (3.03)
	CA OVARIO	22	1.82 (1.37)	3 (2.09)	3 (2.37)	6 (3.39)	4.05 (4.35)
	CA TESTICULAR	19	1.26 (1.1)	1.74 (1.45)	2.53 (2.12)	4.26 (2.45)	1.68 (2.19)
	CA CERVIX	15	2.13 (1.85)	3.47 (3.11)	4.47 (3.42)	7.93 (6.03)	6.53 (6.89)
	CA TIROIDES	13	3 (2.24)	5.62 (4.46)	3.85 (3.11)	9.46 (6.96)	7.62 (8.26)
	LA	13	1 (1.08)	1.62 (1.5)	2 (1.83)	3.62 (2.96)	1.46 (2.44)
	TCG	13	1.15 (1.14)	1.85 (1.95)	2.92 (2.72)	4.77 (4.04)	2.46 (4.1)
	LMA	12	0.67 (0.65)	1.33 (1.23)	2.33 (1.72)	3.67 (2.67)	0.67 (0.65)
	CA TEJIDO BLANDO	11	1.09 (1.14)	2.45 (2.88)	3.18 (2.86)	5.64 (5.12)	2.45 (5.52)
	CA HUESO	9	1.33 (1.41)	2.22 (2.82)	2.11 (1.9)	4.33 (4.3)	2.56 (4.77)
	CA UTERO	9	1.67 (1.22)	3.11 (2.67)	3.44 (2.51)	6.56 (4.33)	4.11 (4.86)
	LPA	8	0.75 (0.71)	1.63 (1.92)	3.88 (2.53)	5.5 (4.14)	0.75 (0.71)
	CA PROSTATA	7	1.17 (2.4)	2.71 (3.86)	3.86 (2.67)	6.67 (5.82)	3 (6.87)
	CA CEREBRO	6	2.8 (2.59)	3.5 (4.51)	3.83 (5)	8.8 (9.76)	8.2 (10.28)
	CA CUELLO UTERINO	6	2.83 (2.48)	5.17 (4.54)	3.17 (0.75)	8.33 (5.01)	5.67 (6.86)
	CA RECTO	5	1.8 (1.64)	3.4 (3.65)	3.6 (1.95)	7 (4.53)	5 (5.96)
	CA RENAL	5	1.6 (1.14)	2 (1.22)	2.6 (2.61)	4.6 (3.58)	2.8 (3.56)
LH	5	1.6 (1.82)	2.6 (2.3)	3.4 (2.51)	6 (3.54)	3.2 (4.09)	
LNH	5	1.2 (1.3)	2 (1.58)	5 (5.1)	7 (4.85)	4 (5.61)	
Grado de instrucción	Sin estudios	14	1.07 (0.92)	2.07 (1.44)	1.79 (2.12)	3.86 (2.28)	1.43 (1.83)
	Primaria completa	61	1.87 (1.79)	3.34 (3.13)	3.72 (3.02)	7.07 (5.46)	5.18 (6.39)
	Secundaria completa	215	1.4 (1.46)	2.48 (2.53)	2.97 (2.33)	5.45 (4.22)	2.92 (4.59)
	Técnica completa	55	1.53 (1.6)	2.58 (2.8)	3.13 (2.6)	5.71 (4.43)	3.04 (4.67)
	Superior completa	105	1.42 (1.52)	2.44 (2.68)	3.26 (2.66)	5.7 (4.67)	3.13 (5.05)
Situación laboral	Desempleado	309	1.52 (1.5)	2.73 (2.71)	3.09 (2.47)	5.82 (4.47)	3.29 (4.93)
	Empleado	38	1.18 (1.41)	1.92 (2.02)	3 (2.9)	4.92 (4.21)	2.34 (4.27)
	Independiente/Ama del hogar	103	1.44 (1.64)	2.42 (2.72)	3.25 (2.67)	5.67 (4.8)	3.44 (5.35)
Fase de Tto	Pre-Tratamiento	78	1.94 (1.72)	3.4 (3.09)	3.13 (3.03)	6.65 (5.49)	4.78 (6.13)
	En Tratamiento	296	1.39 (1.46)	2.41 (2.47)	3.09 (2.4)	5.56 (4.2)	3 (4.65)
	Post-Tratamiento	60	1.23 (1.45)	2.16 (2.79)	3.03 (2.47)	5.27 (4.37)	2.35 (4.43)
Estado Civil	Casado/Conviviente	222	1.61 (1.6)	2.78 (2.87)	3.09 (2.66)	5.87 (4.82)	3.77 (5.35)
	Divorciado/Separado	44	1.48 (1.62)	2.52 (2.47)	3.66 (3.07)	6.18 (4.89)	3.75 (5.49)
	Soltero	166	1.25 (1.35)	2.21 (2.26)	2.96 (2.27)	5.16 (3.86)	2.32 (4)
	Viudo	18	1.89 (1.78)	3.94 (3.46)	3.72 (2.16)	7.67 (4.98)	4.06 (6.12)
Edad	18 a 35 años	186	1.44 (1.41)	2.51 (2.52)	3.13 (2.46)	5.63 (4.41)	2.95 (4.79)
	36 a 50 años	116	1.63 (1.68)	2.91 (2.92)	3 (2.4)	5.91 (4.62)	3.91 (5.31)
	51 a 65 años	118	1.36 (1.51)	2.34 (2.59)	3.34 (2.82)	5.68 (4.57)	3.21 (5.01)
	66 años a más	29	1.45 (1.74)	2.66 (2.87)	2.59 (2.57)	5.24 (4.79)	2.59 (4.76)

Anexo 10. Resumen de poder diagnóstico del PHQ-9, PHQ-ALG, y el PHQ-2

Gráfico 8. Poder diagnóstico de los puntos de corte del PHQ-9

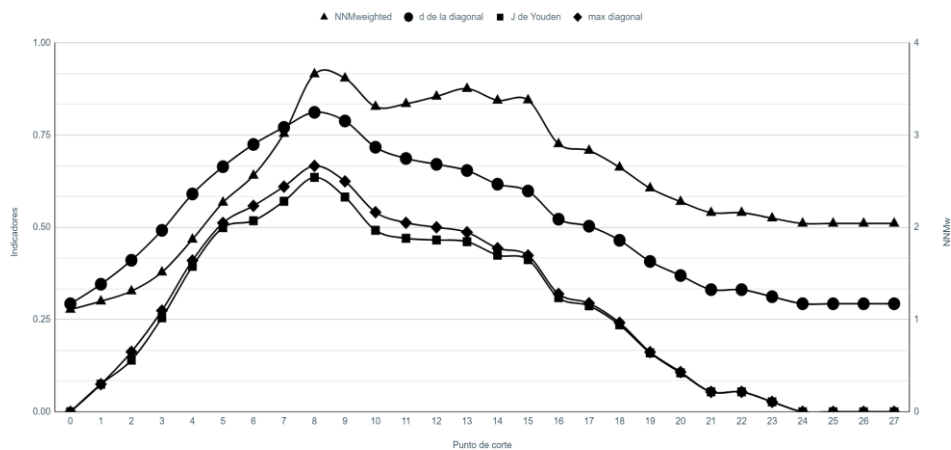


Gráfico 9. Poder diagnóstico de los puntos de corte del PHQ-ALG

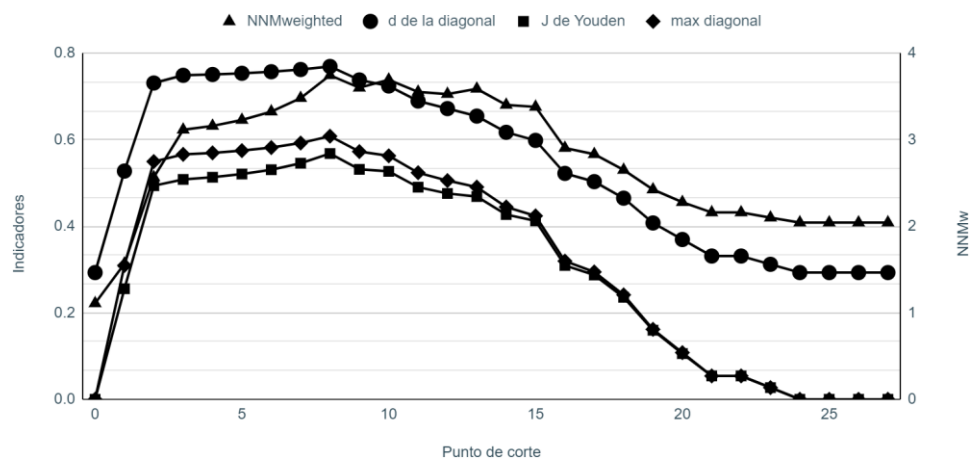
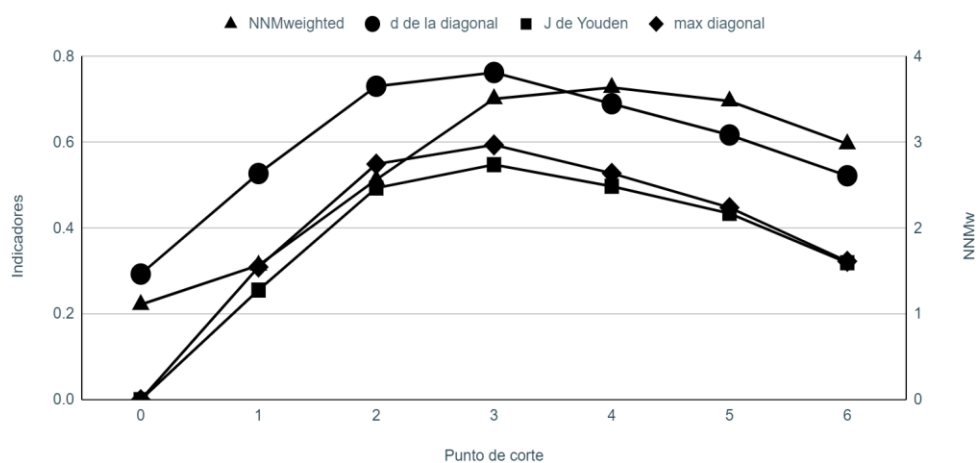


Gráfico 10. Poder diagnóstico de los puntos de corte del PHQ-2



Anexo 11. Análisis Factorial Exploratorio del PHQ-9 según grupo de edad

	18 a 36 años		36 a 50 años		51 a 65 años		66 ≤ años	
	F1	F2	F1	F2	F1	F2	F1	F2
Ítem 1: Anhedonia	0.539	.	0.737	.	0.684	.	0.697	.
Ítem 2: Distimia	0.58	.	.	0.443	0.462	.	0.733	.
Ítem 3: Sueño	.	0.415	0.498	.	.	0.6	0.335	1.069
Ítem 4: Cansancio	.	0.485	0.837	.	.	0.41	0.499	.
Ítem 5: Apetito	.	0.822	.	.	.	0.499	0.669	.
Ítem 6: Culpa	0.98	.	.	0.725	0.699	.	.	.
Ítem 7: Desconcentración	.	.	.	0.393	0.403	.	0.541	.
Ítem 8: Psicomotricidad	0.389	0.56	0.65	.
Ítem 9: Ideación suicida	0.452	.	.	0.808	0.533	.	.	.

Nota: Método de extracción Factorización de ejes principales (FEP), con rotación promax con normalización Kaiser. Se obviaron las cargas factoriales insignificantes ($\lambda < 0.2$)