

Perbandingan Akurasi Diagnostik *PHQ-2* dan *PHQ-9* sebagai Instrumen Skrining Depresi di Layanan Primer: *Systematic Literature Review*

Moch. Tabriz Azenta^{*}, Burhanudin Ichsan, Shalsa Difa Permatasari, Nabila Mutia Salsabila, Nabila Humaira, Chevalerie Daffa Prawito, Natasya Pramono

Program Studi Kedokteran, Fakultas Kedokteran, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia

INFORMASI ARTIKEL

Diterima Redaksi: 04 Desember 2025
Revisi Akhir: 07 Februari 2026
Diterbitkan Online: 08 Februari 2026

KATA KUNCI

Depresi
Layanan Primer
PHQ-2
PHQ-9
Skrining

KORESPONDENSI

Phone: +62 812-4904-3224
E-mail: mochtabrizazenta@gmail.com

A B S T R A K

Deteksi dini depresi di layanan primer membutuhkan instrumen skrining yang akurat sekaligus praktis. Tinjauan sistematis ini membandingkan akurasi diagnostik *Patient Health Questionnaire-2 (PHQ-2)* dan *PHQ-9* untuk skrining depresi. Pencarian dilakukan di *Google Scholar* dan *PubMed* dengan kata kunci terkait (publikasi 2020-2025). Seleksi menggunakan alur *PRISMA* dengan kriteria inklusi studi komparatif pada dewasa di layanan primer. Lima studi (6.500 subjek) dari berbagai populasi memenuhi kriteria. *PHQ-9* menunjukkan sensitivitas lebih tinggi (0,88-0,90) dibandingkan *PHQ-2* (0,75-0,80), dengan spesifisitas masing-masing 0,85 dan 0,78. *PHQ-2* efektif untuk skrining awal dengan administrasi lebih cepat, sedangkan *PHQ-9* memberikan evaluasi komprehensif untuk diagnosis definitif. Pendekatan bertingkat mengoptimalkan deteksi dini dengan menggunakan *PHQ-2* sebagai skrining awal dilanjutkan *PHQ-9* untuk evaluasi mendalam. Disarankan menggunakan pendekatan dua tahap (*two-step screening*): *PHQ-2* untuk skrining cepat semua pasien, lalu lanjutkan *PHQ-9* pada hasil positif untuk konfirmasi.

PENDAHULUAN

Depresi merupakan salah satu gangguan kesehatan mental yang paling sering dijumpai dan berkontribusi besar terhadap penurunan kualitas hidup serta peningkatan beban sistem kesehatan di seluruh dunia (Asim dkk., 2023; Camargo dkk., 2022; Levis dkk., 2020). Deteksi dini depresi di layanan primer menjadi krusial karena fasilitas ini biasanya merupakan pintu pertama pasien berinteraksi dengan sistem Kesehatan (Mohamed Khalifa dkk., 2023; Pellas & Damberg, 2021; Sun dkk., 2020; Teusen dkk., 2022). Identifikasi yang cepat dan tepat di tahap awal memungkinkan intervensi yang lebih efektif serta mencegah berkembangnya komplikasi yang lebih berat (Adams dkk., 2022; Gómez-Gómez dkk., 2023; Kendrick dkk., 2024; Lee dkk., 2022). Namun, keterbatasan waktu konsultasi dan sumber daya di layanan primer sering kali menjadi hambatan dalam pelaksanaan skrining yang komprehensif (Beswick dkk., 2022; Bianchi dkk., 2022; Dalecká dkk., 2023; Rosario-Hernández dkk., 2023). Oleh karena itu, diperlukan instrumen skrining yang tidak hanya akurat, tetapi juga ringkas, mudah digunakan, dan sesuai dengan kebutuhan praktik sehari-hari (Poole dkk., 2020; Tibubos dkk., 2021; Wisting dkk., 2021).

Patient Health Questionnaire (PHQ) adalah instrumen yang paling banyak digunakan untuk skrining depresi di berbagai setting klinis. Dua bentuk yang paling populer adalah *PHQ-9*, yang terdiri dari sembilan item berdasarkan kriteria *DSM*, dan versi singkatnya, *PHQ-2*, yang hanya mencakup dua pertanyaan inti mengenai suasana hati dan anhedonia (Kim dkk., 2023; Wickramasinghe dkk., 2023; Wu dkk., 2020). Berbagai penelitian menunjukkan bahwa kedua instrumen memiliki

validitas dan reliabilitas yang baik, namun berbeda dalam hal sensitivitas, spesifisitas, dan kepraktisan (Errazuriz dkk., 2022; Liang dkk., 2023; Tack dkk., 2023). Meta-analisis oleh Levis dkk. (2020) menunjukkan bahwa *PHQ-2* cukup sensitif untuk deteksi awal, sedangkan *PHQ-9* mampu memberikan gambaran yang lebih menyeluruh terkait tingkat keparahan depresi (Levis dkk., 2020). Studi lain menegaskan perlunya validasi dan penyesuaian *cut-off* skor sesuai populasi dan konteks budaya (Cumbe dkk., 2020; Ghazisaeedi dkk., 2022; Liang dkk., 2023; Sun dkk., 2022). Namun, hasil perbandingan langsung antara kedua instrumen di layanan primer masih menunjukkan variasi, terutama dalam konteks keterbatasan sumber daya dan tingginya beban pasien (Harshfield dkk., 2020; Kendrick dkk., 2024; Slavin dkk., 2020).

Berdasarkan temuan tersebut, penelitian ini bertujuan meninjau secara sistematis bukti terbaru dalam lima tahun terakhir terkait akurasi *PHQ-2* dibandingkan *PHQ-9* dalam skrining depresi di layanan primer. Tujuan spesifik review ini adalah memvalidasi penggunaan pendekatan bertahap (dua tahap) yang mengoptimalkan efisiensi dengan menggunakan *PHQ-2* sebagai skrining awal cepat untuk semua pasien, dilanjutkan *PHQ-9* untuk evaluasi komprehensif pada hasil positif, bukan untuk merekomendasikan penggantian salah satu instrumen. Kajian ini memberikan pemahaman komprehensif mengenai kelebihan dan keterbatasan masing-masing instrumen untuk membantu praktisi mengimplementasikan strategi skrining yang sesuai dengan kondisi klinis dan sumber daya tersedia.

Hipotesis yang diajukan adalah bahwa meskipun *PHQ-9* memiliki akurasi diagnostik yang lebih tinggi, *PHQ-2* tetap berpotensi menjadi alat skrining awal yang efektif dan efisien di layanan primer, terutama dalam kondisi dengan keterbatasan waktu dan tenaga. Kebaruan penelitian ini terletak pada pendekatan komparatif yang sistematis terhadap kedua instrumen dengan mempertimbangkan konteks layanan primer, karakteristik populasi yang beragam, serta integrasi hasil dari berbagai negara dan budaya. Dengan demikian, kajian ini diharapkan dapat berkontribusi pada pengembangan praktik skrining depresi yang lebih tepat sasaran, berbasis bukti, serta mendukung penyusunan kebijakan kesehatan di layanan primer.

METODOLOGI

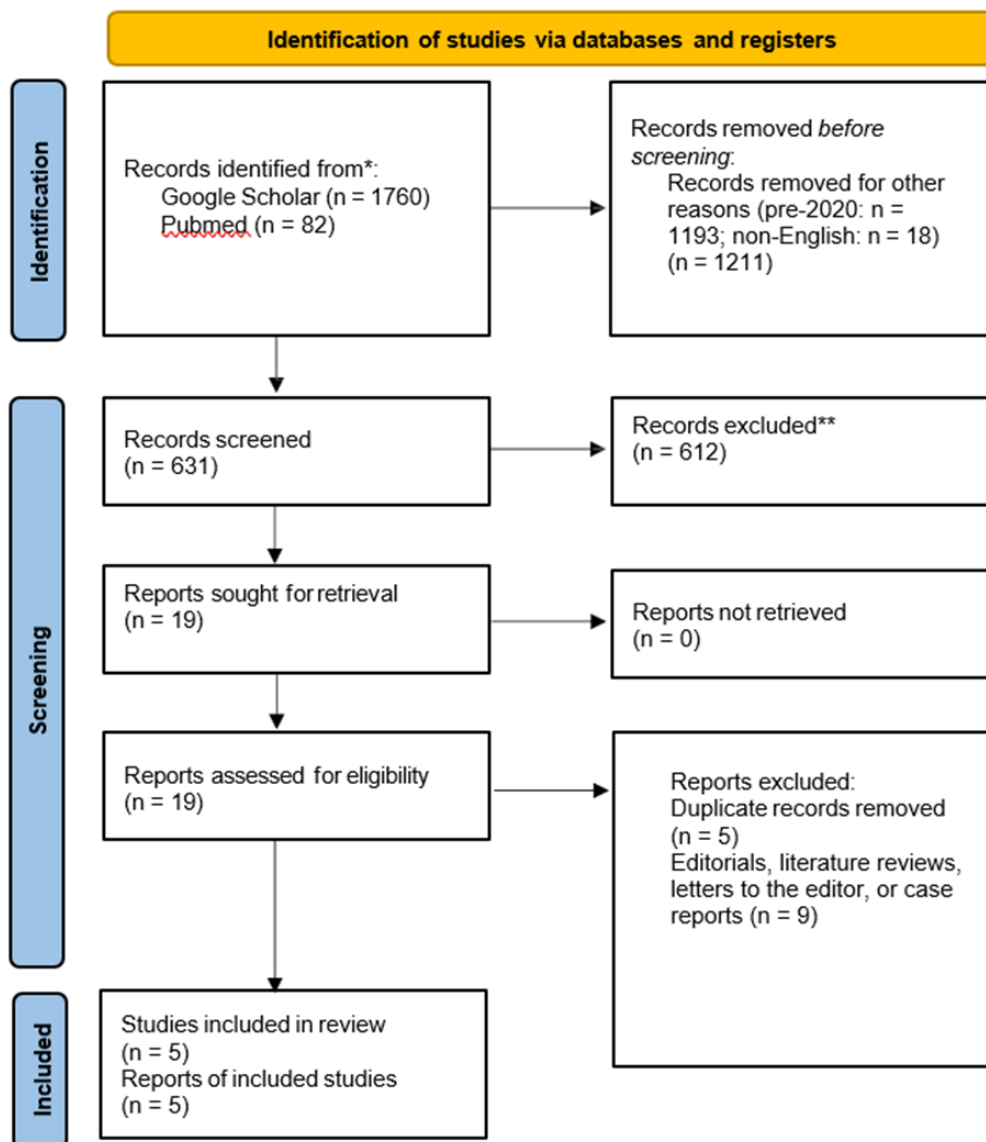
Pencarian literatur dalam penelitian ini dilakukan secara sistematis dengan menggunakan dua basis data utama, yaitu *Google Scholar* dan *PubMed*, guna memperoleh cakupan publikasi yang seluas mungkin terkait dengan skrining depresi di layanan primer. Kata kunci dirancang dengan kombinasi Boolean operator yang meliputi istilah "*PHQ-2*" atau "*Patient Health Questionnaire-2*" dipadukan dengan "*PHQ-9*" atau "*Patient Health Questionnaire-9*", serta kata "depresi" atau "depression" yang dikombinasikan dengan istilah "skrining", "screening", maupun "primary care". Pada basis data *PubMed*, strategi pencarian juga disesuaikan dengan penggunaan Medical Subject Headings (*MeSH*) agar hasil pencarian lebih terfokus dan relevan. Batasan pencarian ditetapkan pada publikasi yang terbit sejak Januari 2020 hingga September 2025, dengan pertimbangan bahwa periode tersebut mencerminkan perkembangan mutakhir dalam kajian instrumen skrining depresi. Hanya artikel yang dipublikasikan dalam bahasa Inggris atau bahasa Indonesia yang disertakan dalam tahap awal penapisan.

Proses seleksi dilakukan berdasarkan kriteria yang telah ditentukan sebelumnya. Artikel dianggap layak untuk dimasukkan apabila meneliti penggunaan *PHQ-2* atau *PHQ-9* dalam konteks layanan primer dengan melibatkan subjek manusia, baik melalui desain observasional seperti potong lintang, kohort, dan kasus kontrol, maupun melalui studi intervensi. Selain itu, publikasi juga harus menyediakan data yang berkaitan dengan akurasi diagnostik, sensitivitas, spesifisitas, atau validitas dari kedua instrumen. Sementara itu, penelitian dikeluarkan dari tinjauan jika tidak menempatkan *PHQ-2* atau *PHQ-9* sebagai fokus utama, membahas kondisi selain depresi, berupa artikel ulasan atau editorial, laporan kasus, dan komentar, hanya melibatkan penelitian berbasis laboratorium tanpa data klinis, tidak tersedia dalam bentuk teks lengkap, atau memiliki keterbatasan jumlah sampel serta tidak menyertakan kelompok pembanding. Proses penyaringan studi dijelaskan lebih lanjut melalui alur *PRISMA*. Dari total 1842 artikel yang teridentifikasi pada tahap awal (1760 dari *Google Scholar* dan 82 dari *PubMed*), sebanyak 1211 publikasi dieliminasi karena tidak sesuai dengan kriteria tahun penerbitan maupun bahasa yang dipersyaratkan. Sebanyak 631 artikel kemudian masuk tahap penyaringan, namun 612 di antaranya kembali dikeluarkan karena tidak relevan dengan topik penelitian. Tahap berikutnya menghasilkan 19 artikel yang ditelaah secara penuh, namun 14 artikel harus dikeluarkan dengan alasan adanya duplikasi, bentuk publikasi berupa ulasan atau editorial. Akhirnya, hanya lima artikel yang memenuhi kriteria kelayakan dan dimasukkan dalam tinjauan literatur ini.

Variabel utama dalam penelitian ini adalah penggunaan *PHQ-2* dan *PHQ-9* sebagai instrumen skrining depresi. *PHQ-2* didefinisikan sebagai kuesioner singkat dengan dua item pertanyaan yang dirancang untuk mendeteksi gejala depresi secara cepat dan praktis. Sementara itu, *PHQ-9* merupakan kuesioner yang lebih komprehensif dengan sembilan item pertanyaan yang tidak hanya digunakan untuk mendeteksi depresi mayor, tetapi juga menilai tingkat keparahannya sesuai dengan kriteria diagnostik *DSM*. Variabel luaran yang diperhatikan dalam tinjauan ini meliputi nilai sensitivitas, spesifisitas, nilai prediktif, serta tingkat akurasi diagnostik dari kedua instrumen ketika diterapkan di layanan primer.

Data yang diperoleh dari studi yang terpilih diekstraksi secara sistematis, meliputi rancangan penelitian, lokasi penelitian, karakteristik dan jumlah sampel, pendekatan metodologis, serta hasil utama yang dilaporkan terkait efektivitas *PHQ-2* dan *PHQ-9*. Analisis dilakukan melalui sintesis tematik untuk menemukan pola, perbedaan, serta potensi kontradiksi antar penelitian, sehingga dapat memberikan gambaran yang lebih mendalam mengenai kelebihan dan keterbatasan kedua instrumen.

Diagram *PRISMA* yang digunakan dalam penelitian ini menegaskan alur proses seleksi literatur. Dari ratusan publikasi awal, sebagian besar dieliminasi pada tahap identifikasi awal akibat faktor bahasa dan tahun publikasi, sementara seleksi lanjutan menunjukkan bahwa hanya sedikit artikel yang benar-benar sesuai dengan fokus penelitian. Pada akhirnya, lima artikel yang lolos seleksi inilah yang menjadi dasar pembahasan, menekankan pentingnya kriteria metodologis yang ketat dalam menjaga kualitas dan relevansi sebuah tinjauan literatur.



Gambar 1. Diagram *PRISMA*

HASIL DAN PEMBAHASAN

Tabel 1. Ringkasan hasil dari studi-studi yang ditinjau

No	Studi	Desain	Populasi	Perbandingan <i>PHQ-2</i> dan <i>PHQ-9</i>	Negara	Temuan Utama
1	Liang dkk. (2023)	<i>Cross-sectional, network analysis</i> untuk identifikasi item inti <i>PHQ-9</i>	1.200 remaja usia 13-18 tahun	<i>PHQ-2</i> yang dikembangkan dari item inti <i>PHQ-9</i> menunjukkan sensitivitas dan spesifisitas yang hampir setara dengan <i>PHQ-9</i> penuh, dengan kemudahan penggunaan lebih tinggi	China	<i>PHQ-2</i> baru efektif untuk skrining depresi pada remaja, memudahkan deteksi dini dengan waktu singkat tanpa mengorbankan akurasi
2	Gómez dkk. (2022)	<i>Cross-sectional, validasi diagnostik</i>	1.500 pasien dewasa usia 18-65 tahun di layanan primer	<i>PHQ-9</i> memiliki sensitivitas lebih tinggi (0.88) dibanding <i>PHQ-2</i> (0.75), namun <i>PHQ-2</i> tetap memiliki nilai prediktif positif yang baik untuk skrining awal	Spanyol	<i>PHQ-9</i> lebih akurat untuk diagnosis, tetapi <i>PHQ-2</i> dapat digunakan sebagai alat skrining awal yang efisien di layanan primer dengan keterbatasan waktu
3	Errazu rizet dkk. (2022)	<i>Cross-sectional, validasi alat skrining</i>	800 imigran Spanyol usia 20-60 tahun	<i>PHQ-9</i> menunjukkan reliabilitas internal lebih tinggi (<i>Cronbach's alpha</i> 0.89) dibanding <i>PHQ-2</i> (0.78), namun <i>PHQ-2</i> tetap valid untuk skrining cepat	Chile	<i>PHQ-2</i> dapat digunakan sebagai alat skrining awal yang praktis, terutama dalam populasi dengan hambatan bahasa dan akses layanan
4	Ghazis aeedi dkk. (2021)	<i>Cross-sectional, validasi dan penentuan cut-off</i>	1.000 mahasiswa universitas usia 18-25 tahun	<i>PHQ-2</i> efektif untuk skrining awal di populasi muda, namun <i>PHQ-9</i> diperlukan untuk evaluasi lebih mendalam	Iran	<i>PHQ-2</i> efektif untuk skrining awal di populasi muda, namun <i>PHQ-9</i> diperlukan untuk evaluasi lebih mendalam
5	Kim dkk. (2023)	<i>Cross-sectional, validasi populasi umum</i>	2.000 orang dewasa usia 19-70 tahun	<i>PHQ-9</i> menunjukkan sensitivitas 0.90 dan spesifisitas 0.85, sedangkan <i>PHQ-2</i> sensitivitas 0.80 dan spesifisitas 0.78; <i>PHQ-2</i> lebih cepat namun kurang detail	Korea Selatan	<i>PHQ-9</i> direkomendasikan untuk diagnosis definitif, <i>PHQ-2</i> cocok untuk skrining awal di layanan primer dengan beban kerja tinggi

Secara umum, tinjauan literatur menunjukkan bahwa baik *PHQ-2* maupun *PHQ-9* merupakan instrumen yang valid dan reliabel untuk skrining depresi di layanan primer, meskipun keduanya memiliki peran yang berbeda. *PHQ-9*, dengan sembilan item penilaian, terbukti memberikan sensitivitas dan spesifisitas yang lebih tinggi sehingga lebih unggul dalam mengidentifikasi depresi mayor secara akurat. Sementara itu, *PHQ-2* yang hanya berfokus pada dua gejala inti (mood depresif dan anhedonia) dinilai lebih praktis dan efisien, sehingga sesuai digunakan sebagai tahap awal skrining, terutama di fasilitas kesehatan primer yang sering menghadapi keterbatasan waktu serta tingginya jumlah pasien.

Beberapa penelitian (misalnya Gómez-Gómez & Alleau, 2022; Kim dkk., 2023) menegaskan keunggulan *PHQ-9* sebagai alat diagnostik yang lebih komprehensif, sedangkan studi lain (Liang dkk., 2023; Ghazisaeedi dkk., 2021; Errazuriz dkk., 2022) menekankan manfaat *PHQ-2* dalam mempermudah deteksi awal serta rujukan untuk evaluasi lebih lanjut. Dengan demikian, kombinasi penggunaan kedua instrumen secara berjenjang yaitu *PHQ-2* untuk penyaringan awal diikuti *PHQ-9* atau evaluasi klinis mendalam dapat menjadi strategi yang optimal dalam praktik layanan primer.

Selain itu, literatur menyoroti pentingnya adaptasi *cut-off* skor dan validasi lokal pada berbagai kelompok populasi dan budaya. Studi di kalangan mahasiswa Iran, remaja Tiongkok, maupun imigran berbahasa Spanyol menunjukkan bahwa perbedaan konteks demografis, bahasa, dan budaya dapat memengaruhi hasil skrining. Penyesuaian ini krusial untuk meningkatkan akurasi serta mengurangi kemungkinan salah klasifikasi, baik berupa *false positive* maupun *false negative*.

Kendati demikian, sebagian besar penelitian yang ditinjau masih menggunakan desain *cross-sectional*, sehingga belum dapat menilai efektivitas skrining dalam jangka panjang. Selain itu, variasi ukuran sampel dan fokus pada populasi khusus membatasi generalisasi hasil pada populasi umum. Oleh karena itu, penelitian prospektif dengan cakupan sampel yang lebih heterogen sangat diperlukan untuk memperkuat bukti dan menyusun rekomendasi praktik yang lebih kokoh.

Efektivitas Skrining dalam Berbagai Konteks Layanan Primer

Berbagai penelitian menunjukkan bahwa efektivitas *PHQ-2* dan *PHQ-9* dalam mendeteksi depresi sangat dipengaruhi oleh konteks layanan primer di mana alat tersebut digunakan. Di negara dengan sistem layanan kesehatan yang padat dan waktu konsultasi terbatas, seperti yang dilaporkan oleh Kim dkk. (2023) di Korea Selatan, *PHQ-2* menjadi pilihan yang lebih praktis untuk skrining awal karena kemudahan dan kecepatan pengisian. Sebaliknya, di setting yang memungkinkan evaluasi lebih mendalam, *PHQ-9* lebih disarankan karena memberikan gambaran yang lebih lengkap mengenai kondisi pasien. Hal ini menegaskan bahwa pemilihan alat skrining harus disesuaikan dengan kondisi operasional dan kebutuhan klinis di layanan primer.

Perbandingan Akurasi PHQ-2 dan PHQ-9

Hasil dari berbagai studi menunjukkan bahwa *PHQ-9* secara konsisten memiliki tingkat sensitivitas dan spesifisitas yang lebih tinggi dibandingkan *PHQ-2*. *PHQ-9*, yang terdiri dari sembilan item, memberikan gambaran yang lebih lengkap mengenai gejala depresi, sehingga lebih mampu mengidentifikasi kasus depresi mayor dengan akurasi yang lebih baik. Sebagai contoh, Gómez-Gómez dan Alleau (2022) melaporkan sensitivitas *PHQ-9* sebesar 88%, sedangkan *PHQ-2* hanya 75%. Namun demikian, *PHQ-2* tetap memiliki nilai klinis yang signifikan sebagai alat skrining awal yang cepat dan mudah digunakan, terutama dalam situasi di mana waktu konsultasi terbatas. Studi Liang dkk. (2023) bahkan mengembangkan versi *PHQ-2* yang disesuaikan untuk remaja Tiongkok dengan hasil yang menjanjikan, menunjukkan bahwa *PHQ-2* dapat dioptimalkan untuk populasi tertentu tanpa kehilangan akurasi secara signifikan.

Peran PHQ-2 dan PHQ-9 dalam Deteksi Dini dan Rujukan Lanjutan

PHQ-2 berperan penting sebagai alat skrining awal yang dapat mengidentifikasi individu yang berisiko mengalami depresi sehingga dapat segera dirujuk untuk pemeriksaan lebih lanjut menggunakan *PHQ-9* atau evaluasi klinis. Studi Liang dkk. (2023) menegaskan bahwa *PHQ-2* yang disesuaikan dengan populasi remaja mampu mendeteksi gejala inti depresi secara efektif, sehingga memfasilitasi intervensi dini. Namun, *PHQ-9* tetap menjadi standar dalam memastikan diagnosis definitif dan menentukan tingkat keparahan depresi, yang sangat penting untuk perencanaan pengobatan dan tindak lanjut.

Kelebihan dan Keterbatasan PHQ-2 dalam Skrining Awal

PHQ-2 memiliki keunggulan utama berupa kemudahan dan kecepatan administrasi, yang sangat berguna di layanan primer dengan beban pasien yang tinggi. Alat ini memungkinkan tenaga kesehatan untuk melakukan skrining awal secara efisien dan menentukan pasien yang perlu evaluasi lebih lanjut menggunakan *PHQ-9* atau pemeriksaan klinis mendalam. Namun, keterbatasan *PHQ-2* terletak pada kemampuannya yang lebih rendah dalam mendeteksi depresi ringan atau gejala yang lebih kompleks, sehingga berpotensi menghasilkan *false negative*. Hal ini ditegaskan oleh Errazuriz dkk. (2022) dan Ghazisaeedi dkk. (2021) yang menemukan reliabilitas internal *PHQ-2* lebih rendah dibanding *PHQ-9*, meskipun tetap valid untuk skrining cepat.

Implikasi Budaya dan Penyesuaian Skor Cut-off

Penelitian juga menyoroti pentingnya penyesuaian skor *cut-off* *PHQ-2* dan *PHQ-9* sesuai dengan karakteristik budaya dan demografis populasi yang diuji. Misalnya, Ghazisaeedi dkk. (2021) menyesuaikan *cut-off* untuk mahasiswa Iran, sementara Liang dkk. (2023) mengadaptasi *PHQ-2* untuk remaja Tiongkok. Penyesuaian ini bertujuan untuk meningkatkan akurasi skrining dan mengurangi kesalahan diagnosis yang mungkin timbul akibat perbedaan ekspresi gejala depresi antar budaya. Oleh karena itu, penggunaan *PHQ-2* dan *PHQ-9* harus mempertimbangkan konteks lokal agar hasil skrining lebih valid dan dapat diandalkan.

Implikasi Praktis untuk Tenaga Kesehatan di Layanan Primer

Dari perspektif praktis, penggunaan *PHQ-2* dan *PHQ-9* harus mempertimbangkan beban kerja tenaga kesehatan dan kebutuhan pasien. *PHQ-2* yang singkat memungkinkan skrining cepat tanpa membebani tenaga kesehatan, sehingga dapat diterapkan secara rutin pada semua pasien yang datang ke layanan primer. Namun, untuk pasien yang menunjukkan hasil positif pada *PHQ-2*, pemeriksaan lanjutan dengan *PHQ-9* atau wawancara klinis mendalam sangat dianjurkan untuk memastikan diagnosis dan menentukan intervensi yang tepat. Pendekatan bertingkat ini dapat meningkatkan efisiensi skrining sekaligus menjaga kualitas diagnosis. Strategi ini dikenal sebagai metode “*Screen-and-Confirm*” atau skrining dua tahap (*two-step screening*), dimana semua pasien di layanan primer terlebih dahulu menjalani *PHQ-2* sebagai penyaring cepat karena hanya membutuhkan waktu sekitar 1-2 menit, kemudian pasien dengan hasil positif dilanjutkan dengan *PHQ-9* untuk konfirmasi dan penilaian keparahan depresi. Dari segi kemampuan diskriminasi, *PHQ-9* umumnya menunjukkan nilai *Area Under the Curve* (AUC) yang lebih besar dibandingkan *PHQ-2*. AUC merupakan metrik statistik yang mengukur kemampuan tes diagnostik membedakan antara individu dengan dan tanpa kondisi tertentu, dengan nilai berkisar 0 hingga 1. Nilai AUC yang lebih tinggi menunjukkan kemampuan diskriminasi yang lebih baik, dimana nilai 0,9-1,0 dianggap sangat baik, 0,8-0,9 baik, dan 0,7-0,8 cukup. Secara konsisten, *PHQ-9* mencapai AUC di atas 0,85, mencerminkan akurasi diagnostik yang superior dalam mengidentifikasi kasus depresi mayor.

Diskusi

Hasil analisis menunjukkan perbedaan karakteristik mendasar antara kedua instrumen skrining. *PHQ-9* memperlihatkan performa diagnostik superior dengan sensitivitas 0,88-0,90 dan spesifisitas 0,85, sementara *PHQ-2* mencapai sensitivitas 0,75-0,80 dengan spesifisitas 0,78. Disparitas ini mencerminkan perbedaan desain fundamental dimana *PHQ-9* menyediakan evaluasi komprehensif melalui sembilan item, sedangkan *PHQ-2* mengutamakan efisiensi deteksi via dua gejala kardinal depresi.

Temuan dari lima studi lintas negara mengkonfirmasi pentingnya adaptasi kontekstual. Penelitian Liang di China memvalidasi efektivitas *PHQ-2* termodifikasi untuk remaja, sementara Errazuriz di Chile membuktikan validitas pada komunitas imigran. Ghazisaeedi di Iran menunjukkan perlunya penyesuaian ambang batas untuk populasi mahasiswa, mengindikasikan variabilitas manifestasi depresif antar kelompok demografis.

Hasil ini konsisten dengan Levis dkk. yang menyimpulkan *PHQ-2* efektif untuk deteksi awal sementara *PHQ-9* memberikan evaluasi mendalam. Namun, penelitian ini menawarkan perspektif baru dengan menunjukkan potensi pendekatan bertingkat yang mengoptimalkan keunggulan kedua instrumen. Gómez-Gómez di Spanyol memvalidasi bahwa meskipun *PHQ-9* superior dalam akurasi, *PHQ-2* mempertahankan nilai prediktif positif memadai untuk skrining awal.

Evaluasi psikometrik menunjukkan *PHQ-9* memiliki reliabilitas internal superior (*Cronbach's alpha* 0,89) dibanding *PHQ-2* (0,78). Perbedaan ini dapat dijelaskan melalui prinsip pengukuran klasik dimana item lebih banyak meningkatkan konsistensi internal. Namun, untuk tujuan skrining cepat, reliabilitas *PHQ-2* tetap acceptable sesuai standar klinis.

Kim di Korea Selatan mendemonstrasikan bahwa dalam sistem kesehatan berbeban tinggi, *PHQ-2* menawarkan solusi pragmatis meskipun dengan kompromis pada sensitivitas. Implementasi strategi bertingkat menggunakan *PHQ-2* sebagai filter awal diikuti *PHQ-9* untuk konfirmasi dapat mengoptimalkan alokasi sumber daya sambil mempertahankan kualitas diagnostik.

Keterbatasan metodologis meliputi dominasi desain *cross-sectional* yang membatasi pemahaman stabilitas temporal instrumen dan variabilitas ukuran sampel yang membatasi generalisabilitas. Penelitian longitudinal diperlukan untuk memahami responsivitas instrumen terhadap perubahan klinis dan intervensi terapeutik.

Tinjauan ini berkontribusi pada evidence base dengan menghasilkan *framework* konseptual untuk implementasi skrining bertingkat yang *evidence-based*. Temuan menunjukkan *PHQ-2* dan *PHQ-9* memiliki peran komplementer dalam deteksi depresi di layanan primer. Strategi bertingkat menawarkan solusi praktis yang mengkombinasikan efisiensi *PHQ-2* dengan akurasi *PHQ-9*, memberikan panduan implementasi untuk layanan primer dengan keterbatasan sumber daya. Implikasi kebijakan mencakup pengembangan *guideline* klinis yang mengakomodasi variabilitas konteks layanan dan pelatihan tenaga kesehatan dalam aspek *cultural competency* serta *clinical judgment*.

KESIMPULAN DAN SARAN

Tinjauan ini menunjukkan bahwa *PHQ-9* memiliki akurasi diagnostik lebih baik dibanding *PHQ-2* dalam mendeteksi depresi di layanan primer karena mampu menggambarkan gejala secara lebih komprehensif. Meski demikian, *PHQ-2* tetap bermanfaat sebagai skrining awal yang cepat dan praktis, terutama pada situasi dengan keterbatasan waktu dan sumber daya. Pendekatan bertahap dengan memanfaatkan keunggulan kedua instrumen dapat meningkatkan efektivitas deteksi dini depresi.

Penelitian selanjutnya disarankan mengevaluasi penerapan jangka panjang *PHQ-2* dan *PHQ-9* melalui studi longitudinal, serta memperhatikan faktor pelatihan tenaga kesehatan, penerimaan pasien, dan validasi lintas budaya untuk memastikan relevansi dan akurasi skrining.

DAFTAR PUSTAKA

- Adams, D. P., Holt, J. R., Martin, J. A., Houpy, D. M., & Hollenbach, K. A. (2022). The Effect of COVID-19 Lockdown on PHQ Depression Screening Scores for High School Athletes. *International Journal of Environmental Research and Public Health*, 19(16). <https://doi.org/10.3390/ijerph19169943>
- Asim, R., Asim, M., Reddy, R., Chepulis, L., & Lawrenson, R. (2023). Screening for depression as part of annual diabetic review using *PHQ-9* scores: A pilot study. *Primary Health Care Research and Development*, 24. <https://doi.org/10.1017/S1463423623000294>
- Beswick, E., Quigley, S., Macdonald, P., Patrick, S., Colville, S., Chandran, S., & Connick, P. (2022). The Patient Health Questionnaire (*PHQ-9*) as a tool to screen for depression in people with multiple sclerosis: a cross-sectional validation study. *BMC Psychology*, 10(1). <https://doi.org/10.1186/s40359-022-00949-8>
- Bianchi, R., Verkuilen, J., Toker, S., Schonfeld, I. S., Gerber, M., Brähler, E., & Kroenke, K. (2022). Is the *PHQ-9* a Unidimensional Measure of Depression? A 58,272-Participant Study. *Psychological Assessment*, 34(6), 595–603. <https://doi.org/10.1037/pas0001124>
- Camargo, L., Soto-Añari, M., Caldichoury-Obando, N., & Ramos-Henderson, M. (2022). Shelach Salomón 6e , Carol Saldías-Solis 7f , Pascual Gargiulo 8 , Norman López 9g REV. In *CHIL NEURO-PSIQUIAT* (Vol. 60, Issue 3). www.sonepsyn.cl
- Cumbe, V. F. J., Muanido, A., Manaca, M. N., Fumo, H., Chiruca, P., Hicks, L., De Jesus Mari, J., & Wagenaar, B. H. (2020). Validity and item response theory properties of the Patient Health Questionnaire-9 for primary care depression screening in Mozambique (*PHQ-9-MZ*). *BMC Psychiatry*, 20(1). <https://doi.org/10.1186/s12888-020-02772-0>
- Dalecká, A., Dalecká, A., Pikhart, H., Apková, C. ě, & Bobá, M. (2023). Abstract citation ID: ckac131.485 Pandemic-related determinants of depression in Czech older persons: Evidence from the HAPIEE cohort. *European Journal Of Public Health*, 32. https://academic.oup.com/eurpub/article/32/Supplement_3/ckac131.487/6766315
- Errazuriz, A., Beltrán, R., Torres, R., & Passi-Solar, A. (2022). The Validity and Reliability of the *PHQ-9* and *PHQ-2* on Screening for Major Depression in Spanish Speaking Immigrants in Chile: A Cross-sectional Study. *International Journal of Environmental Research and Public Health*, 19(21). <https://doi.org/10.3390/ijerph192113975>
- Ghazisaeedi, M., Mahmoodi, H., Arpaci, I., Mehrdar, S., & Barzegari, S. (2022). Validity, Reliability, and Optimal Cut-off Scores of the WHO-5, *PHQ-9*, and *PHQ-2* to Screen Depression Among University Students in Iran. *International Journal of Mental Health and Addiction*, 20(3), 1824–1833. <https://doi.org/10.1007/s11469-021-00483-5>
- Gómez-Gómez, I., Benítez, I., Bellón, J., Moreno-Peral, P., Oliván-Blázquez, B., Clavería, A., Zabaleta-Del-Olmo, E., Llobera, J., Serrano-Ripoll, M. J., Tamayo-Morales, O., & Motrico, E. (2023). Utility of *PHQ-2*, *PHQ-8* and *PHQ-9* for detecting major depression in primary health care: a validation study in Spain. *Psychological Medicine*, 53(12), 5625–5635. <https://doi.org/10.1017/S0033291722002835>
- Harshfield, E. L., Pennells, L., Schwartz, J. E., Willeit, P., Kaptoge, S., Bell, S., Shaffer, J. A., Bolton, T., Spackman, S., Wassertheil-Smoller, S., Kee, F., Amouyel, P., Shea, S. J., Kuller, L. H., Kauhanen, J., Van Zutphen, E. M., Blazer, D. G., Krumholz, H., Nietert, P. J., ... Davidson, K. W. (2020). Association between Depressive Symptoms and Incident Cardiovascular Diseases. *JAMA - Journal of the American Medical Association*, 324(23), 2396–2405. <https://doi.org/10.1001/jama.2020.23068>
- Kendrick, T., Dowrick, C., Lewis, G., Moore, M., Leydon, G. M., Geraghty, A. W. A., Griffiths, G., Zhu, S., Yao, G. L., May, C., Gabbay, M., Dewar-Haggart, R., Williams, S., Bui, L., Thompson, N., Bridewell, L., Trapasso, E., Patel,

- T., McCarthy, M., ... Stuart, B. (2024). Depression follow-up monitoring with the *PHQ-9*: an open cluster-randomised controlled trial. *British Journal of General Practice*, 74(744), e456–e465. <https://doi.org/10.3399/BJGP.2023.0539>
- Kim, M., Jung, S., Park, J. E., Sohn, J. H., Seong, S. J., Kim, B. S., Chang, S. M., Hong, J. P., Hahm, B. J., & Yeom, C. W. (2023). Validation of the *Patient Health Questionnaire-9* and *Patient Health Questionnaire-2* in the General Korean Population. *Psychiatry Investigation*, 20(9), 853–860. <https://doi.org/10.30773/pi.2023.0100>
- Lee, J., Lee, K. E., Park, S., & Jun, K. Y. (2022). Validation of the *Patient Health Questionnaire-9* for Screening Depressive Disorders among Korean Employees: A Longitudinal Study of the National Health Examination Data. *International Journal of Environmental Research and Public Health*, 19(7). <https://doi.org/10.3390/ijerph19073780>
- Levis, B., Sun, Y., He, C., Wu, Y., Krishnan, A., Bhandari, P. M., Neupane, D., Imran, M., Brehaut, E., Negeri, Z., Fischer, F. H., Benedetti, A., & Thombs, B. D. (2020). Accuracy of the *PHQ-2* Alone and in Combination with the *PHQ-9* for Screening to Detect Major Depression: Systematic Review and Meta-analysis. In *JAMA - Journal of the American Medical Association* (Vol. 323, Issue 22, pp. 2290–2300). American Medical Association. <https://doi.org/10.1001/jama.2020.6504>
- Liang, K., Chen, S., Zhao, Y., Ren, Y., Ren, Z., & Chi, X. (2023). A new *PHQ-2* for Chinese adolescents: identifying core items of the *PHQ-9* by network analysis. *Child and Adolescent Psychiatry and Mental Health*, 17(1). <https://doi.org/10.1186/s13034-023-00559-1>
- Mohamed Khalifa, A. F., Ahmed Hussamuldin, A. B., Alkhatran, R. M., Alghamdi, A. A., Rifaey, A. A., Alabdullah, A. M., Raddawi, B. Z., Almansour, M. M., Almogheer, R. S., Alawi, S. A., Malaekkah, K. R., Balla Elnour, M. A., & Omer, H. I. (2023). *PHQ-9* to screen for depression in Riyadh, Saudi Arabia. *Medical Science*, 27(135), 1–7. <https://doi.org/10.54905/disssi/v27i135/e204ms2974>
- Pellas, J., & Damberg, M. (2021). Accuracy in detecting major depressive episodes in older adults using the Swedish versions of the GDS-15 and *PHQ-9*. *Uppsala Journal of Medical Sciences*, 126. <https://doi.org/10.48101/ujms.v126.7848>
- Poole, D. N., Liao, S., Larson, E., Hedt-Gauthier, B., Raymond, N. A., Bärnighausen, T., & Smith Fawzi, M. C. (2020). Sequential screening for depression in humanitarian emergencies: A validation study of the *Patient Health Questionnaire* among Syrian refugees. *Annals of General Psychiatry*, 19(1). <https://doi.org/10.1186/s12991-020-0259-x>
- Rosario-Hernández, E., Rovira-Millán, L. V., Merino-Soto, C., & Angulo-Ramos, M. (2023). Review of the psychometric properties of the *Patient Health Questionnaire-9 (PHQ-9)* Spanish version in a sample of Puerto Rican workers. *Frontiers in Psychiatry*, 14. <https://doi.org/10.3389/fpsy.2023.1024676>
- Slavin, V., Creedy, D. K., & Gamble, J. (2020). Comparison of screening accuracy of the *Patient Health Questionnaire-2* using two case-identification methods during pregnancy and postpartum. *BMC Pregnancy and Childbirth*, 20(1). <https://doi.org/10.1186/s12884-020-02891-2>
- Sun, Y., Fu, Z., Bo, Q., Mao, Z., Ma, X., & Wang, C. (2020). The reliability and validity of *PHQ-9* in patients with major depressive disorder in psychiatric hospital. *BMC Psychiatry*, 20(1). <https://doi.org/10.1186/s12888-020-02885-6>
- Sun, Y., Kong, Z., Song, Y., Liu, J., & Wang, X. (2022). The validity and reliability of the *PHQ-9* on screening of depression in neurology: a cross sectional study. *BMC Psychiatry*, 22(1). <https://doi.org/10.1186/s12888-021-03661-w>
- Tack, L., Maenhoudt, A. S., Ketelaars, L., De Zutter, J., Pinson, S., Keunebrock, L., Haaker, L., Deckmyn, K., Gheysen, M., Kenis, C., Wildiers, H., Depoorter, L., Geerts, P. J., Chandler, R., Boterberg, T., Schofield, P., Parris, C. N., & Debruyne, P. R. (2023). Diagnostic Performance of Screening Tools for Depressive Symptoms in Vulnerable Older Patients with Cancer Undergoing Comprehensive Geriatric Assessment (CGA): Results from the SCREEN Pilot Study. *Current Oncology*, 30(2), 1805–1817. <https://doi.org/10.3390/curroncol30020140>
- Teusen, C., Hapfelmeier, A., von Schrottenberg, V., Gökce, F., Pitschel-Walz, G., Henningsen, P., Gensichen, J., Schneider, A., Dreischulte, T., Falkai, P., Gensichen, J., Henningsen, P., Bühner, M., Jung-Sievers, C., Krcmar, H., Lukaschek, K., Pitschel-Walz, G., Schneider, A., Vukas, J., ... von Nessen-Lapp, R. W. (2022). Combining the GP's assessment and the *PHQ-9* questionnaire leads to more reliable and clinically relevant diagnoses in primary care. *PLoS ONE*, 17(10 October). <https://doi.org/10.1371/journal.pone.0276534>
- Tibubos, A. N., Otten, D., Zöllner, D., Binder, H., Wild, P. S., Fleischer, T., Johar, H., Atasoy, S., Schulze, L., Ladwig, K. H., Schomerus, G., Linkohr, B., Grabe, H. J., Kruse, J., Schmidt, C. O., Münzel, T., König, J., Brähler, E., & Beutel, M. E. (2021). Bidimensional structure and measurement equivalence of the *Patient Health Questionnaire-9*: sex-sensitive assessment of depressive symptoms in three representative German cohort studies. *BMC Psychiatry*, 21(1). <https://doi.org/10.1186/s12888-021-03234-x>

- Wickramasinghe, A., Essén, B., Surenthirakumaran, R., & Axemo, P. (2023). Prevalence of depression among students at a Sri Lankan University: A study using the *Patient Health Questionnaire-9 (PHQ-9)* during the COVID-19 pandemic. *BMC Public Health*, 23(1). <https://doi.org/10.1186/s12889-023-15427-y>
- Wisting, L., Johnson, S. U., Bulik, C. M., Andreassen, O. A., Rø, Ø., & Bang, L. (2021). Psychometric properties of the Norwegian version of the *Patient Health Questionnaire-9 (PHQ-9)* in a large female sample of adults with and without eating disorders. *BMC Psychiatry*, 21(1). <https://doi.org/10.1186/s12888-020-03013-0>
- Wu, Y., Levis, B., Riehm, K. E., Saadat, N., Levis, A. W., Azar, M., Rice, D. B., Boruff, J., Cuijpers, P., Gilbody, S., Ioannidis, J. P. A., Kloda, L. A., McMillan, D., Patten, S. B., Shrier, I., Ziegelstein, R. C., Akena, D. H., Arroll, B., Ayalon, L., ... Thombs, B. D. (2020). Erratum: Equivalency of the diagnostic accuracy of the PHQ-8 and *PHQ-9*: a systematic review and individual participant data meta-analysis - ERRATUM (*Psychological medicine* (2020) 50 8 (1368-1380)). In *Psychological medicine* (Vol. 50, Issue 16, p. 2816). NLM (Medline). <https://doi.org/10.1017/S0033291719002137>