

General Health Questionnaire-12 (GHQ-12) sebagai Instrumen Skrining Gangguan Penyesuaian

Indira Primasari¹

¹Fakultas Psikologi Universitas Indonesia

Rahmat Hidayat²

²Fakultas Psikologi Universitas Gadjah Mada

Abstract. Adjustment Disorder (AD) is a common mental health problem in primary care setting. This study aimed to examine the validity of GHQ-12 as Adjustment Disorder screening instrument in Public Health Center (Puskesmas). The subjects were 243 adult outpatients from twenty five Puskesmas in Sleman, Yogyakarta. The prevalence of AD was 12.9%. A cronbach's Alpha of .863 (likert's scoring method), .841 (bimodal's scoring method), .832 (CGHQ's scoring method) were obtained, Sensitivity and specificity for GHQ-12 were .81 and .62 (for the optimum cut-off point ≥ 11 in Likert's scoring method), .81 and .57 (for the optimum cut-off point ≥ 2 in bimodal's scoring method), .81 and .55 (for the optimum cut-off point ≥ 4 in CGHQ's scoring method). Positive Likelihood Ratio and negative Likelihood Ratio for GHQ-12 were 2.12 and 0.31 (Likert's scoring method), 1.90 and 0.34 (Bimodal's scoring method), and 1.80 and 0.35 (CGHQ's scoring method). The GHQ-12 is valid and reliable in Puskesmas population, and can be employed as a screening instrumen in this population.

Keywords: GHQ-12, SCID, adjustment disorder, public health center (puskesmas), validity

Abstrak. Gangguan Penyesuaian (GP) banyak dijumpai pada pasien yang datang di layanan kesehatan primer, terutama Pusat Kesehatan Masyarakat (Puskesmas). Penelitian ini bertujuan untuk menguji validitas klinik GHQ-12, sebagai instrumen skrining GP pada pasien Puskesmas. Penelitian ini melibatkan 243 pasien di 25 Puskesmas di Kabupaten Sleman, Yogyakarta. Hasil analisis menunjukkan prevalensi GP di Puskesmas sejumlah 12,9%. Analisis reliabilitas menunjukkan hasil yang memuaskan ($r_{xx} = 0,863; 0,841; 0,832$). Analisis ROC dan LR menghasilkan nilai sensitivitas sebesar 0,81, spesivitas 0,62, LR+ 2,12, dan LR- 0,31, dengan titik potong optimum ≥ 11 (Likert), sensitivitas 0,81, spesivitas 0,57, LR+ 1,90, dan LR- 0,34, dengan titik potong optimum ≥ 2 (Bimodal), serta spesivitas 0,81, spesivitas 0,55, LR+ 1,80, dan LR- 0,35 dengan titik potong ≥ 4 (CGHQ). Dapat disimpulkan bahwa GHQ-12 valid, reliabel dan akurat sebagai instrumen skrining GP.

Kata kunci: GHQ-12, SCID, gangguan penyesuaian, puskesmas, validitas

Salah satu gangguan mental emosional yang banyak dijumpai di pusat

layanan kesehatan primer, seperti Puskesmas, adalah gangguan penyesuaian (GP). Hasil penelitian menunjukkan bahwa tingkat prevalensi GP dapat mencapai 11%-18% (Casey dan Doherty,

¹ Korespondensi mengenai isi artikel ini dapat dilakukan melalui: indira.primasari@gmail.com dan rhidayat262@ugm.ac.id

2012). Secara klinis, GP muncul sebagai akibat dari adanya perubahan hidup yang berarti, yang berkembang dalam waktu tiga bulan sejak onset stresor terjadi pertama kali. Keberadaan GP ditandai oleh adanya gejala distress psikologis, penurunan fungsi sosial, dan harus tidak menunjukkan adanya kedukaan (Carta, Balestrieri, Murru & Hardoy, 2009 dan First, 1994).

Meskipun secara statistik terdapat tingkat prevalensi GP yang cukup tinggi di pusat layanan kesehatan primer, namun beberapa studi menunjukkan bahwa penderita GP belum teridentifikasi dan tertangani secara optimal. Kondisi tersebut disebabkan oleh beberapa hal, yaitu: (a) kesulitan paramedis dalam mendiagnosis pasien yang sebagian besar datang dengan keluhan gejala-gejala somatik, dan bukan gejala psikologis, (b) kurangnya pengetahuan dan kepercayaan diri paramedis dalam menemukan kriteria diagnostik yang tepat, (c) terbatasnya waktu untuk menegakkan diagnosis akibat kesibukan pada jam kerja dan banyaknya jumlah pasien dibandingkan dengan jumlah paramedis, (d) adanya stigma masyarakat terkait dengan gangguan mental, (e) resistensi pasien terhadap hasil diagnosis gangguan jiwa dan rujukan untuk konsultasi ke psikolog, serta (f) adanya persamaan gejala neurologis dan perilaku dari gangguan GP dan depresi (Menderscheid, Spitzer & Katzelnick, dalam Schmitz, Kruse, Heckrath, Alberti & Tress, 1999; Patel dkk., 2007; Primasari, 2012; CPMH, 2012; Carta dkk, 2009).

Kurang optimalnya identifikasi dan penanganan pasien GP di pusat layanan kesehatan primer, khususnya Puskesmas, tentu memerlukan solusi yang konkret dan relevan. Razavi, Delvaux, Farvacques dan Robave (1990); Gardner, Kelleher dan Pajer (2002), dan Patel dkk. (2007)

merekomendasikan perlunya pengembangan instrumen skrining gangguan mental, khususnya GP, yang sensitif, spesifik, dan sederhana di pusat pelayanan kesehatan primer. Perlu diketahui bahwa instrumen psikodiagnostik yang digunakan di Puskesmas saat ini sudah beragam, akan tetapi masih rumit dan lama dalam pengadministrasian (Rahardian dalam Oktaviana dan Wimbari, 2014)

Instrumen skrining, sebagaimana dijelaskan oleh Mitchel (2010), merupakan instrumen pendalaman yang dibuat secara sistematis dan bertujuan untuk mengidentifikasi individu yang berisiko mengalami suatu gangguan tertentu. Pengembangan instrumen skrining diperlukan untuk menjamin agar individu tersebut mendapatkan pertolongan medis atas gangguan yang dideritanya. Instrumen skrining GP juga sangat penting dikembangkan mengingat tingginya risiko perilaku bunuh diri penderitanya. Runeson dkk. (dalam Strain & Friedman, 2011) melaporkan jarak interval yang pendek (satu bulan) antara ditegakkannya diagnosis GP dengan timbulnya perilaku bunuh diri, dibandingkan dengan gangguan depresi (tiga bulan), gangguan kepribadian bipolar (30 bulan), dan skizofrenia (47 bulan).

Salah satu instrumen skrining yang banyak dipakai secara internasional di pusat pelayanan kesehatan dasar adalah *General Health Questionnaire* (GHQ). Instrumen ukur ini disusun oleh David Goldberg dan dipublikasikan pertama kali pada tahun 1970. Dalam publikasi tersebut, Goldberg dan Blackwell (1970) menyebutkan bahwa GHQ telah terbukti valid dalam mengidentifikasi gangguan psikiatrik pada pasien-pasien di layanan kesehatan dasar, bahkan ketika diuji secara independen dengan metode ases-

men klinis yang lainnya, yaitu wawancara oleh psikiater. Hasil penelitian menunjukkan bahwa terdapat korelasi yang baik antara skor kuesioner dengan rating wawancara klinis ($r=+0.80$) dan 91.5% pasien berhasil diklasifikasi apakah sungguh-sungguh terganggu atau tidak, secara tepat. Saat ini, GHQ telah diterjemahkan ke 38 bahasa setelah melalui serangkaian tes validitas dan reliabilitas sehingga dapat digunakan di lebih dari 50 negara (Bell, Watson, Sharp, Lyons, & Lewis, 2005; Makowska, Merez, Moscicka & Kolasa, 2002; Allison dkk., 2005; Nitsche, Koch dan Kallert, 2010; Kapur dkk., 2004; Mellor, Braydon & Steptoe, 2010; Uher dan Goodman, 2010; Pevalin, 2000; Evans, Kessler, Lewis, Peters & Sharp, 2004; Jacob, Bhugra & Manh, 1997; Berardi, dkk., 1999; Belek, 2000; Kuruvilla dkk., 1999; Patel dkk., 2007; Sanchez-Lopez dan Dresch, 2008; Fellingner dkk, 2005; Gureje, 2002).

General Health Questionnaire (GHQ) terdiri dari 60 butir pertanyaan, dan sering dikenal dengan sebutan GHQ-60. GHQ versi 60 kemudian mengalami berbagai modifikasi, yaitu versi 30, 28, dan 12 pertanyaan. Sementara itu, versi GHQ yang sesuai dengan kebutuhan di Puskesmas. adalah GHQ versi 12 atau GHQ-12, dengan pertimbangan versi ini adalah versi yang paling banyak dipakai, memiliki validitas yang mapan, bahasa yang sederhana, ringkas dan waktu pengerjaan relatif cepat (Yusoff, Rahim & Yacoob, 2009).

Berbagai penelitian menunjukkan bahwa GHQ-12 telah diuji secara psikometrik untuk dipakai sebagai instrumen skrining distres psikologis di pelayanan kesehatan dasar di berbagai negara, dan memiliki validitas dan reliabilitas yang tinggi (Belek, 2000; Idaiani dan Suhardi, 2006; Patel dkk., 2007; Sanchez-Lopez dan Dresch, 2008; Fellingner dkk., 2005; Uher

dan Goodman, 2010), saat diuji dengan instrumen lain sebagai standar baku emas (Makowska, dkk., 2002; Pevalin, 2000; Berardi dkk., 1999; Uher dan Goodman, 2010; Fellingner dkk., 2005; Patel dkk, 2007). Instrumen skrining ini juga telah digunakan dalam berbagai jenis populasi, sebagai contoh penderita tuna rungu (Fellingner dkk., 2005) dan pasien dengan penyakit kardiovaskular (Mellor, dkk., 2010), serta diadministrasikan dalam berbagai cara selain *self-report*, yaitu wawancara (Kuruvilla dkk, 1999) dan melalui komputer (Fellingner dkk, 2005).

Terkait dengan permasalahan di atas dan untuk menjawab kebutuhan yang ada di Puskesmas, tujuan dari penelitian ini adalah untuk menguji validitas, menentukan titik potong optimum, dan menentukan jenis skoring yang paling tepat dari GHQ-12, sehingga dapat digunakan sebagai instrumen skrining GP pada pasien Puskesmas. Tujuan penelitian ini selaras dengan rekomendasi yang diusulkan dalam FGD Kelompok Pembuatan Sistem Program *Inter-University Partnership To Improve Mental Health Care In Indonesia* (CPMH, 2012) dan prioritas dari Rencana Kegiatan Pemerintah Direktorat Bina Kesehatan Jiwa Republik Indonesia tahun 2011 (Irmansyah, 2012). Hasil penelitian ini bermanfaat untuk membantu meningkatkan kepekaan paramedis dalam mendeteksi GP di Puskesmas, sehingga para pasien yang mengalami GP akan menjadi lebih cepat tertangani.

Metode

Kriteria subjek penelitian adalah pasien Balai Pelayanan Umum (BPU) Puskesmas dan berusia 18 – 65 tahun, dengan ukuran sampel sebesar 250 orang, di 25 Puskesmas di Kabupaten Sleman. Metode sampling yang digunakan adalah

non probability accidental sampling, yaitu pasien yang memiliki nomor urut pendaftaran BPU kelipatan 10 (nomor, 10, 20, 30, dan seterusnya), ditawarkan untuk menjadi subjek penelitian. Apabila pasien tersebut tidak bersedia, maka ditawarkan kepada pasien dengan nomor urut selanjutnya. Subjek yang terpilih sebagai sampel diberi penjelasan mengenai prosedur penelitian, hak dan kewajiban pasien dalam penelitian yang tertera dalam *informed consent*. Persetujuan subjek mengenai keikutsertaan dan proses penelitian dibuktikan dengan tanda tangan yang tertuang dalam *informed consent*.

Proses pengambilan data dilakukan oleh 25 perawat puskesmas, sebagai petugas administrasi GHQ-12 dan 25 psikolog puskesmas sebagai petugas administrasi SCID-I. Baik perawat maupun psikolog Puskesmas telah mendapatkan pembekalan pengambilan data yang diberikan oleh peneliti dan psikolog senior. Khusus di dalam pembekalan administrasi SCID-I bagi psikolog puskesmas, dilaksanakan pula uji kualitas diagnosis psikolog sebagai petugas pengumpul data. Uji ini penting untuk dilakukan, karena hasil diagnosis psikolog menggunakan SCID-I digunakan sebagai standar baku emas dalam tahap analisis data. Dalam uji ini, para psikolog puskesmas memberikan rating kriteria dan diagnosis terhadap tayangan film yang disiapkan oleh peneliti. Hasil rating para psikolog puskesmas kemudian dibanding-

kan dengan hasil rating psikolog klinis senior, dengan metode *inter-rater reliability*. Dari 25 psikolog puskesmas yang mendapatkan pelatihan, terdapat 2 psikolog yang memiliki nilai *Koefisien Cohen's Kappa* kurang dari 0,75 (mengacu pada Ventura, dkk, 1998), sehingga data dari kedua psikolog tersebut tidak dipergunakan dalam penelitian ini.

Seluruh prosedur dalam pengambilan data menggunakan prosedur *blinding*, yaitu baik perawat maupun psikolog puskesmas saling tidak mengetahui hasil pengisian kuesioner dan wawancara. Oleh karena itu, baik lembar GHQ-12 maupun SCID-I, keduanya dimasukkan dalam amplop terpisah dan tertutup, serta hanya boleh dibuka oleh peneliti. Dahlan (2009) menyebutkan bahwa pada penelitian diagnostik, pengukuran sebaiknya dilakukan secara tersamar (*blinding*). Tujuan penggunaan metode *blinding* adalah menghindari bias pengukuran.

Penelitian ini menggunakan GHQ-12 versi Bahasa Indonesia sebagai instrumen penelitian. Instrumen ini terdiri dari 12 aitem, dengan 4 pilihan jawaban untuk masing-masing aitem. Seluruh jawaban diberi skor dengan menggunakan 3 jenis skoring, yaitu skala bimodal, skala likert, dan skala CGHQ. Adapun perincian skor untuk masing-masing skala dipaparkan pada tabel 1.

Instrumen lain yang digunakan dalam penelitian ini adalah *Structured Clinical Interview for DSM IV Disorder* atau SCID-I

Tabel 1.
Rincian model skoring GHQ-12

Model skoring	Jenis Respon				
	Respon 1	Respon 2	Respon 3	Respon 4	
Skala Bimodal	0	0	1	1	
Skala Likert	0	1	2	3	
Skala CGHQ	0	0	1	1	Khusus untuk aitem positif
	0	1	1	1	Khusus untuk aitem negatif

versi Bahasa Indonesia Januari 2000 V.1.01. Instrumen tersebut sudah diadaptasi dan divalidasi oleh Bagian Psikiatri Fakultas Kedokteran Universitas Indonesia dalam rangka penelitian genetik pada psikosis di Indonesia. Dalam penelitian ini SCID berperan sebagai standar baku emas (*gold standard*) yang menjadi patokan dalam menjamin keakuratan diagnosis psikolog.

Data penelitian diolah dengan bantuan program SPSS dan Mplus. Analisis reliabilitas pendekatan konsistensi internal dengan metode *Alpha Cronbach* digunakan untuk memeriksa tingkat koefisien reliabilitas dari GHQ-12, ditinjau dari masing-masing model skoring (bimodal, likert, dan CGHQ). Analisis struktur konstrak menggunakan metode analisis faktor dengan teknik *Confirmatory Factor Analysis* (CFA) digunakan untuk meninjau model konstrak yang dimiliki oleh GHQ-12. Model yang digunakan sebagai pembanding dibagi menjadi tiga tipe yaitu: (1) model unidimensi, (2) model dua faktor (Idaiani & Suhardi, 2006; Politi, Piccinelli, dan Wilkinson, 1994; Jacob, Nhugra, dan Mann, 1997), (3) model tiga faktor (Graetz, 1991; Farrell, 1998); Ibrahim dkk, 2014). Seluruh model dianalisa dengan menggunakan metode *mean and variance corrected weighted least squares* (WLSMV). Tingkat kecocokan dengan model diukur dengan kriteria yang ditetapkan oleh Yu (2002), yaitu $< .06$ untuk *Root Mean Square Error Approximation* (RMSEA), $\geq .95$ untuk *Nonnormed Fit Index* (NNFI), dan $\geq .96$ untuk *Comparative Fit Index* (CFI). Sementara itu, analisis dengan metode *Receiver Operating Curve* (ROC) digunakan untuk menentukan *Area Under The Curve* (AUC), sensitivitas, spesifisitas dan nilai batas pisah optimum untuk skrining GP. Sedangkan metode analisis *Likelihood Ratio* (LR) digunakan untuk pemeriksaan likelihood ratio pada hasil tes

positif (LR +) dan pada hasil tes negatif (LR -).

Hasil

Dari 251 data yang terkumpul, terdapat 243 subjek yang memenuhi syarat rentang usia dalam penelitian ini, yaitu 18 – 65 tahun, namun hanya terdapat 202 data yang memenuhi syarat untuk dilakukan analisis kualitas diagnosis (ROC dan LR). Hasil analisis deskriptif menunjukkan bahwa rata-rata usia subjek adalah 37,62 dengan mayoritas subjek adalah perempuan (70,80%), bekerja (54,50%), menikah (74,30%), dan memiliki pendidikan SMA (40,10%). Sementara itu, 26 orang (12,9%) terdiagnosis mengalami GP.

Berdasarkan analisis pengestimasi reliabilitas pendekatan konsistensi internal dengan metode *Alpha Cronbach* terhadap GHQ-12, diperoleh hasil seperti yang tertera pada Tabel 2 berikut.

Tabel 2.
Koefisien Reliabilitas GHQ-12

Model Skoring	Koefisien Reliabilitas (r_{xx})
Likert	0,863
Bimodal	0,841
CGHQ	0,832

Berdasarkan ketiga hasil analisis reliabilitas di atas maka dapat disimpulkan bahwa GHQ-12, baik ditinjau dari model skoring Likert, bimodal, maupun CGHQ, memiliki tingkat reliabilitas yang memuaskan (r_{xx} = 0,863; 0,841; 0,832). Tingkat reliabilitas tertinggi dicapai oleh model skoring Likert.

Analisa CFA terhadap 12 aitem GHQ-12 dengan tiga jenis skoring (bimodal, likert, dan CGHQ) menunjukkan adanya kecocokan dengan model teoritik yang berbeda. Baik jenis skoring bimodal maupun CGHQ secara teoritik memiliki tingkat kecocokan yang baik dengan

model yang dimiliki oleh Idaiani dan Suhardi (2006), Politi, dkk (1994), dan Graetz (1991). Hanya saja, tingkat kecocokan yang paling baik untuk jenis skoring bimodal adalah dengan model yang dimiliki oleh Idaiani dan Suhardi (2006), dan Politi, dkk (1994), sedangkan jenis skoring CGHQ memiliki tingkat kecocokan paling baik dengan model teori dari Graetz (1991). Disisi lain, hasil analisa terhadap jenis skoring likert menunjukkan tidak adanya kecocokan dengan model teori manapun. Beberapa model memenuhi nilai kriteria nilai $\geq .96$ untuk CFI dan $\geq .95$ untuk NNFI, akan tetapi tidak satupun model yang memenuhi batas minimal nilai RSMEA yang ditetapkan, yaitu $< .06$. Hasil penelitian ini juga menunjukkan bahwa model dua faktor dan tiga faktor memiliki tingkat kecocokan yang lebih baik bila dibandingkan dengan model faktor tunggal. Hal tersebut nampak dari tidak ada satupun kriteria kecocokan model, baik RMSEA, CFI, maupun NNFI, yang terpenuhi oleh

seluruh jenis skoring GHQ-12 untuk model faktor tunggal. Keterangan lebih lanjut dapat dilihat pada Tabel 3.

Analisis kualitas diagnosis dilakukan dengan dua metode analisis yaitu *Receiver Operating Curve* (ROC) dan *Likelihood Ratio* (LR). Resume seluruh hasil analisis ROC dan LR menggunakan tiga model skoring dapat dilihat pada Tabel 4.

Diskusi

Hasil estimasi reliabilitas dengan metode *Alpha Cronbach* menunjukkan bahwa GHQ-12, memiliki tingkat reliabilitas yang memuaskan, baik ditinjau dari model skoring Likert, bimodal, maupun CGHQ ($r_{xx'} = 0,863; 0,841; \text{ dan } 0,832$). Nilai koefisien reliabilitas tersebut melebihi apa yang disyaratkan oleh Kehoe dalam Ihsan (2004), yang menyatakan bahwa sebuah tes yang kecil (10-15 item) mempunyai koefisien reliabilitas sebesar 0,50 adalah cukup.

Tabel 3.
Hasil *Confirmatory Analysis* terhadap 3 jenis skoring GHQ-12

Model	Bimodal			Likert			CGHQ		
	RMSEA	CFI	NNFI	RMSEA	NNFI	TLI	RMSEA	CFI	NNFI
<i>Unitary</i>	0.065	0.955	0.945	0.152	0.864	0.834	0.086	0.932	0.917
Idaiani dan Suhardi (2006)	0.036	0.987	0.983	0.089	0.954	0.943	0.041	0.985	0.981
Politi, dkk (1994)	0.036	0.987	0.983	0.089	0.954	0.943	0.041	0.985	0.981
Ibrahim, dkk (2014)	0.060	0.964	0.954	0.142	0.888	0.855	0.072	0.955	0.941
Graetz (1991)	0.039	0.985	0.981	0.09	0.955	0.942	0.04	0.986	0.982
Jacob, dkk (1997)	0.067	0.954	0.943	0.154	0.864	0.83	0.086	0.932	0.915
Farrell (1998)	0.062	0.962	0.951	0.146	0.881	0.847	0.078	0.946	0.93

Tabel 4.
Hasil Analisis ROC dan *Likelihood Ratio*

Model Skoring	Titik potong optimum	AUC	Sensitivitas	Spesifisitas	LR +	Prosentase LR +	LR -	Prosentase LR -
Bimodal	≥ 2	76,6%	81%	57%	1,90	13 - 22%	0,34	13 - 5%
Likert	≥ 11	74,1%	81%	62%	2,12	13 - 24%	0,31	13-4%
CGHQ	≥ 4	77%	81%	55%	1,80	13 = 21%	0,35	13-5%

Hasil koefisien reliabilitas yang memuaskan pada GHQ-12 juga ditemukan dalam berbagai penelitian di berbagai populasi di berbagai negara. Kuruvilla, dkk (1999) dalam penelitiannya di pusat layanan kesehatan primer Tamil, India, menemukan nilai *Alpha Cronbach* 0,86 dan nilai Spearman Brown 0,83, sedangkan Montazeri, dkk (2003), dalam penelitian pada populasi umum di Iran, menemukan nilai *Alpha Cronbach* 0,87. Sementara itu Sanchez-Lopez dan Dresch (2008) dan Abeysena, Jayawardana dan Peiris (2012) dalam penelitian pada populasi umum di Spanyol dan Sri Lanka, menemukan nilai *Alpha Cronbach* > 0,70 dan 0,88. Sedangkan Gouveia, Barbosa, Andrade, dan Carneiro (2010), yang meneliti GHQ-12 pada populasi paramedic di Brazil, menemukan nilai *Alpha Cronbach* 0,89. Dengan demikian dapat disimpulkan bahwa berbagai penelitian dari berbagai negara, dengan hasil yang relatif sama tersebut, semakin meneguhkan bahwa GHQ-12 memiliki reliabilitas yang mapan dan memuaskan.

Sementara itu, jika mengacu pada struktur faktor yang dimiliki, banyak penelitian yang berusaha mengungkap apakah GHQ-12 merupakan instrumen pengukuran yang bersifat unidimensi atau multidimensi. Secara praktis, GHQ-12 banyak diperlakukan sebagai instrumen yang unidimensi, akan tetapi banyak penelitian yang menemukan bahwa GHQ-12 terdiri dari dua atau tiga faktor yang berbeda.

Idaiani dan Suhadi (2006) dalam penelitiannya yang dilakukan pada populasi umum di Indonesia, mengidentifikasi adanya dua faktor yang berbeda yaitu distres psikologi dan disfungsi sosial. Penelitian lain yang juga mengidentifikasi jumlah faktor yang sama adalah penelitian Toyabe, dkk (2007) pada penyintas gempa bumi di Jepang, dengan nama faktor

disphoria dan disfungsi sosial, penelitian, Gouveia dkk., (2010) dengan nama faktor depresi dan disfungsi sosial, serta penelitian yang dilakukan oleh Abeysena, Jayawardana dan Peiris (2012), dengan nama faktor depresi dan kecemasan (faktor 1) dan disfungsi sosial (faktor 2).

Disisi lain, penelitian lain yang mengidentifikasi 3 faktor sebagai pembentuk konstruk GHQ-12 antara lain adalah penelitian milik Werneke, dkk (2000) yang mengidentifikasi 3 faktor berbeda dengan nama faktor kecemasan dan depresi, disfungsi sosial, dan *loss of confidence*. Sementara itu, penelitian Younes, Montazeri, Ismail dan Roncin (2009) menemukan 3 faktor yang disebut dengan kecemasan dan depresi, disfungsi sosial, dan kehilangan kepercayaan diri. Sedangkan penelitian yang dilakukan oleh Ibrahim, Osman, Jan, Ismail, Kan, Mukhtar, dan Sidik (2014) pada populasi mahasiswa juga menemukan tiga faktor berbeda yang disebut dengan gejala depresi, harga diri, dan *perceived abilities*.

Sementara itu, hasil analisa CFA penelitian ini menunjukkan bahwa model dua faktor dan tiga faktor memiliki tingkat kecocokan yang lebih baik bila dibandingkan dengan model faktor tunggal. Hal tersebut nampak dari tidak ada satupun kriteria kecocokan model, baik RMSEA, CFI, maupun NNFI, yang terpenuhi oleh seluruh jenis skoring GHQ-12 untuk model faktor tunggal. Hal tersebut senada dengan hasil penelitian dari Graetz (1991) yang menyimpulkan bahwa penggunaan GHQ-12 akan lebih bermanfaat jika menggunakan model multidimensi dibandingkan dengan unidimensi. Graetz (1991) sendiri melakukan pemeriksaan ulang terhadap struktur faktor GHQ-12 dalam sebuah studi yang mana responden diwawancarai dalam jangka waktu interval satu tahun dalam 4 kesempatan yang

terpisah. Hal tersebut ditegaskan kembali oleh Smith, Fallowfield, Stark, Velikova, dan Jenkins (2010) bahwa GHQ-12 merupakan instrumen yang bersifat multidimensi, dan bukan unidimensi.

Bila ditinjau dari jenis skoring yang digunakan, hasil penelitian ini menunjukkan bahwa model yang dimiliki oleh Idaiani dan Suhardi (2006), Politi, dkk (1994), dan Graetz (1991) secara konsisten cocok diterapkan baik untuk jenis skoring bimodal maupun CGHQ. Model yang dimiliki oleh Idaiani dan Suhardi (2006), dan Politi, dkk (1994) merupakan model yang paling cocok diterapkan pada model skoring bimodal. Adapun model teori dari Graetz (1991) paling cocok untuk diterapkan pada jenis skoring CGHQ. Hasil tersebut sekaligus memperkuat temuan dari Campbell, Walker dan Farrel (2003) yang menunjukkan hasil senada. Hanya saja, tidak ada satupun model yang cocok diterapkan pada jenis skoring Likert. Dengan demikian tidak ada satupun model yang cocok diterapkan untuk seluruh lintas jenis skoring, baik bimodal, likert, maupun CGHQ.

Hasil analisis ROC terhadap GHQ-12 dengan model skoring bimodal menemukan nilai AUC sebesar 76% dan batas pisah optimum ≥ 2 , dengan sensitivitas sebesar 0,81 dan spesifisitas 0,57. Sedangkan hasil analisis ROC terhadap GHQ-12 dengan model skoring Likert dan CGHQ menemukan nilai AUC sebesar 74,1% dan batas pisah optimum ≥ 11 , dengan sensitivitas sebesar 0,81 dan spesifisitas 0,62, serta nilai AUC sebesar 77% dan batas pisah optimum ≥ 4 , dengan sensitivitas sebesar 0,81 dan spesifisitas 0,55. Metz (1978) mengkategorikan nilai AUC pada seluruh model skoring tersebut ke dalam kategori nilai diagnostik sedang. Bila mengacu pada pendapat Warner (2004), nilai diagnostic pada kategori

sedang antara lain disebabkan oleh isu aplikasi GHQ-12 pada setting Puskesmas saat terjadinya penelitian. Penggunaan GHQ-12 untuk skrining GP baru pertama kali dilaksanakan di Puskesmas, sehingga baik perawat maupun pasien, masih beradaptasi untuk menerima dan menggunakannya, karena tentu akan merubah sistem manajemen pelayanan pasien.

Meski demikian, bila mengacu pada ketiga hasil di atas maka dapat disimpulkan pula bahwa penggunaan GHQ-12 dengan model skoring CGHQ memiliki nilai diagnostik yang lebih kuat dibandingkan dengan model skoring yang lainnya, untuk mendeteksi GP. Oleh karena itu, dapat direkomendasikan penggunaan model skoring CGHQ bagi skrining GP di Puskesmas. Kesimpulan tersebut dikuatkan oleh saran yang diberikan oleh Dahlan (2009) bahwa sensitivitas yang tinggi cocok untuk tujuan skrining, yaitu "menangkap" sebanyak-banyaknya pasien yang diduga menderita suatu gangguan. Dalam penelitian ini terlihat bahwa model skoring CGHQ memiliki nilai diagnostik yang lebih tinggi dibandingkan dengan model skoring lain.

Sementara itu, hasil penelitian lain terkait dengan skrining gangguan mental di masyarakat, terutama di *setting* pelayanan kesehatan primer, menemukan batas pisah optimum, sensitivitas dan spesifisitas yang berbeda satu sama lainnya. Penelitian Idaiani dan Suhardi (2006) di Indonesia, yang membandingkan antara GHQ-12 dengan SCL-90, memperoleh batas pisah optimum 7/8 bila menggunakan skala Likert, dengan sensitivitas 0,67 dan spesifisitas 0,74. Masih dengan menggunakan standar baku emas SCL-90 ditambah dengan SCID, penelitian Schmitz, dkk (1999) di Jerman, memperoleh batas pisah optimum 1,5 bila menggunakan skoring bimodal, dengan

sensitivitas 0,68 dan spesifisitas 0,65. Sedangkan penelitian Kuruvilla, dkk (1999), di India, yang menggunakan International Clasification of Diseases-10: Primary Care Version (ICD-10 PHC) sebagai standar baku emas, merekomendasikan batas pisah optimum 2/3, dengan sensitivitas 87,4% dan spesifisitas 79,2%.

Dalam ranah metode penelitian psiko-diagnostika, perbedaan nilai batas pisah dalam berbagai penelitian merupakan fenomena yang wajar. Perbedaan tersebut dapat disebabkan oleh jenis alat ukur yang dijadikan sebagai standar baku emas, jenis populasi dan subjek, kondisi budaya setempat, serta latar belakang sosiodemografi lainnya (Idaiani dan Suhadi, 2006). Pendapat tersebut juga ditegaskan oleh Wilson dan Jungner (1968) bahwa tingkat sensitivitas dan spesifisitas instrumen skrining dapat bervariasi satu sama lain, bergantung pada *setting* skrining.

Sementara itu, dari hasil analisis Likelihood Ratio terhadap GHQ-12, dengan menggunakan 3 model skoring, didapatkan hasil LR + 1,90 dan LR - 0,34 untuk skoring bimodal, LR + 2,12 dan LR - 0,31 untuk skoring Likert, dan LR + 1,80 dan LR - 0,35 untuk skoring CGHQ. Dengan demikian, dapat disimpulkan bahwa individu dengan gangguan penyesuaian memiliki kemungkinan antara 1 sampai 2 kali lebih besar untuk memperoleh hasil positif pada GHQ-12. Sedangkan individu dengan gangguan penyesuaian memiliki kemungkinan antara 0,31-0,35 untuk memiliki hasil negatif pada GHQ-12 dibandingkan dengan individu tanpa gangguan penyesuaian. Adapun persentase peningkatan kemungkinan pasien mengalami GP bila memiliki hasil GHQ-12 positif berkisar antara 13 - 24% untuk seluruh model skoring, dan persentase penurunan kemungkinan pasien mengalami GP bila

memiliki hasil GHQ-12 negatif, berkisar antara 4 - 13% untuk seluruh model skoring.

Terkait dengan hasil perhitungan Likelihood Ratio di atas, Akobeng (2007) menyatakan bahwa sebuah instrumen skrining atau diagnostik harus memiliki nilai LR+ ≥ 1 dan LR- ≤ 1 untuk dapat dipakai. Dengan demikian dapat disimpulkan bahwa dalam penelitian ini, GHQ-12 memenuhi persyaratan tersebut.

Kesimpulan

Berdasarkan paparan hasil dan diskusi di atas, maka dapat disimpulkan beberapa hasil sebagai berikut: 1) Berdasarkan analisis CFA, terbukti bahwa GHQ-12 merupakan instrumen pengukuran yang bersifat multidimensi dan bukan unidimensi. Meski demikian, tidak ada satu pun model multidimensi yang secara konsisten cocok diterapkan untuk seluruh jenis skoring, baik bimodal, likert, maupun CGHQ. 2) Berdasarkan analisis kualitas diagnosis ROC dan LR, GHQ-12 terbukti valid dalam menskrining GP di Puskesmas. Ketiga model skoring GHQ-12 memiliki tingkat keakuratan yang berbeda namun ketiganya tetap dapat digunakan sebagai instrumen skrining GP karena memiliki nilai AUC lebih besar dari 70% dan memiliki nilai sensitivitas yang tinggi untuk seluruh titik potong optimum pada setiap model skoring (sensitivitas >80%). 3) Hasil penelitian ini dapat merekomendasikan titik potong optimum untuk masing-masing model skoring, yaitu ≥ 2 untuk model skoring bimodal, ≥ 11 untuk model skoring likert, dan ≥ 4 untuk model skoring CGHQ. Hasil penelitian ini sekaligus juga merekomendasikan model skoring CGHQ sebagai model skoring yang paling tepat digunakan dalam skrining GP di Puskesmas, karena

memiliki nilai AUC yang paling tinggi dibandingkan dengan model skoring yang lainnya.

Saran

Tidak dapat dipungkiri bahwa masih terdapat keterbatasan dalam penelitian ini, yaitu penelitian ini belum sampai pada tahap analisis yang lebih spesifik mengenai sub tipe GP. Oleh karena itu, pada penelitian selanjutnya disarankan untuk mengkaji dan menganalisa data berdasarkan sub tipe GP, agar dapat memberikan elaborasi yang lebih komprehensif dalam strategi deteksi dini dan penanganan GP.

Kepustakaan

- Abeysena, C., Peiris, U., Jayawardana, P., & Rodrigo, A. (2012). Validation of the Sinhala version of 30-item General Health Questionnaires. *International Journal of Collaborative Research on Internal Medicine & Public Health*, 4(7), 1373-1381. doi: 10.4038/jpgim.7859
- Akobeng, A. K. (2007). Understanding diagnostic tests 2: likelihood ratios, pre-and post-test probabilities and their use in clinical practice. *Acta Paediatrica*, 96(4), 487-491. doi: 10.1111/j.1651-2227.2006.00179.x
- Allison, K.R., Adlaf, E.M., Irving, H.M., Hatch, J.L., Smith, T.F., Dwyer, J.J.M., & Goodman, J. (2005). Relationship of vigorous physical activity to psychological distress among adolescents. *Journal of Adolescent Health*, 37, 164-166. doi: 10.1016/j.jadohealth.2004.08.017
- Belek, I. (2000). Social Class, income, education, area of residence and psychology distress: does social class have an independent effect on psychological distress in Antalya, Turkey? *Social Psychiatry and Psychiatric Epidemiology*, 3, 94-101. doi: 10.2307/2095178
- Bell, T., Watson, M., Sharp, D., Lyons, I., & Lewis, G. (2005). Factor associated with being a false positive on the General Health Questionnaire. *Social Psychiatry and Psychiatric Epidemiology*. 40, 402-407. doi: 10.1007/s00127-005-0881-6.
- Berardi, D., Ceroni, G.B., Leggieri, G., Rucci, P., Ustun, B., & Ferarri, G. (1999). Mental, physical, and functional status in primary care attenders. *International Journal of Psychiatry in Medicine*, 29 (2), 133-148. doi: 10.2190/3D0C-QREW-1M5W-VDUU
- Campbell, A., Walker, J., & Farrell, G. (2003). Confirmatory factor analysis of the GHQ-12: can I see that again?. *Australian and New Zealand Journal of Psychiatry*, 37(4), 475-483.
- Carta, M.G., Balestrieri, M., Murru, A., & Hardoy, M.C. (2009). Adjustment disorder: Epidemiology, diagnosis, and treatment. *Clinical Practice and Epidemiology in Mental Health Review*, 5(15), 1-15. doi: 10.1186/1745-0179-5-15
- Casey, P., & Doherty, C.A. (2012). Adjustment disorder: Diagnostic and treatment issues. *Psychiatric Times*, 1, 93-96. doi: 10.1186/1745-0179-5-15
- Center for Public Mental Health (CPMH). (2012). *Laporan Hasil Diskusi Kelompok Terarah Lokakarya Pengembangan Kerjasama Penelitian Aksi untuk Pengembangan Sistem Layanan Kesehatan Mental Masyarakat*. Data mentah yang belum dipublikasikan.
- Dahlan, M.S. (2009). *Penelitian diagnostik: Dasar-dasar teoritis dan aplikasi dengan program SPSS dan stata*. Jakarta: Penerbit Salemba Medika.

- Evans, M., Kessler, D., Lewis, G., Peters, T.J., & Sharp, D. (2004). Assessing mental health in primary care research using standardized scales: Can I be carried out over the telephone? *Psychological Medicine*, *34*, 157-162. doi: 10.1017/S0033291703008055
- First, M.B. (Ed.). (1994). *Diagnostic and statistical manual of mental disorders 4th edition*. Washington, DC: American Psychiatric Association
- Farrell, G.A. (1998). The mental health of hospital nurses in Tasmania as measured by the 12-item General Health Questionnaire. *Journal of advanced nursing*, *28*(4), 707-712. doi: 10.1046/j.1365-2648.1998.00735.x
- Fellinger, J., Holzinger, D., Dobner, U., Gerich, J., Lehner, R., Lenz, G., & Goldberg, D. (2005). Mental distress and quality of life in a deaf population. *Social Psychiatry and Psychiatric Epidemiology*, *40*, 737-742. doi: 10.1007/s00127-005-0936-8
- Gardner, W., Kelleher, K.J., & Pajer, K.A. (2002). Multidimensional adaptive testing for mental health problems in primary care. *Medical Care*, *40* (9), 812-823. doi: 10.1097/01.MLR.0000025436.30093.77.
- Goldberg, D.P., & Blackwell, B. (1970). Psychiatric illness in general practice: a detailed study using a new method of case identification. *British Medical Journal*, *2*(5707), 439-443.
- Gouveia, V.V., Barbosa, G.A., Andrade, E.D.O., & Carneiro, M.B. (2010). Factorial validity and reliability of the General Health Questionnaire (GHQ-12) in the Brazilian physician population. *Cadernos de Saúde Pública*, *26*(7), 1439-1445. doi: <http://dx.doi.org/10.1590/S0102-311X2010000700023>
- Graetz, B. (1991). Multidimensional properties of general health questionnaire. *Social Psychiatry and Psychiatric Epidemiology*, *26*, 132-138. doi: 10.1007/BF00782952
- Gureje, O. (2002). Psychological disorders and symptoms in primary care: Association with disability and service use after 12 month. *Social Psychiatry and Psychiatric Epidemiology*, *37*, 220-224. doi: 10.1007/s00127-002-0544-9
- Ibrahim, N., Osman, Z. J., Jan, K. O. N. N., Ismail, S. I. F., Kar, P. C., Mukhtar, F., & Sidik, S. M. (2014). Reliability and Factor structure of the General Health Questionnaire-12 among university students. *Malaysian Journal of Medicine and Health Sciences*, *10* (2), 53-60. Diunduh dari http://www.medic.upm.edu.my/dokumen/FKUSK1_Jurnal_Text_Article_7.pdf
- Idaiani, S., & Suhardi (2006). Validitas dan reliabilitas General Health Questionnaire untuk skrining distres psikologik dan disfungsi sosial di masyarakat. *Buletin Penelitian Kesehatan*, *34* (4), 161-173. Diunduh dari <http://ejournal.litbang.depkes.go.id/index.php/BPK/article/view/1204>
- Ihsan, H. (2004). *Reliabilitas, validitas konstruk dan validitas prediktif ujian masuk PGSD D-II Universitas Pendidikan Indonesia Bandung* (Tesis Tidak Diterbitkan). Universitas Gadjah Mada, Yogyakarta
- Irmansyah. (2012). *Refleksi, capaian, masalah, dan kebutuhan serta prioritas program kesehatan jiwa tingkat nasional dan masa depan puskesmas dalam penguatan layanan kesehatan jiwa berbasis masyarakat*. Dipresentasikan dalam Lokakarya Pengembangan Kerjasama Penelitian Kasi untuk Pengembangan Sistem Layanan

- Kesehatan Mental Masyarakat. Fakultas Psikologi Universitas Gadjah Mada, Yogyakarta.
- Jacob, K.S., Bhugra, D., & Mann, A.H. (1997). General Health Questionnaire-12 psychometric properties and factor structure among Indian women living in the United Kingdom. *Indian Journal of Psychiatric*, 39 (3), 196-199. doi: 10.4103/0972-6748.132918.
- Kapur, N., Hunt, I., Lunt, M., Mcbeth, J. Creed., & Macfarlane, G. (2004). Psychosocial and illness related predictors of consultation rates in primary care-a cohort study. *Psychological Medicine*, 34, 719-728. doi: 10.1017/S0033291703001223
- Kuruvilla, A., Pothan, M., Philip, K., Braganza, D., Joseph, A., & Jacob, K.B. (1999). The validation of the tamil version of the 12 item general health questionnaire. *Indian Journal of Psychiatry*, 41(3), 217-221. doi: 10.5812/ircmj.11794.
- Makowska, Z., Merecz, D., Moscicka, A., & Kolasa. (2002). The validity of General Health Questionnaires, GHQ-12 and GHQ-28, in mental health studies of working people. *International Journal of Occupational Medicine and Environment Health*, 15(4), 353-362. doi: 10.1016/j.psychres.2010.10.010.
- Mellor, S.G., Brydon, L., & Steptoe, A. (2010) Psychological distress and circulating inflammatory markers in healthy young adults. *Psychological Medicine*, 40, 2079-2087. doi: 10.1017/S003329170710000267
- Metz, C.E. (1978). Basic principles of ROC analysis. *Seminars of Nuclear Medicine*, 4, 283-298. doi: 10.1016/S0001-2998(78)80014-2.
- Mitchell, A.J. (2010). How do we know when a screening test is clinically usefull? Dalam Mitchell, A.J., & Coyne, J.C (Eds), *Screening for Depression in Clinical Practice: a Evidance Based Guide* (pp. 99-112). New York: Oxford University Press. Diunduh dari <http://www/gen.lib.rus>
- Montazeri, A., Harirchi, A.M., Shariati, M., Garmaroudi, G., ebadi, M., & Fatch, A. (2003). The 12-item General Health Questionnaire (GHQ-12): Translation and validation study of Iranian version. *Health and Quality Life Outcomes*, 1 (1), 66. doi: 10.1186/1477-7525-1-66.
- Nitsche, I., Koch, R., & Kallert, T. W. (2010). Which factors influence the psychological distress among relatives of patients with chronic functional psychoses? An exploratory study in a community mental health care setting. *Journal of Public Health*, 18(2), 105-117. doi: 10.1007/s10389-009-0292-3
- Oktaviana, M., & Wimbari, S. (2014). Validasi klinik Strenghts and Difficulties Questionnaire (SDQ) sebagai instrumen skrining gangguan tingkah laku. *Jurnal Psikologi*, 41(1), 101-104. doi: 10.22146/jpsi.6961
- Patel, V., Araya, R., Chowdhary, N., King, M., Kirkwood, B., Nayak, S., Simon, G., & Weiss, H.A. (2007). Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. *Psychological Medicine*, 38, 221-228, doi: 10.1017/S0033291707002334.
- Pevalin, D.J. (2000). Multiple applications of the GHQ-12 in a general population sample: an investigation of long term test effects. *Social Psychiatry and Psychiatric Epidemiology*, 35. 508-512. doi: 10.1007/s001270050272.
- Politi, P.L., Piccinelli, M., & Wilkinson, G. (1994). Reliability, validity and factor

- structure of the 12-item General Health Questionnaire among young males in Italy. *Acta Psychiatrica Scandinavica*, 90(6), 432-437.
- Primasari, I. (2012). *Kumpulan Wawancara Praktisi Kesehatan Mental*. Yogyakarta. Naskah tidak dipublikasikan.
- Razavi, D., Delvaux, N., Farvacques, C., & Robaye, E. (1990). Screening for adjustment disorders and major depressive disorders in cancer in-patients. *The British Journal of Psychiatry*, 156, 79-83. doi: 1192/bjp.156.1.79
- Sanchez-Lopez, M.P., & Dresch, V. (2008). The 12-item General Health Questionnaire (GHQ-12): reliability, external validity and factor structure in The Spanish population. *Psicothema*, 20 (4), 839-843. doi:10.1002/ajmg.b.32158
- Schmitz, N., Kruse, J., Heckrath, C., Alberti, L., & Tress, W. (1999). Diagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. *Social Psychiatry and Psychiatric Epidemiology*, 34, 360-366. doi: 10.1007/s001270050156
- Smith, A.B., Fallowfield, L.J., Stark, D.P., Velikova, G., & Jenkins, V. (2010). A Rasch and confirmatory factor analysis of the General Health Questionnaire (GHQ)-12. *Health Quality Life Outcome*, 8, 45. doi: 10.1186/1477-7525-8-45.
- Strain, J.J., & Friedman, M.J. (2011). Considering adjustment disorder as stress response syndromes for DSM-5. *Depression & Anxiety Review*, 28, 818-823. doi: 10.1002/da.20782.
- Toyabe, S., Shioiri, T., Kobayashi, K., Kuwabara, H., Koizumi, M., Endo, T., Ito, M., Honma., Fukushima, N., Someya, T., & Akazawa, H. (2007). Factor structure of The General Health Questionnaire (GHQ-12) in subjects who had suffered from the 2004 Nigata_chuetsu earthquake in japan: Community Based Study. *BMC Public Health*, 7 (175), 1-17. doi: 10.1186/1471-2458/7/175
- Uher, R., & Goodman, R. (2010). The everyday feeling questionnaire: The structure and validation of a measure of general psychological wellll being and distress. *Social Psychiatry and Psychiatric Epidemiology*, 45, 413-423, doi: 10.1017/s00127-009-0074-9.
- Ventura, J., Liberman, R.P., Green, M.F., Shaner, A., & Mintz, J. (1998). Training and quality insurance with the structured clinical interview for DSM-IV (SCID-I/P). *Psychiatry Research*, 79, 163-173. doi: http://dx.doi.org/10.1016/S0165-1781(98)00038-9.
- Werneke, U., Goldberg, P., Yalcin, L., & Ustun, B.T. (2000).The stability of the factor structure of the General Health Questionnaire. *Psychological Medicine*, 30, 823-829. doi: http://dx.doi.org/
- Wilson, J.M.G., & Jungner, G. (1968). *Principles and Practice of Screening for Disease*. Geneva: World Health Organization.
- Younes, M.S., Montazeri, A., Ismail, A., & Roncin, C. (2009). Factor structure and internal consistency of the 12-item General Health Questionnaire (GHQ-12) and the subjective Vitality Scale (VS), and the relationship between them: a study from France. *Health and Quality of Life Outcomes*, 7 (22), 1-6. doi:10.1186/1477-7525-7-22
- Yu, C.Y. (2002). *Evaluating cutoff criteria of model fits indices for latent variable models with binary and continuous*

- outcomes* (Unpublished doctoral dissertation). University of California, Los Angeles.
- Yusoff, M.S.B., Rahim, A.F.A., & Yacoob, M.J. (2009). The sensitivity, specificity and reliability of the malay version 12-items General Health Questionnaire (GHQ-12) in Detecting Distressed Medical Students. *ASEAN Journal of Psychiatry*, 11 (1). 1-8. doi:10.5959/eimj.2.1.2010.or.