

Metodologi Penelitian Psikodiagnostika

*Rahmat Hidayat*¹

*Indira Primasari*²

Fakultas Psikologi
Univeritas Gadjah Mada

Abstract

The measurements conducted in the discipline of psychology are related to the goals of pure and applied sciences. Measurements taken for pure science is in general carried out within scientific researches. Regarding those in the applied science, there are two methods, namely screening and diagnosis that are commonly used in the practice of clinical measurement. This paper aims to address the comprehensive measurements required in the field of clinical measurement. In particular, it will discuss if psychometric properties only are already sufficient or they are, in fact, not needed in this field. In addition, it will also briefly describe the methodology of measuring the quality of instruments of clinical measurement.

Keywords: screening, diagnosis, clinical measurement

Pengukuran dalam disiplin psikologi dilakukan terkait dengan tujuan-tujuan keilmuan murni dan terapan. Pengukuran dengan tujuan keilmuan murni pada umumnya dilakukan terkait dengan kegiatan penelitian. Tujuan utamanya adalah untuk mendapatkan besaran yang handal (*valid*) dan ajeg (*reliable*) dari sebuah konstruk yang diteliti. Karenanya validitas dan reliabilitas merupakan isu utama pada pengukuran psikologis dalam penelitian. Validitas biasanya ditegakkan berdasarkan keselarasan antara instrumen pengukuran (item, skala) dengan bangunan teoretikal yang dirujuk dalam konstruksi alat. Keselarasan diuji baik dalam tahap pengembangan alat maupun pada tahap penggunaan empirik. Pada tahap pengembangan alat validitas isi diukur salah satunya dengan *judgment content-*

validity ratio (CVR). Sementara itu dalam penelitian lapangan validitas konstruk diuji lagi dengan cara mengukur keselarasan konstruksi teoritik yang telah tercapai pada tahap pengembangan alat dengan pola respon subjek terhadap alat ukur. Analisis *confirmatory analysis* merupakan salah satu instrumen utama untuk menguji validitas konstruk empirik ini.

Properti psikometrika kedua yang tidak kalah penting adalah keajegan alat ukur dalam penelitian. Konsep keajegan menekankan pentingnya hasil pengukuran psikologis dapat dipercaya lintas waktu, dan dalam batas tertentu lintas konteks pengukuran. Dapat dipahami bahwa tingkat keajegan dari hasil pengukuran menentukan tingkat kepercayaan terhadap instrumen tersebut. Manusia cenderung menunjukkan konsistensi dalam perilakunya; atribut-atribut mental-psikologis berfungsi secara ajeg, seperti inteligensi, nilai-nilai, dan dalam batas tertentu sikap manusia terhadap objek-objek di dalam

¹ Korespondensi mengenai isi artikel ini dapat melalui: r.hidayat@ugm.ac.id

² Atau melalui: indira.primasari@hotmail.com atau indira.primasari@gmail.com

dan di luar dirinya sendiri. Tingkat keajegan sebuah instrumen diukur melalui berbagai teknik: *tes-retest*, *parallel forms*, dan *split half*.

Di samping kesahihan dan keajegan, feasibilitas penggunaan alat ukur juga merupakan sebuah kriteria penentuan alat ukur di dalam penelitian psikologi. Parameter-parameter ini, terutama dua parameter yang pertama, telah mendapatkan perhatian yang memadai dalam dunia penelitian dan akademik psikologi di Indonesia. Hampir tidak ada penelitian kuantitatif yang tidak mendiskusikan kedua parameter psikometrika itu di dalam laporannya. Hal ini mengindikasikan tingkat pemahaman yang meluas dan relatif mendalam oleh segenap akademisi psikologi di Indonesia tentang kelengkapan psikometrika tersebut. Ini berbeda dengan tingkat 'keterbiasaan' terhadap properti yang dituntut dari alat pengukuran psikologi yang digunakan di dalam ranah praktik. Ranah ini meliputi asesmen untuk kepentingan tindakan individual, kelompok, maupun organisasional. Salah satu ranah pengukuran terapan adalah skrining dan diagnostika dalam praktek klinik.

Baik pada ranah skrining maupun diagnostika, tindakan pengambilan keputusan menjadi suatu hal yang esensial. Dalam proses pengambilan keputusan, seorang praktisi perlu mempertimbangkan risiko dari berbagai alternatif keputusan, sehingga dapat mengambil keputusan terbaik sesuai dengan kebutuhan di lapangan. Sebagai contoh, seorang praktisi yang akan melakukan skrining perlu mempertimbangkan biaya dan akurasi hasil. Oleh karena itu, sebuah instrumen pengukuran diharapkan mampu memberikan informasi yang secara aplikatif menjamin kualitas pengambilan keputusan. Kebutuhan ini yang tampaknya belum cukup dipenuhi oleh parameter psikome-

trika tradisional. Oleh karena itu, makalah ini bertujuan untuk mendiskusikan kelengkapan pengukuran apa yang dituntut dalam ranah pengukuran klinik. Secara khusus akan didiskusikan apakah properti psikometrika saja memadai? Atau justru properti psikometrika bukan kebutuhan utama dalam ranah pengukuran ini. Di samping itu akan diuraikan secara singkat metodologi pengukuran kualitas instrumen pengukuran klinik.

Ruang Lingkup Pengukuran pada Praktik Klinik

Pada praktik klinik, pengukuran erat kaitannya dengan pengiraan psikologis (*psychological assessment*), yang meliputi kegiatan mengukur, menilai, mengira-ngira, dan memprakirakan perilaku yang akan diperlihatkan oleh seorang individu pada situasi tertentu (Prawitasari, 2011). Dalam proses pengiraan psikologis terdapat beberapa langkah penting, yang saling berkait erat satu dengan lainnya. Sebagai contoh, Prawitasari (2011), membagi proses pengiraan psikologis menjadi lima langkah yang berurutan, yaitu klarifikasi masalah, pemahaman hal ikhwal masalah, pengumpulan data, interpretasi data, dan terakhir, laporan psikologis atau rujukan. Di dalam proses tersebut, pengukuran menjadi salah satu sarana mendapatkan gambaran utuh mengenai profil psikologis individu, sebagai dasar pengambilan keputusan terhadap keadaan atau status psikologis individu.

Secara umum, pengukuran klinis cenderung digolongkan menjadi dua kategori besar, yaitu pengukuran yang dipakai untuk kepentingan skrining dan untuk kepentingan diagnosis. Pengukuran skrining didesain untuk mengidentifikasi kemungkinan gangguan barangkali muncul dan untuk mendorong evaluasi yang

lebih jauh terhadap klien yang memiliki hasil skrining positif. Sedangkan pengukuran diagnosis disusun untuk memberikan jaminan kepada para klinisi bahwa klien benar-benar menderita gangguan (Warner, 2004).

Pengukuran, baik skrining maupun diagnosis, dapat diterapkan dalam lingkup sempit dan luas. Pada lingkup sempit, pengukuran klinis dapat diterapkan dalam praktik konsultasi pribadi di klinik psikologi, maupun dalam skrining kesehatan mental di komunitas, dengan target sasaran yang beragam, seperti pada para penyintas bencana, lansia, korban konflik, anak jalanan, dan penderita penyakit kronis. Sebagai contoh, para psikolog Puskesmas di Kabupaten Sleman dan Yogyakarta terbiasa menggunakan *Beck Depression Inventory* (BDI) dan *Beck Anxiety Inventory* (BAI) dalam praktik individu. Sedangkan Toyabe, Shioiri, Kobayashi, Kuwabara, Koizumi, Endo, Ito, Honma, Fukushima, Someya, dan Akazawa (2007) menggunakan *General Health Questionnaire-12* (GHQ-12) sebagai instrumen skrining distress psikologis pada para penyintas gempa bumi di regio Niigata-Chuetsu, Jepang. Adapun instrumen pengukuran lain yang juga banyak dipakai dalam praktik klinis adalah *Symptom Check List* (SCL-90-R), *Kessler Psychological Distress Scale* (K10), dan *Clinical Assessment Questionnaire* (CAQ).

Di samping lingkup sempit, pengukuran klinis juga dapat diterapkan pada lingkup luas, yaitu pengiraan psikologis dalam praktik profesi psikolog lainnya, di luar ranah klinis. Beberapa contoh penerapan di lingkup luas antara lain, praktik konsultasi pribadi psikolog pendidikan dan psikolog perkembangan, skrining tumbuh kembang anak, dan skrining permasalahan perkembangan anak. Adapun instrumen yang banyak dipakai adalah

Ages and Stages Questionnaire (ASQ), *Denver Developmental Screening Test II* (DDST-II), *Early Screening Inventory-Revised* (ESI-R), *Infant Toddler Developmental Assessment* (IDA), *Hawaii Early Learning Profile* (HELP), *Carolina Curriculum for Infants and Toddlers*, *Peabody Picture Vocabulary Test*, *Rossetti Infant-Toddler Language Scale*, dan *Batelle Developmental Inventory*. Beberapa instrumen, seperti *Strengths and Difficulties Questionnaire* (SDQ) dan *Child Behavior Checklist* (CBCL) bahkan dapat diadministrasikan oleh pengguna selain klinis, seperti orang tua dan guru.

Dalam pengertian sempit dan luas ini pengukuran di ranah klinik erat dengan kepentingan diagnostika. Tujuannya adalah untuk mengidentifikasi dan menganalisis faktor penyebab dan mekanisme yang mendasari permasalahan klien (Haynes, Smith, & Hunsley, 2011; Westhoff, Hagemester, & Strobel, 2007 dalam Groenier, Beerthus, Pieter, Witherman, dan Swinkels, 2011). Hasil dari proses diagnosis adalah sebuah representasi integratif dari keluhan dan permasalahan klien, yang disertai dengan sebuah penjelasan terhadap permasalahan tersebut dan sebuah rencana intervensi (Nelson-Gray, 2003 dalam Groenier, Beerthus, Pieter, Witherman, dan Swinkels, 2011). Oleh karena itu, keakuratan diagnosis menjadi poin utama yang tidak dapat ditawar untuk dapat menyusun rencana intervensi yang memadai. Di samping itu, juga diperlukan informasi tentang kapasitas subjek. Dengan demikian intervensi dirancang berdasarkan masalah yang ada, sesuai dengan potensi atau kapasitas untuk penanganan masalah.

Standar Kualitas Pengukuran Klinik

Dari paparan di atas disimpulkan bahwa pengukuran di ranah klinik berin-

tikan kepentingan diagnostika. Sejumlah standar kualitas dapat dirumuskan, yakni: kesahihan, keajegan, feasibilitas. Ketiganya terlihat sama dengan parameter kualitas pengukuran penelitian yang disebut sebagai properti psikometrika. Namun sebenarnya ada pengertian yang spesifik, sebagaimana akan diuraikan pada bagian-bagian berikut ini.

Kesahihan. Azwar (2003) menyebutkan bahwa validitas berasal dari kata *validity* yang berarti sejauh mana ketepatan dan kecermatan suatu alat ukur dalam melakukan fungsi ukurnya. Suatu instrumen dikatakan valid apabila instrumen tersebut mampu mencapai tujuan pengukuran yang dikehendaki dengan tepat dan mampu memberikan gambaran yang cermat mengenai data yang disampaikannya. Instrumen yang cermat dapat diartikan bahwa instrumen tersebut mampu memberikan gambaran mengenai perbedaan yang sekecil-kecilnya diantara subjek yang satu dengan yang lainnya, suatu instrumen perlu valid agar mampu memberikan informasi yang dapat dipercaya, sehingga dapat diambil keputusan yang tepat.

Di samping instrumen harus mampu memberikan informasi yang akurat, perlu diingat pula bahwa, sebuah penelitian psikodiagnostika dilakukan dengan meniru praktik psikolog sehari-hari dalam menegakkan diagnosis. Dalam konteks tersebut, kualitas nilai diagnosis dari sebuah instrumen menjadi poin yang tidak dapat ditawar. Oleh karena itu, diperlukan pemeriksaan validitas yang berbasis klinis. Dalam pengukuran klinis, sebuah instrumen psikodiagnostika dikatakan sah apabila mampu memberikan hasil pemeriksaan yang dapat dikonfirmasi oleh sebuah prosedur diagnostika yang terpercaya (Dahlan, 2009; Wilson & Jungner, 1968). Secara praktis, instrumen dikatakan sah apabila mampu membedakan antara

klien yang terganggu dengan klien yang tidak terganggu. Lebih jauh lagi, kesahihan sebuah instrumen psikodiagnostika juga terkait dengan seberapa mampu instrumen memberikan hasil pemeriksaan yang positif, bila klien positif mengalami gangguan (*sensitivitas*), dan seberapa mampu instrumen memberikan hasil pemeriksaan yang negatif, bila klien tidak mengalami gangguan (*spesivitas*).

Keajegan. Sebuah instrumen pengukuran psikologi menjadi berguna apabila secara konsisten mampu memproduksi hasil yang kurang lebih sama bila digunakan pada satu orang (Cooley, 2010). Prinsip ini erat kaitannya dengan prinsip reliabilitas dalam pengukuran. Reliabilitas yang tinggi berarti sebuah pengukuran memberikan hasil yang relatif sama dalam kesempatan yang berbeda. Akan tetapi, tidak semua pengukuran akan menghasilkan pengukuran yang relatif sama, selalu ada celah untuk adanya eror. Berkaitan dengan adanya eror dalam konsep reliabilitas, Azwar (2003) menyebutkan bahwa konsep reliabilitas dibagi menjadi dua yaitu reliabilitas alat ukur dan reliabilitas hasil ukur. Reliabilitas alat ukur berkaitan dengan masalah eror pengukuran (*error of measurement*), sedangkan reliabilitas hasil ukur berkaitan dengan eror dalam pengambilan sampel (*sampling error*) yang mengacu kepada inkonsistensi hasil ukur apabila pengukuran dilakukan pada kelompok individu yang berbeda.

Terkait dengan adanya eror dalam pengukuran dan juga penggunaan instrumen di lapangan, Cook dan Beckman (2006) menegaskan bahwa parameter reliabilitas memang diperlukan, namun tidak cukup hanya mengandalkan reliabilitas saja untuk menarik kesimpulan yang valid. Oleh karena itu, hal esensial yang penting dilakukan adalah mengumpulkan bukti-bukti yang memadai dari

berbagai sumber untuk menegakkan reliabilitas sebelum instrumen digunakan dalam praktik. Terlebih lagi mengingat bahwa instrumen yang sama, apabila digunakan dalam setting yang berbeda, dengan subjek yang berbeda pula, dapat memperlihatkan variasi nilai reliabilitas yang luas (Feldt & Brennan, 1989 dalam Cook & Beckman, 2006).

Feasibilitas. Secara praktis, instrumen yang diteliti dalam penelitian psikodiagnostika, tidak hanya akan digunakan oleh klinisi (psikolog, dokter umum, atau psikoter), tetapi juga oleh non klinisi, baik itu di ranah psikologi (ilmuwan psikologi), profesi lain (perawat, bidan, guru, pekerja sosial, dan lain sebagainya), maupun pelaksana yang tidak memiliki keahlian khusus (kader kesehatan, orang tua, dan lain sebagainya). Oleh karena itu, untuk mencapai kebermanfaatan yang optimal, sebuah instrumen dituntut untuk dapat digunakan secara lebih cepat, sederhana, aman, dan mudah, dengan biaya serendah mungkin (Dahlan, 2009; Wilson & Jungner, 1978).

Di samping ketiga parameter diatas, terdapat pula parameter yang lebih spesifik pada dua kategori besar pengukuran klinis; skrining dan diagnosis. Wilson dan Jungner (1968) menyebutkan bahwa instrumen skrining sebaiknya mudah untuk diadministrasikan, dapat diterima oleh pasien, memiliki sensitivitas yang tinggi, dan mampu mengidentifikasi gangguan, baik yang dapat disembuhkan, maupun gangguan yang mana terapi bersifat perbaikan. Sedangkan bagi instrumen diagnosis, meskipun tidak ada instrumen diagnosis yang 100% akurat, namun perlu dibandingkan dengan standar baku emas (*gold standard*). Abroraya, France, Young, Curci, dan LePage (2005) mendefinisikan standar baku emas sebagai standar yang memanfaatkan seluruh validitas

kriteria. Klinisi harus menggunakan seluruh validitas kriteria yang tersedia untuk mencapai diagnosis yang paling akurat. Prosedur diagnostik, yang dijadikan sebagai standar baku emas, dapat berupa instrumen lain, wawancara diagnosis terstruktur, seperti DSM dan ICD, atau wawancara dengan klinisi (Warner, 2004; Rogers, 2001). Meski demikian, apapun instrumen yang dipilih, perlu diasumsikan bahwa instrumen tersebut 100% akurat (Warner, 2004).

Dapat disimpulkan bahwa parameter kualitas instrumen diagnostika sama dengan psikometrika penelitian, yaitu validitas, reliabilitas, dan feasibilitas. Namun terdapat perbedaan spesifik pada masing-masing parameter tersebut sesuai dengan ruang lingkupnya. Yang paling mendasar adalah pada konsepsi validitas pada pengukuran klinik, di mana kesahihan pengukuran ditunjukkan oleh nilai sensitivitas dan spesivitas instrumen berdasarkan standar baku emas yang dipergunakan.

Trade-off antara Spesifisitas dan Sensitivitas

Sensitivitas dan spesifisitas sebuah alat ukur seringkali tidak mandiri satu dari yang lain. Terdapat hubungan yang bersifat *trade-off* antara keduanya. Karakteristik hubungan antara sensitivitas dan spesifisitas terangkum di dalam *Receiver Operating Characteristic Curve* (ROC), yang merupakan sebuah penghitungan statistika untuk menilai akurasi dari sebuah prediksi. Sebuah prediksi dibuat sebelum nilai dari entitas yang diprediksi tersebut dikenal. Oleh karena itu, diperlukan sebuah metode untuk mengevaluasi akurasi dari berbagai prediksi tersebut (Gonen, 2007). Metode ROC pertama kali dikembangkan untuk meningkatkan kemampuan sinyal radar dalam mendeteksi pengebom. Seiring dengan berjalannya

waktu, metode ini kemudian diaplikasikan secara luas dalam berbagai bidang, antara lain pengecekan kecacatan material, pengecekan penghindaran pemasukan pajak, pengecekan kelayakan kredit, prakiraan cuaca, ilmu kedokteran, psikologi eksperimen, dan psikofisik (Swets, 1973; Collinson, 1998; Gonen, 2007).

Dalam bidang psikologi khususnya, salah satu contoh penelitian yang menggunakan ROC sebagai metode analisis statistik adalah penelitian Mickes, Flowe, dan Wixted (2012) mengenai ingatan saksi mata. Dalam penelitian ini Mickes, *et al.* (2012) berhasil membuktikan bahwa ROC terbukti akurat dalam menentukan akurasi prosedur pemeriksaan saksi mata. Hasil penelitian ini juga menunjukkan bahwa prosedur sekuensial memiliki akurasi yang lebih rendah dibandingkan dengan prosedur simultan, dalam membedakan saksi mata yang bersalah dan tidak bersalah. Sementara itu, di Indonesia, ROC juga dipakai sebagai salah satu metode analisis statistik, antara lain oleh Idaiani dan Suhardi (2005), Emeldah (2012), dan Nurwanti (2012), yang kesemuanya meneliti mengenai proses validasi klinik GHQ-12 sebagai instrumen skrining gangguan jiwa.

Perlu disadari bahwa dalam dunia medis dan psikologi, istilah prediksi sekiranya jarang digunakan, namun sebuah diagnosis merupakan sebuah prediksi atas apa yang mungkin salah dengan simptom-simtom dan keluhan-keluhan yang ditampilkan oleh seorang pasien (Gonen, 2007), dan isu akurasi menjadi isu yang menonjol dalam konteks ini. Dalam konteks diagnostika, istilah akurasi direpresentasikan oleh dua hal, yaitu sensitivitas dan spesivitas. Attia (2003) mendefinisikan sensitivitas sebagai proporsi dari jumlah penderita gangguan yang memiliki hasil tes positif. Sedangkan spesivitas didefinisikan

sebagai proporsi dari jumlah orang yang tidak menderita gangguan yang memiliki hasil tes negatif. Sensitivitas dan spesivitas inilah yang menjadi karakteristik utama dari sebuah instrumen diagnostika (Warner, 2004). Di samping sensitivitas dan spesivitas, terdapat pula istilah nilai duga positif dan nilai duga negatif. Nilai duga positif (NDP) mengarah pada seberapa besar hasil positif yang diberikan oleh instrumen sungguh-sungguh positif, sedangkan nilai duga negatif (NDN) mengarah pada seberapa besar hasil negatif yang diberikan oleh instrumen sungguh-sungguh negatif (Dahlan, 2010). Gambaran mengenai jumlah sensitivitas, spesivitas, NDP, dan NDN biasanya terangkum dalam tabel 2x2, seperti yang ada pada Tabel 1.

Tabel 1

Tabel 2x2 Hasil Penelitian Diagnostika

Hasil Tes	Hasil Standar Baku Emas	
	Hasil Positif	Hasil Negatif
Hasil Positif	a True Positif	b False Positive
Hasil Negatif	c False Negatif	d True Negatif

Rumus

$$\text{Sensitivitas} = a / (a + c)$$

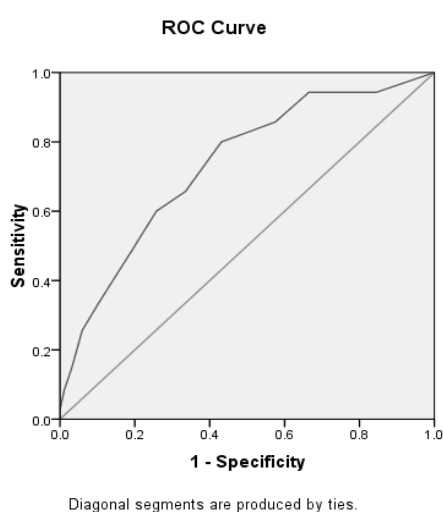
$$\text{Spesivitas} = d / (b + d)$$

$$\text{Nilai duga positif} = a / (a + b)$$

$$\text{Nilai duga negatif} = d / (c + d)$$

Meskipun sensitivitas dan spesivitas merupakan indikator utama dari sebuah instrumen diagnostika, namun nilai keduanya tidak selalu harus berada pada titik absolut dan tetap. Pada praktiknya, terjadi tarik ulur antara nilai sensitivitas dan spesivitas, berdasarkan tujuan penggunaan instrumen diagnostika yang sedang diuji. Semakin tinggi nilai sensitivitas, maka akan diiringi dengan penurunan nilai spesivitas, dan begitu pula sebaliknya. Tarik ulur antara sensitivitas pada berbagai titik potong inilah yang kemu-

dian tergambar dalam kurva ROC (Dahlan, 2010). Selanjutnya pengguna instrumen harus melakukan tawar-menawar (*trade-off*) antara sensitivitas dan spesifisitas dari instrumen tersebut (Fletcher, Fletcher, dan Wagner, 1991; Attia, 2003), dan tawar menawar tersebut biasanya dijadikan sebagai bahan pertimbangan klinisi pada saat menentukan titik potong optimum skoring instrumen psikodiagnostika. Warner (2004) menyebutkan bahwa sebuah instrumen yang ideal akan memiliki satu titik potong dengan nilai sensitivitas 1 (100% berhasil mengidentifikasi kasus yang sesungguhnya) dan nilai spesivitas 1 (100% berhasil mengeksklusi kasus negatif). Pada kenyataannya, instrumen dengan nilai sensitivitas dan spesivitas 1 sulit didapatkan, dan klinisi perlu melakukan tawar menawar antara nilai sensitivitas dan spesivitas sesuai dengan kebutuhan penggunaan instrumen.



Gambar 1. Kurva ROC

Berikut ini adalah contoh kasus yang menggambarkan tawar menawar antara nilai sensitivitas dan spesivitas sebuah instrumen skrining gangguan pemusatan perilaku dan perhatian (GPPH) di sekolah. Hasil analisis ROC terhadap instrumen X menghasilkan titik potong optimum ≥ 5 dengan nilai sensitivitas 0.65 dan spesivitas 0.66. Dari hasil tersebut diketahui

bahwa 34% pasien digolongkan sebagai false negatif (sebenarnya positif menderita gangguan, namun hasil skrining menunjukkan nilai negatif). Secara aplikatif, penggunaan titik potong ≥ 5 tentu akan cenderung merugikan klinisi dalam proses skrining, karena prosentase false negatif yang besar, sehingga menyebabkan hasil skrining menjadi tidak akurat. Apabila klinisi menurunkan titik potong optimum menjadi ≥ 4 , maka nilai sensitivitas yang diperoleh adalah 0.80 dan spesivitas 0.56. Bila dibandingkan dengan titik potong ≥ 5 , tampak bahwa terjadi penurunan nilai spesivitas dan kenaikan nilai sensitivitas pada titik potong ≥ 4 . Disamping itu, hasil false negatif juga mengalami penurunan menjadi 20%. Penggunaan titik potong optimum ≥ 4 pada instrumen X tentu akan lebih bermanfaat secara praktis karena meningkatkan keakuratan hasil skrining.

Metz (1978) menyebutkan terdapat tiga kemungkinan penentuan titik potong, dengan menggunakan ambang pengambilan keputusan yang paling sesuai dengan penggunaan instrumen, yaitu *strict threshold* (sebuah kasus disebut positif bila hasil penilaian menunjukkan hasil yang sungguh-sungguh positif), *moderate threshold*, dan *lax threshold* (sebuah kasus disebut positif bila terdapat kecurigaan bahwa terdapat gangguan dalam kasus tersebut). Sementara itu, Dahlan (2010) menambahkan bahwa penentuan titik potong menggunakan ROC dapat menggunakan dua metode, yaitu secara klinis dan statistik. Penentuan titik potong secara klinis merupakan penentuan titik potong yang ditetapkan oleh klinisi sesuai dengan harapan peneliti akan kepentingan klinis. Instrumen yang akan digunakan sebagai alat skrining direkomendasikan untuk memiliki nilai sensitivitas yang tinggi, sehingga hampir seluruh kasus yang berpotensi mengalami gangguan

akan dapat diidentifikasi dan diases lebih jauh. Sedangkan instrumen yang akan digunakan untuk kepentingan diagnosis cenderung memiliki nilai spesivitas yang lebih besar, untuk menghindari pemberian intervensi yang tidak sesuai (Warner, 2004). Tujuan yang lain adalah untuk menghindari penggunaan biaya intervensi yang tidak rasional.

Di samping penentuan titik potong secara klinis, klinis juga dapat menggunakan penentuan titik potong secara statistik. Dahlan (2010) menyebutkan bahwa metode ini dapat digunakan oleh klinisi saat mengalami kesulitan menentukan titik potong secara klinis. Adapun *software* yang digunakan antara lain adalah *Microsoft Office Excel*. Program ini akan menghasilkan keluaran berupa titik potong optimum berdasarkan kesetimbangan antara sensitivitas dan spesivitas. Metode penentuan titik potong secara statistik juga dapat digunakan klinisi untuk mendapatkan gambaran awal, sebelum akhirnya melakukan penentuan titik potong secara klinis. Proses ini akan mempermudah klinisi dalam memperkirakan kesetimbangan antara sensitivitas dan spesivitas dalam satu titik potong.

Di samping nilai titik potong optimal, analisis ROC juga dapat menghasilkan nilai *Area Under The Curve (AUC)*. Nilai AUC memiliki rentang antara 50% (0.5) sampai dengan 100% (1). Nilai AUC 50% merupakan nilai AUC terburuk, sedangkan nilai AUC 100% merupakan nilai yang terbaik (Dahlan, 2010). Secara lebih detil, interpretasi nilai AUC diklasifikasikan menjadi lima bagian yang berbeda (Metz, 1978), yaitu 0.5 – 0.6 (tingkat akurasi sangat lemah), 0.6 – 0.7 (tingkat akurasi lemah), 0.7 – 0.8 (tingkat akurasi sedang), 0.8 – 0.9 (tingkat akurasi tinggi), dan 0.9 – 1 (tingkat akurasi tinggi). Meski demikian, perlu disadari bahwa sulit bagi klinis

untuk mendapatkan nilai AUC yang sempurna (100%).

Besarnya area AUC akan menunjukkan seberapa benar kesimpulan yang dapat diberikan oleh instrumen diagnostika bila digunakan dalam kepentingan klinis. Sebagai contoh, sebuah instrumen diagnosis depresi menghasilkan nilai AUC sebesar 80%, maka dapat disimpulkan bahwa apabila instrumen tersebut dipergunakan untuk 100 orang pasien Puskesmas, maka jumlah pemeriksaan yang menghasilkan kesimpulan benar dalam menentukan ada tidaknya gangguan depresi, akan didapati pada 80 orang pasien.

Meskipun sensitivitas dan spesivitas menjadi indikator utama dari kegunaan instrumen diagnostika dalam praktik klinik, namun secara aplikatif keduanya masih memiliki keterbatasan. Keterbatasan utama adalah bahwa sensitivitas dan spesivitas merupakan pengukuran yang berbasis populasi, sehingga akan sulit untuk diterapkan dalam interpretasi hasil untuk pasien secara individual (Attia, 2003; Akobeng, 2006). Dengan hanya berpatokan pada sensitivitas dan spesivitas, menurut Attia (2003), akan sulit bagi praktisi untuk mengetahui apakah sebuah hasil tes yang positif atau negatif memang sungguh-sungguh positif, sungguh-sungguh negatif, *false positive* atau *false negatif*.

Salah satu metode yang kemudian direkomendasikan untuk mengatasi keterbatasan tersebut adalah *Likelihood Ratio*. *Likelihood Ratio* (LR) merupakan pengukuran tunggal yang mengkombinasikan antara hasil sensitivitas dan spesivitas. Dalam ranah praktis, penggunaan LR akan lebih berguna karena LR mampu menyediakan ringkasan yang menunjukkan seberapa banyak atau seberapa sedikit kemungkinan klien yang mengalami suatu gangguan memiliki hasil tertentu dibandingkan dengan klien yang tidak meng-

alami gangguan. Di samping itu, LR juga dapat digunakan oleh praktisi untuk mengkalkulasi probabilitas adanya gangguan pada pasien secara individual (Deeks & Altman dalam Akobeng (2006).

Secara umum, LR dapat dipergunakan pada hasil tes positif maupun negatif. Penggunaan LR pada hasil tes positif biasa disebut dengan LR+, sedangkan penggunaan LR pada hasil tes positif disebut dengan LR-. LR+ didefinisikan sebagai probabilitas dari seorang individu dengan gangguan, yang memiliki hasil tes positif dibagi dengan probabilitas dari seorang individu tanpa gangguan yang memiliki hasil tes positif. Sedangkan LR- merupakan probabilitas dari seorang individu dengan gangguan yang memiliki hasil tes negatif, dibagi dengan probabilitas dari seorang individu tanpa gangguan yang memiliki hasil tes negatif (Akobeng, 2006).

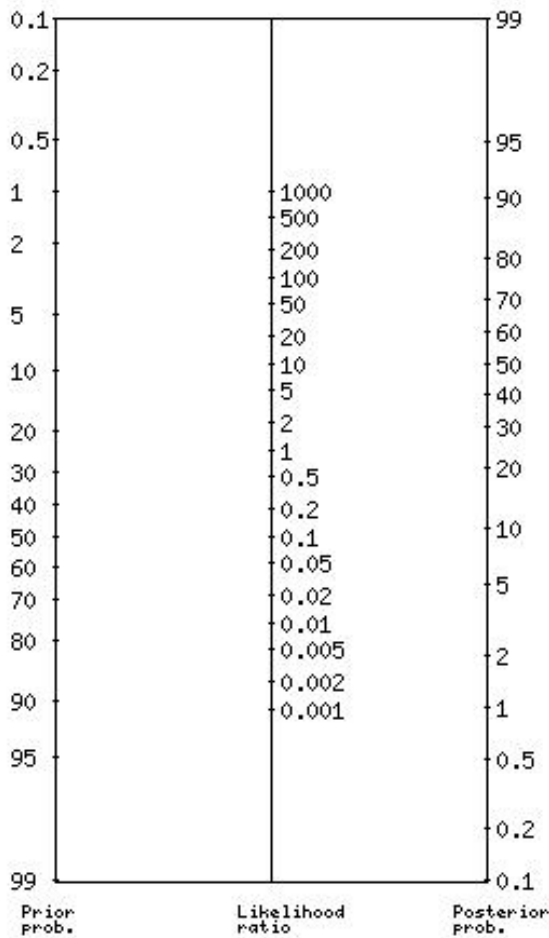
Berikut ini adalah contoh dari perhitungan dan penggunaan LR+ dan LR-. Tes X, sebuah tes yang biasa digunakan untuk mendiagnosis gangguan A, memiliki sensitivitas sebesar 81% atau 0.81 dan spesivitas 62% atau 0.62. Bila dikalkulasi, maka hasil perhitungan LR+ yang didapat adalah $0.81 / 1 - 0.62 = 2.12$. Sedangkan hasil perhitungan LR- yang didapat adalah $1 - 0.81 / 0.62 = 0.31$. Dengan demikian, dapat disimpulkan bahwa (1) pasien dengan gangguan A memiliki kemungkinan 2.12 kali lebih tinggi untuk memiliki hasil tes X positif, daripada pasien yang tidak mengalami gangguan A, (2) pasien yang dengan gangguan A memiliki kemungkinan 0.31 kali lebih rendah untuk mendapatkan hasil tes negatif dibandingkan dengan pasien yang tidak memiliki gangguan A.

Di samping memiliki data LR+ dan LR-, untuk penggunaan LR yang lebih

optimal, penting bagi praktisi untuk melakukan estimasi, yang dapat dipertanggungjawabkan, terhadap ada atau tidaknya suatu gangguan, sebelum tes dilakukan (Attia, 2003; Akobeng, 2006; Warner, 2004). Secara teknis, estimasi terhadap probabilitas keberadaan gangguan, yang dilakukan sebelum tes diberikan kepada pasien disebut dengan *pre-test probability*. Estimasi dalam konteks *pre-test probability* dapat dilakukan praktisi berdasarkan bukti (evidence), seperti prevalensi gangguan dan publikasi ilmiah, dan atau intuisi praktisi setelah melakukan asesmen, yang didasarkan pada pengalaman praktisi (Attia, 2003; Akobeng, 2006; Warner, 2004).

Sementara itu, setelah melakukan tes, seorang praktisi juga perlu melakukan estimasi ulang berdasarkan hasil tes. Estimasi ulang praktisi yang menghasilkan probabilitas pasien dalam menderita gangguan, setelah hasil tes didapatkan disebut dengan *post-test probability*. Proses ini nantinya akan membantu praktisi untuk mengambil keputusan, baik itu dalam menegakkan diagnosis, menyingkirkan diagnosis, atau mengadministrasikan instrumen tes yang lain (Akobeng, 2006).

Untuk mempermudah penggunaan *pre-test probability*, LR, maupun *post-test probability* dalam praktik sehari-hari, praktisi dapat memanfaatkan sarana yang disebut dengan Nomogram Fagan (The Fagan's Nomogram). Nomogram ini dapat memfasilitasi praktisi untuk mengestimasi *post-test probability*, berdasarkan estimasi *pre-test probability* dan LR, tanpa perlu melakukan perhitungan matematika. Secara fisik, nomogram ini terdiri dari tiga aksis, yaitu *pre-test probability* (di sebelah kiri), LR (di tengah), dan *post-test probability* (kanan).



Gambar 2. Nomogram Fagan

Berikut ini adalah panduan penggunaan Nomogram Fagan:

- Estimasi *pre-test probability* dari diri pasien. Estimasi ini biasanya berupa prosentase, dan kemudian tandailah estimasi tersebut pada aksis di sebelah kiri nomogram
- Hitunglah nilai LR + dan atau LR-, dan kemudian tandailah hasil tersebut pada aksis di tengah nomogram.
- Tarik garik lurus yang menghubungkan antara nilai *pre-test probability*, LR, dan *post-test probability* di aksis bagian kanan nomogram, dan berilah tanda di tempat tersebut. Bagian yang diberi tanda merupakan nilai *post-test probability*. Perlu diingat bahwa nilai *post-test probability* dapat lebih tinggi atau lebih rendah dari nilai *pre-test probability*.

Berikut ini adalah contoh aplikasi dari penggunaan Nomogram Fagan di ranah klinik. Seorang laki-laki, berusia 25 tahun, datang ke Balai Pelayanan Umum Puskesmas dengan keluhan merasa sulit tidur, sering sakit kepala, mudah tersinggung, cemas, sulit berkonsentrasi dalam bekerja, dan merasa kurang percaya diri dalam pergaulan. Dari hasil wawancara awal didapatkan informasi bahwa berbagai keluhan tersebut sudah dialami selama tiga bulan sejak bercerai dengan istrinya yang berselingkuh dengan laki-laki lain. Hasil pengisian GHQ-12 menunjukkan skor 15, yang berarti klien memiliki indikasi mengalami gangguan. Dokter yang menangani mengestimasi klien tersebut memiliki kemungkinan mengalami gangguan penyesuaian sebesar 50%. Diketahui bahwa nilai LR+ dari GHQ-12 adalah 2.12, dan apabila ditarik garis lurus antara *pre-test probability* 50%, LR+ 2.12 pada Nomogram Fagan, maka akan didapatkan nilai *post-test probability* sebesar 60%. Terjadi kenaikan nilai kemungkinan klien menderita gangguan penyesuaian sebesar 10%. Oleh karena itu, dokter kemudian merujuk klien tersebut untuk berkonsultasi kepada psikolog Puskesmas.

Daftar Pustaka

- Aboraya, A., France, C., Young, J., Curci, K., & LePage, J. (2005). The validity of psychiatric diagnosis revisited: The clinician's guide to improve the validity of psychiatric diagnosis. *Psychiatric Epidemiology*, 2(9), 48-55. Diunduh dari: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2993536/>
- Akobeng, A. K. (2006). Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. *Acta*

- Paediatrica*, 96, 487-491. Doi: 10.1111/j.1651-2227.226.00179x
- Attia, J. (2003). Moving beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests. *Australian Prescriber*, 26(5), 111-113.
- Azwar, S. (2003). *Dasar – Dasar Psikometri*. Yogyakarta: Pustaka Pelajar Offset.
- Azwar, S. (2003). *Reliabilitas dan Validitas*. Yogyakarta: Pustaka Pelajar Offset.
- Coaley, K. (2010). *An Introduction to Psychological Assessment and Psychometrics*. London: Sage Publication, Ltd.
- Collinson, R. (1998). Statistical note of bombers, radiologists, and cardiologists: time to ROC. *Heart*, 80, 215-217. Doi:10.1136/hrt.80.3.215
- Cook, D. A., & Beckman, T. J. (2006). Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. *The American Journal of Medicine*. 199(2), 166.e7-166.e16. Doi: 10.1016/j.amjmed.2005.10.036
- Dahlan, M. S. (2009). *Penelitian Diagnostik: Dasar-dasar Teoritis dan Aplikasi dengan Program SPSS dan Stata*. Jakarta: Penerbit Salemba Medika.
- Emeldah. (2012). *Validasi klinik General Health Questionnaire-12 sebagai instrumen skrining gangguan kecemasan menyeluruh di Puskesmas*. (Tesis, tidak dipublikasikan). Fakultas Psikologi UGM, Yogyakarta.
- Fletcher, R., Fletcher, S., & Wagner, E. (1991). *Sari Epidemiologi Klinik*. Yogyakarta: Gadjah Mada University Press.
- Gonen, M. (2007). *Analyzing Receiver Operating Characteristic Curves with SAS®*. USA: SAS Institute, Inc.
- Gronieer, M., Berthuis, V. R. J., Pieters, J. M., Witteman, C. L. M., & Swinkels, J. A. (2011). Psychologists' Diagnostic Processes during a Diagnostic Interview. *Psychology*, 2(9), 917-924. Doi:10.4236/psych.2011.29138
- Idaiani, S., & Suhardi, S. (2006). Validitas dan reliabilitas General Health Questionnaire untuk skrining distress psikologik dan disfungsi sosial di masyarakat. *Buletin Penelitian Kesehatan*, 34(4), 161-173. Diunduh dari: <http://www.google.com/url?sa=t&rct=j&q=Validitas%20dan%20reliabilitas%20PMrQfmjYGoBw&usg=AFQjCNFe3Kasrsi5U9Uorx4uw2Z9LDynEQ&bvm=bv.49967636,d.bmk>
- Metz, C. E. (1978). Basic Principles of ROC Analysis. *Seminars of Nuclear Medicine*, 4, 283-298. Diunduh dari: <ftp://norbf.uio.no/pub/outgoing/runeho/KR/Metz78SeminNuclMed8-283.pdf>
- Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver Operating Characteristic Analysis of Eyewitness Memory: Comparing the Diagnostic Accuracy of Simultaneous vs. Sequential Lineups. *Journal of Experimental Psychology Applied*, 18(4), 361-76. Doi: 10.1037/a0030609.
- Nurwanti, R. (2012). *Validasi klinik General Health Questionnaire-12 sebagai instrumen skrining depresi di Puskesmas*. (Tesis, tidak dipublikasikan). Fakultas Psikologi UGM, Yogyakarta.
- Prawitasari, J. E. (2011). *Psikologi Klinik: Pengantar Terapan Mikro dan Makro*. Jakarta: Penerbit Erlangga.
- Rogers, R. (2001). *Handbook of Diagnostic and Structured Interviewing*. New York: The Guildford Press.
- Swets, J. A. (1973). The Relative Operating Characteristics in Psychology. *Science*, 182, 990-1000. Diunduh dari: <http://www.google.com/url?sa=t&rct=j&q=T>

he%20Relative%20Operating%20Char
acteristics

Warner, J. (2004). Clinicians' Guide to evaluating diagnostika and screening tests in psychiatry. *Journal of Continuing*

Professional Development, 10, 446-454.
Doi: 1192/apt.10.6.446

Wilson, J.M.G., & Jungner, G. (1968). *Principles and Practice of Screening for Disease*. Geneva: World Health Organization