



Modul 13

SFS 510-Metodologi Penelitian Fisioterapi II

Materi 13

Pengukuran Validitas

Disusun Oleh

Wahyuddin

UNIVERSITAS ESA UNGGUL

2020

Pendahuluan

Pengukuran validitas adalah komponen penting dari praktek berbasis bukti, untuk memastikan bahwa alat asesmen memberikan informasi akurat untuk pengambilan keputusan. Meskipun klinisi sering menghadapi ketidakpastian dalam manajemen pasien, banyak strategi pengambilan keputusan dapat diterapkan untuk mengurangi ketidakpastian ini.

Validitas Tes Diagnostik

Banyak instrumen pengukuran dirancang khusus sebagai screening atau alat diagnostik. Tes tersebut digunakan untuk menentukan ada tidaknya penyakit atau kondisi abnormal. Tes skrining biasanya dilakukan pada individu yang asimtomatik, untuk mengidentifikasi resiko terhadap gangguan tertentu, dan untuk mengklasifikasikan pasien yang mungkin memperoleh manfaat dari strategi intervensi tertentu. Karena prosedur ini melibatkan alokasi sumber daya, ada potensi risiko terhadap pasien dan digunakan untuk pengambilan keputusan klinis, sehingga sangat penting untuk memastikan validitasnya.

Hasil dari suatu diagnostik atau prosedur screening bisa bersifat dichotomous, kategori atau kontinyu. Tes sederhana akan memiliki hasil dichotomous: positif atau negatif, seperti kehamilan atau status HIV. Tes kategori akan melibatkan peringkat pada skala ordinal, seperti ++, +, -, untuk mencerminkan tingkat sensasi atau refleksi. Skala kontinyu menyediakan informasi mengenai hasil, seperti pengukuran ROM atau tingkat pendengaran. Skala ordinal dan kontinyu sering dikonversi ke dichotomous menggunakan cutoff score untuk menunjukkan respon yang "normal" atau "abnormal".

Reference Standard

Tes diagnostik ideal akan selalu akurat dalam membedakan antara mereka yang dengan dan tanpa penyakit atau kondisi; akan selalu memiliki hasil positif untuk orang dengan suatu kondisi, dan hasil yang negatif pada orang lain.

Kita menentukan seberapa baik suatu tes dengan membandingkan hasil temuan tes diagnostik yang diperoleh dengan reference standards. Reference standard akan

mencerminkan status pasien. Asumsi dibuat bahwa individu yang melakukan tes blinding terhadap kondisi sebenarnya, akan menghilangkan kemungkinan bias.

Sensitivity dan Specificity

Validitas tes diagnostik dievaluasi dalam hal kemampuan untuk secara akurat menilai ada tidaknya kondisi. Tes diagnostik dapat memiliki empat kemungkinan hasil, seperti pada tabel 1. Klasifikasi ditetapkan menurut benar tidaknya penyakit (diagnosis positif atau negatif) dibandingkan dengan hasil tes positif atau negatif.

		Reference Standard True Diagnosis		Total
		<i>Dx+</i>	<i>Dx-</i>	
Test result	Positive	<i>a</i> (True positive)	<i>b</i> (False positive)	<i>a + b</i>
	Negative	<i>c</i> (False negative)	<i>d</i> (True negative)	<i>c + d</i>
Total		<i>a + c</i>	<i>b + d</i>	<i>N</i>

Tabel 1. 2x2 Contingency Table

Sensitivity adalah kemampuan tes mendapatkan tes positif ketika kondisi benar-benar ada. Specificity adalah kemampuan tes yang negatif ketika kondisi ini benar-benar tidak ada. Untuk menggambarkan aplikasi pengukuran tersebut, studi validitas Functional Reach Test (FRT) untuk mengidentifikasi para orang tua dengan penyakit Parkinson yang punya resiko jatuh. FRT dirancang untuk menilai stabilitas anterior-posterior dengan mengukur jarak maksimum meraih. Berdasarkan penelitian sebelumnya, cutoff score 10 inci (25,4 cm) digunakan untuk mengklasifikasikan subjek beresiko atau tidak. Hasil screening dibandingkan dengan riwayat jatuh (reference standard) seperti yang ditunjukkan dalam tabel 2 berikut ini.

		Reference Standard		Total
		Hx of Falls	No Hx of Falls	
FRT Result	Reach <10 in. (+ Fall Risk)	9	1	10
	Reach ≥10 in. (- Fall Risk)	21	12	33
Total		30	13	43

Tabel 2. Hasil FRT Dibandingkan dengan Reference Standard

Sensitivity tes pada populasi ini rendah yaitu 30%. Dari 30 pasien yang diidentifikasi memiliki riwayat jatuh, hanya 9 positif dengan menggunakan FRT. Specificity, 92% dimana dari 13 pasien yang tidak memiliki riwayat jatuh, 12 negatif.

Predictive Value

Selain sensitivity dan specificity, kegunaan dari alat skrining klinis dapat dinilai kelayakannya. Suatu tes harus menunjukkan efisien penggunaan waktu dan sumber daya dan bahwa itu menghasilkan sejumlah respon akurat terkait kemaknaan klinis. Karakteristik ini dinilai dari predictive value. Positive predictive value (PV+) memperkirakan kemungkinan bahwa seseorang yang tes positif benar-benar memiliki penyakit. Oleh karena itu, tes dengan nilai positive predictive value yang tinggi akan memberikan perkiraan jumlah sebenarnya pasien yang memiliki kondisi. Demikian pula, nilai negative predictive value (PV -) menunjukkan kemungkinan bahwa seseorang yang tes negatif benar-benar tidak mempunyai penyakit.

Predictive value merupakan hal yang penting dalam memutuskan apakah mengimplementasikan program screening atau tidak. Ketika positive predictive value rendah, hanya sebagian kecil orang-orang yang tes positif benar-benar memiliki kondisi. Oleh karena itu, banyak sumber daya akan diperlukan untuk mengevaluasi orang-orang dengan false positive. Keputusan sering didasarkan pada keseimbangan antara penggunaan sumber daya yang tersedia dan potensi efek berbahaya yang dihasilkan dari tidak mengidentifikasi orang-orang dengan suatu kondisi.

Likelihood Ratios

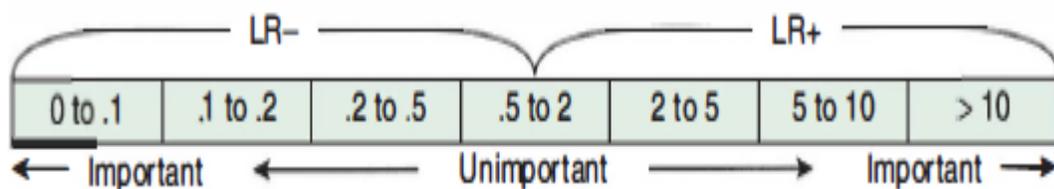
Setelah menetapkan hipotesis bahwa pasien mungkin memiliki diagnosis tertentu, kita ingin menentukan jika tes dapat membuat kita lebih percaya diri dalam menegakkan diagnosis tersebut. Hal ini disebut sebagai likelihood ratios. Likelihood ratios menginformasikan seberapa banyak kemungkinan bahwa seseorang memiliki diagnosis setelah tes dilakukan yang akan membantu menentukan probabilitas posttest. Ini menunjukkan nilai tes untuk meningkatkan kepastian tentang diagnosis atau konfirmasi power. Likelihood ratios sering dilaporkan dalam literatur medis sebagai standar penting untuk praktek berbasis bukti. Likelihood ratios memiliki hal yang lebih besar dibanding sensitivity, specificity, dan predictive values karena independen dari prevalensi penyakit. Kita dapat menentukan likelihood ratio baik positif atau negative dengan perhitungan sebagai berikut:

$$LR+ = \frac{\text{true positive rate}}{\text{false positive rate}} = \frac{\text{sensitivity}}{1 - \text{specificity}}$$

$$LR- = \frac{\text{false negative rate}}{\text{true negative rate}} = \frac{1 - \text{sensitivity}}{\text{specificity}}$$

Interpretasi Likelihood Ratios

Nilai likelihood ratio agak intuitif dimana LR+ yang lebih besar menunjukkan kemungkinan lebih besar adanya penyakit, dan semakin kecil LR- menunjukkan kemungkinan penyakit yang lebih kecil. Nilai-nilai ini telah ditafsirkan sesuai dengan skala berikut:



LR + lebih dari 5 dan LR lebih rendah daripada 0.2 mewakili efek yang relatif penting. likelihood ratio antara 0,2-0,5 dan antara 2 sampai 5 mungkin penting, tergantung pada diagnosis yang sedang dipelajari. Nilai mendekati 1.0 mewakili efek penting. Likelihood ratio 1.0 pada dasarnya berarti tes ini berguna.

Clinical Prediction Rules

Pada bagian sebelumnya telah kita dijelaskan bagaimana sensitivity, specificity dan konsep-konsep yang terkait dapat digunakan untuk mendukung klasifikasi diagnostik atau prognostik berdasarkan skor tes tertentu. Dalam praktek klinis, kompleksitas kondisi pasien sangat penting sehingga kombinasi dari prediktor-prediktor dapat digunakan untuk mendukung hasil klasifikasi. Meskipun pengalaman klinis sering memberikan intuisi dimana temuan dari riwayat pemeriksaan fisik penting untuk akurasi penilaian.

Clinical prediction rules (CPR) adalah alat untuk mengkuantifikasi kontribusi dari variabel yang berbeda untuk diagnosis, prognosis atau respon treatment pada individu pasien. Tujuan CPRs adalah untuk mengurangi ketidakpastian dengan menunjukkan bagaimana kluster tertentu dari temuan klinis dapat digunakan untuk memprediksi hasil.

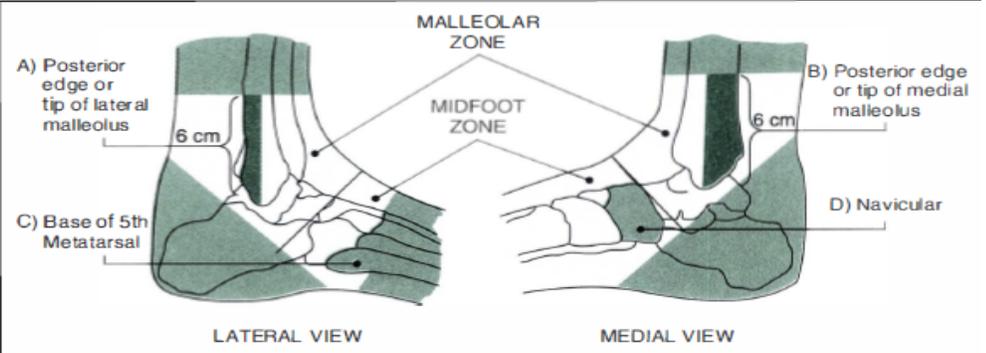
Diagnosis

Salah satu penerapan CPR adalah untuk membantu dalam penegakan diagnosis gangguan yang didasarkan pada tanda-tanda klinis. Misalnya pengembangan CPR untuk penggunaan radiografi pada pasien dengan cedera pergelangan kaki akut. Mereka mencatat bahwa banyak pasien dengan cedera pergelangan kaki tidak mengalami fraktur. Ini dikenal sebagai Ottawa Ankle Rules seperti yang ditampilkan pada tabel 3 berikut ini:

For a patient with an ankle or foot injury:

An ankle X-ray is required only if there is pain in the malleolar zone AND any one of the following:

- Bone tenderness along the distal 6 cm of the posterior edge of the fibula or tip of the lateral malleolus OR
- Bone tenderness along the distal 6 cm of the posterior edge of the tibia or tip of the medial malleolus OR
- Inability to bear weight for four steps at the time of injury and when examined.

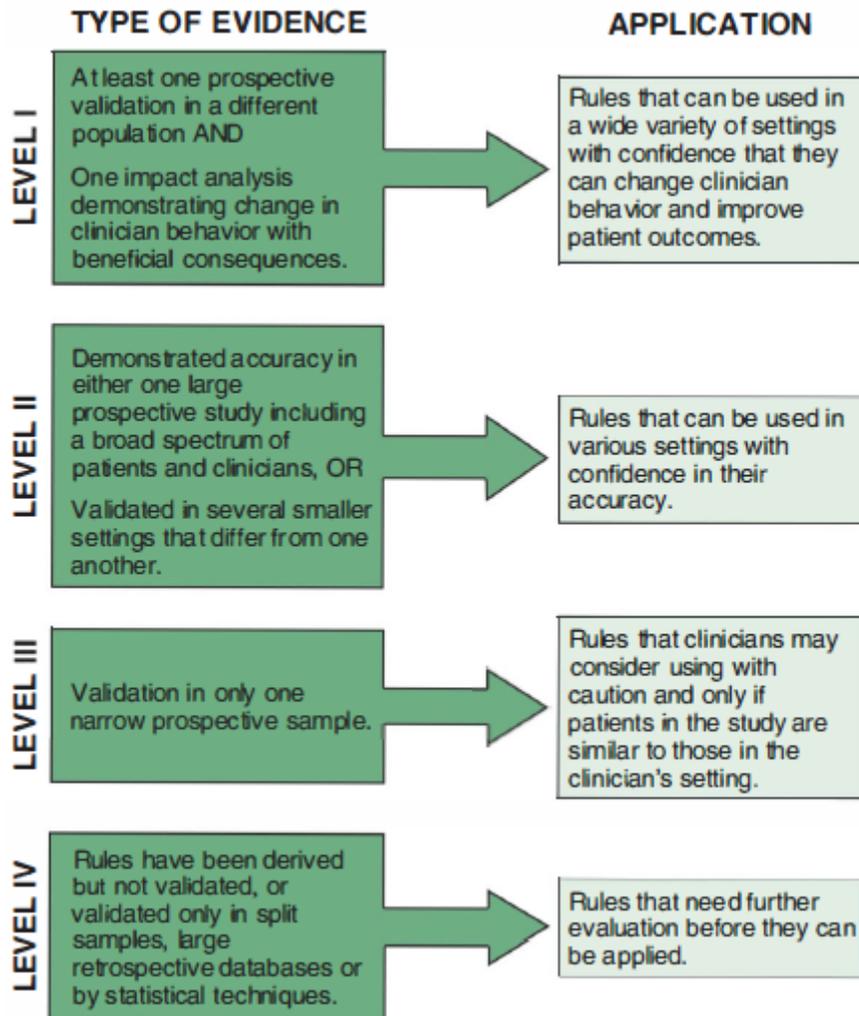


A Foot X-ray is required if there is pain in the midfoot AND

- Bone tenderness at the base of the 5th metatarsal OR at the navicular OR
- Inability to bear weight for four steps at the time of injury and when examined.

Tabel 3. Ottawa Ankle Rules

Tinjauan sistematis dari Ottawa Ankle Rules telah menunjukkan sensitifitas 95-100%, dengan negative likelihood ratio 0.08. Validasi CPR dapat dilakukan melalui langkah-langkah berikut:



Gambar 1. Prosedur Validasi Clinical Prediction Rules

Prognosis dan Resiko

Clinical prediction rules juga dapat digunakan untuk menentukan tingkatan mana individu berada pada risiko tertentu. Sebagai contoh, terdapat pengembangan CPR untuk mengidentifikasi orang dewasa yang mempunyai risiko mengalami diabetes tipe 2. Mereka awalnya berasal cross-sectional cohort, dan kemudian divalidasi menggunakan a prospective cohort di komunitas. Contoh lain adalah identifikasi risiko uterkait penurunan fungsional pada wanita yang lebih tua yang mengidentifikasi pasien risiko komplikasi setelah bedah kardiak, faktor-faktor yang menyebabkan rawat inap dengan asma, dan identifikasi pekerja untuk kembali bekerja dalam kondisi kesehatan yang baik.

Kesimpulan

Mencari makna dalam memahami angka dasar statistik untuk pengukuran validitas sangat penting terkait pengambilan keputusan berbasis bukti. Beberapa perkiraan validitas mudah diakui sebagai metode yang tepat untuk penilaian, seperti yang digunakan untuk mengukur akurasi diagnostik. Hal lain seperti metode untuk mengevaluasi perubahan masih berkembang dan akan selalu berkembang lebih lanjut seiring waktu.

Banyak penulis telah menggunakan istilah "minimal detectable difference" dan "minimal clinically important difference" sebagai istilah-istilah sinonim, meskipun kedua definisi tersebut telah dengan jelas didefinisikan dan dibedakan. Ini adalah dua acuan penting jika kita ingin benar-benar memahami pengukuran yang kita lakukan.

Penilaian klinis mengenai validitas pengukuran harus didasarkan pada beberapa kriteria yang relevan untuk pasien tertentu. Setiap studi tertentu akan menyajikan data dari sampel yang memiliki properti tertentu dan yang telah dipelajari dalam konteks tertentu selama jangka waktu tertentu. Klinisi harus menilai bahwa informasi untuk menentukan jika itu tepat diterapkan untuk pasien mereka. Nilai-nilai memungkinkan kita untuk memprediksi tanggapan terhadap pasien untuk pengambilan keputusan.

Terkait dengan pemahaman kita tentang validitas, kita akan mengkaitkan dengan definisi signifikansi klinis. Praktisi berbasis bukti akan mendapatkan keuntungan dari lebih lengkap pelaporan likelihood ratios, effect sizes dan nilai-nilai perubahan minimal dalam studi klinis. Bila mungkin, confidence interval harus digunakan untuk mencerminkan nilai-nilai populasi. Perkiraan diperlukan untuk pengaturan berbeda, kelompok usia, durasi penyakit dan kondisi pada baseline. Klinisi, pasien dan analis kebijakan kesehatan semua ingin menghargai berapa banyak lebih baik itu benar-benar "lebih baik."

Referensi

Sharma PT, Sieber FE, Zakriya KJ, Pauldine RW, Gerold KB, Hang J, et al. Recovery room delirium predicts postoperative delirium after hip-fracture repair. *Anesth Analg* 2005;101:1215-1220.

Cahan A, Gilon D, Manor O, Paltiel O. Clinical experience did not reduce the variance in physicians' estimates of pretest probability in a cross-sectional survey. *J Clin Epidemiol* 2005;58:1211-1216.

Crichlow RJ, Andres PL, Morrison SM, Haley SM, Vrahas MS. Depression in orthopaedic trauma patients: Prevalence and severity. *J Bone Joint Surg Am* 2006;88:1927-1933.

Stingone JA, Claudio L. Asthma and enrollment in special education among urban school children. *Am J Public Health* 2006;96:1593-1598.

Van den Ende J, Moreira J, Basinga P, Bisoffi Z. The trouble with likelihood ratios. *Lancet* 2005;366:548.

Center for Evidence-Based Medicine. Likelihood ratios. Available at: http://www.cebm.net/likelihood_ratios.asp.

Simon S. Children's Mercy Hospital and Clinics. Likelihood ratio slide rule. Available at: <http://www.childrens-mercy.org/stats/sliderule.asp>.

Center for Evidence-Based Medicine. Nomogram. Available at: <http://www.cebm.net/nomogram.asp>.

CEBM Stats Calculator. Available at: <http://www.cebm.utoronto.ca/practise/ca/statscal/>.

The STARD statement and standards for reporting of diagnostic accuracy studies. Available at: <http://www.stard-statement.org/website%20stard/>.

Kanaya AM, Fyr CL, de Rekeneire N, Shorr RI, Schwartz AV, Goodpaster BH, et al. Predicting the development of diabetes in older adults: The derivation and validation of a prediction rule. *Diabetes Care* 2005;28:404-408.

Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. *Arch Phys Med Rehabil* 2005;86:1753-1762.

Tseng YL, Wang WT, Chen WY, Hou TJ, Chen TC, Lau FK. Predictors for the immediate responders to cervical manipulation in patients with neck pain. *Man Ther* 2005;11:306-315.

Fritz JM, Childs JD, Flynn TW. Pragmatic application of a clinical prediction rule in primary care to identify patients with low back pain with a good prognosis following a brief spinal manipulation intervention. *BMC Fam Pract* 2005;6:29.

Leshner JD, Sutlive TG, Miller GA, Chine NJ, Garber MB, Wainner RS. Development of a clinical prediction rule for classifying patients with patellofemoral pain syndrome who respond to patellar taping. *J Orthop Sports Phys Ther* 2006;36(11):854-866.

Childs JD, Cleland JA. Development and application of clinical prediction rules to improve decision making in physical therapist practice. *Phys Ther* 2006;86:122-131.

Quintana JM, Escobar A, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. *Osteoarthritis Cartilage* 2005;13:1076-1083.

Beninato M, Gill-Body KM, Salles S, Stark PC, Black-Schaffer RM, Stein J. Determination of the minimal clinically important difference in the FIM instrument in patients with stroke. *Arch Phys Med Rehabil* 2006;87:32-39.

de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 2006;4:54.

Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther* 2006;86:735-743.

Fragala-Pinkham MA, Haley SM, Goodgold S. Evaluation of a community-based group fitness program for children with disabilities. *Pediatr Phys Ther* 2006;18:159-167.