



Modul 4

FEB 326-Evidence-Based Practice Fisioterapi

Materi 4

Kajian Kritis

Rumusan Pertanyaan Penentuan Akurasi Tes

Untuk Penegakan Diagnosis

Disusun Oleh

Wahyuddin

UNIVERSITAS ESA UNGGUL

2020

A. Pendahuluan

Kesalahan diagnosis merupakan hal yang sering terjadi dalam praktik klinis. Untuk meminimalisir situasi tersebut, pertanyaan tentang akurasi diagnostik tes harus didasari pada spesifikasi problem pasien dan diagnostik tes yang akan digunakan.

Penentuan diagnosis merupakan langkah penting terkait penentuan kondisi pasien. Dengan penentuan diagnosis yang akurat dan spesifik akan mengarahkan pada penentuan prognosis. Hal ini selanjutnya akan berdampak kepada pemberian intervensi terbaik untuk mendapatkan hasil *outcomes* yang terbaik.

B. Kompetensi Dasar

Kompetensi dasar yang harus dimiliki oleh mahasiswa dan mahasiswi pada materi ini adalah pemahaman tentang aspek kritis pada pertanyaan-pertanyaan studi terkait penentuan diagnosis dalam fisioterapi.

C. Kemampuan Akhir Yang Diharapkan

Setelah mempelajari materi ini, diharapkan semua mahasiswa dan mahasiswi mampu memahami jenis-jenis studi terkait penentuan akurasi tes untuk penegakan diagnosis berupa observasi klinis, cross-sectional, randomized trial, screening, dan systematic review

D. Uraian Materi

Proses Diagnostik Dalam Fisioterapi

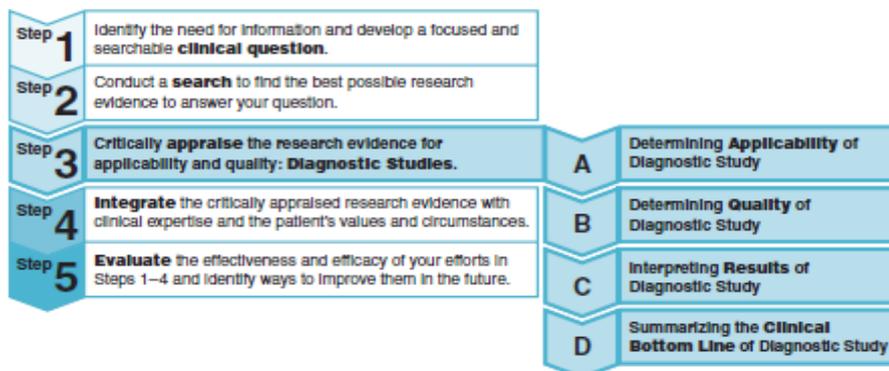
Proses diagnostik dalam fisioterapi meliputi riwayat pasien, *review* sistem serta informasi tentang penggunaan tes dan pengukuran. Tahapan tersebut merupakan langkah ketiga yang dilakukan setelah pada langkah pertama berupa identifikasi kebutuhan informasi dan pengembangan pertanyaan-pertanyaan klinis yang terfokus dan pada tahapan kedua dapat dilakukan pencarian informasi terkait pertanyaan-pertanyaan tersebut.

Dalam proses kritisi *evidence* terkait pengaplikasian dan kualitas terkait studi diagnostik, beberapa langkah-langkah yang dapat dilakukan antara lain:

1. Menentukan kemampuan aplikasi studi diagnostik.

2. Menentukan kualitas studi diagnostik.
3. Menginterpretasikan hasil studi diagnostik.
4. Merangkum *clinical bottom line* dari studi diagnostik.

Tahapan tersebut yang dikenal sebagai proses *appraisal* digambarkan pada gambar 1 berikut ini:



Gambar 1. Proses Appraisal Studi Diagnostik

Pemahaman *evidence* terkait tes diagnostik dan mengintegrasikan hasil tes diagnostik, pengalaman klinis, dan tujuan pada pasien adalah faktor-faktor yang esensial dalam praktik. Mengkomunikasikan hasil proses diagnostik kepada pasien dan profesi lain dilakukan melalui pemahaman pengetahuan tes diagnostik yang valid dan reliabel serta interpretasi yang tepat terhadap tes-tes tersebut. Fisioterapis mengeksaminasi pasien dan mengevaluasi hasil pemeriksaan, menegakkan diagnosis, penentuan prognosis dan perencanaan tindakan. Proses diagnostik telah banyak didiskusikan dalam fisioterapi. Ketika melakukan proses diagnostik, kita telah menentukan intervensi yang akan diberikan serta kebutuhan untuk merujuk kepada profesi lain.

Adanya akses langsung terhadap pelayanan fisioterapi berdampak kepada peningkatan tanggung jawab secara professional. Salah satu bentuk tanggung jawab tersebut adalah akurasi diagnosis kondisi dalam ruang lingkup praktik fisioterapi dan rujukan yang tepat kepada profesi lain. Sebelum akses langsung, fisioterapis mungkin mengandalkan suatu preskripsi atau rujukan dari dokter. Rujukan dokter berlanjut sebagai sumber rujukan utama untuk fisioterapis, tetapi mayoritas rujukan tersebut tidak menyebutkan diagnosis secara spesifik. Sebagai contoh, rujukan mungkin akan ditulis dengan "nyeri

elbow”, “peningkatan kapasitas aerobik” atau “*post-stroke*” tanpa disertai diagnosis yang mengarah kepada etiologi masalah atau kaitan terhadap gerak.

Hal tersebut menjadi tanggung jawab fisioterapis untuk mengoreksi diagnosis terkait masalah gerak, dan jika diperlukan untuk merujuk kepada profesi lain. Salah satu prinsip fundamental pada setiap tes diagnostik yang digunakan fisioterapis adalah hal tersebut dapat membedakan antara mereka yang mengalami masalah gerak dengan yang tidak. Walaupun terdengar berlawanan, hal tersebut lebih kompleks. Jika seseorang mengalami radang tenggorokan, dokter mungkin menggunakan kultur antigen terkait bakteri streptococcus untuk mendiagnosa adanya infeksi. Kita mungkin berasumsi bahwa hasil diagnostik tes tersebut akan menentukan apakah ada radang atau tidak. Namun demikian, setiap diagnostik tes mempunyai rentang nilai apakah nilai tersebut positif atau negatif. Tes cepat untuk mendeteksi adanya infeksi sekitar 80-90 persen, yang berarti ada kesalahan antara 10-20 persen pada kasus. Tes tersebut mungkin mengindikasikan tidak ada infeksi meskipun terdapat kasus (*false-positives*). Tidak ada tes diagnostik yang mempunyai tingkat akurasi sampai 100 persen baik pada *ruling in* atau *ruling out* pada penegakan suatu diagnosis.

Pemberi pelayanan kesehatan menggunakan kultur *rapid* antigen tenggorokan bersamaan dengan riwayat dan gejala-gejala klinis untuk menentukan apakah memerlukan obat anti infeksi dan jika diperlukan dengan menggunakan tes-tes yang dibutuhkan. Setiap pertanyaan dalam riwayat, setiap observasi selama reuvisi sistem, dan setiap pilihan suatu tes atau pengukuran harus diarahkan oleh hipotesis awal yang terkait masalah. Karena itu, pencarian informasi harus diarahkan dari hal yang bersifat umum ke khusus. Kita dapat menentukan jenis tes dan pengukuran yang membantu agar terdapat kejelasan tentang informasi yang dibutuhkan. Sebagai contoh, pertimbangkan seorang pasien dengan diagnosis *non-specific shoulder pain*. Pasien melaporkan riwayat dan gejala yang konsisten misalnya dengan *labral tear*, *supraspinatus tear*, atau *supraspinatus tendonitis*. Hipotesis kita akan terkait dengan persoalan gerak yang berasal dari informasi pasien yang kemudian dites dengan melakukan pemeriksaan. Anterior slide test yang positif mungkin akan menunjang hipotesis adanya *labral tear*. Hasil dari setiap tes yang dipilih seharusnya secara langsung relevan apakah dengan *ruling in* atau *ruling out* dari suatu tes khusus. Seseorang

yang melakukan penegakan diagnostik akan menanyakan seberapa jauh suatu tes dengan hasil yang ada akan membantu untuk mengambil keputusan secara klinis?

Untuk membantu memahami jawaban dari pertanyaan tersebut, kita harus memahami tes yang valid dan reliabel secara klinis. Dalam contoh pada kasus di atas, kita harus mempertimbangkan pertanyaan bagaimana akurasi anterior slide test terkait *ruling in* atau *ruling out* pada kasus *labral tear*? Jawaban pertanyaan ditentukan dari hasil riset terkait tersebut.

Bagaimana kita mendapatkan informasi yang baik tentang akurasi suatu tes diagnostik? Kita dapat memperoleh jawaban tersebut dari observasi klinis atau literatur penelitian.

Observasi klinis

Untuk mendapatkan informasi tentang akurasi suatu tes diagnostik kita harus mengaplikasikan tersebut pada pasien-pasien dan kemudian lihat bagaimana temuan terhadap respon yang dapat memberikan akurasi untuk menegakkan diagnosis yang tepat. Hal tersebut mungkin dilakukan dalam praktik klinis yang bersifat rutin, tetapi lebih sering tidak ada aturan untuk mendapatkan estimasi yang tidak bias terkait akurasi diagnostik tes dalam rutinitas praktik klinis.

Mengapa hal tersebut terjadi? Dalam praktik klinis yang bersifat rutin, diagnosis yang tepat mungkin diperoleh dari investigasi berkelanjutan. Sebagai contoh, seorang klinisi terkesan dengan adanya *rotator cuff tear* yang didasarkan pada O'Brien's test yang dikonfirmasi dengan arthroscopy. Umumnya informasi tentang diagnosis yang tepat tidak tersedia secara rutin karena tidak semua pasien merupakan subjek yang mendapatkan investigasi berkelanjutan. Sebagai konsekuensi, observasi klinis terkait tes klinis dan ketepatan diagnosis hampir selalu berdasarkan (kadang-kadang jumlah sedikit) dari pasien yang dites.

Sejauh ini kemungkinan akurasi tes diagnostik mungkin lebih tinggi atau lebih rendah pada sub grup tersebut. Hal ini berdampak pada observasi klinis terkait akurasi tes diagnostik mungkin akan bersifat *underestimate* atau *overestimate* terkait akurasi tes tersebut. Ini menjadi hal yang menyulitkan untuk menyimpulkan secara umum (generalisasi) akurasi diagnostik tersebut berdasarkan observasi klinis saja yang tidak terstruktur. Estimasi yang lebih baik

tentang akurasi tes diagnostik mungkin diperoleh dari riset klinis berkualitas tinggi.

Riset Klinis

Cross-sectional studies

Seperti pada observasi klinis, studi klinis tentang akurasi tes diagnostik meliputi pengaplikasian pada banyak pasien dan kemudian menentukan seberapa baik temuan dari tes tersebut dengan penegakan diagnosis yang tepat. Beberapa disain studi seperti *cross-sectional studies* untuk membedakan dengan *longitudinal-studies*. Studi untuk menentukan akurasi tes diagnostik adalah *cross-sectional studies* karena konsen terhadap seberapa akurat suatu tes untuk menentukan apakah ada suatu penyakit saat dilakukan tes. Pada *cross-sectional studies* terkait diagnostik tes, suatu grup subjek dilakukan tes. Hal ini dikenal sebagai tes klinis. Subjek yang sama juga dites dengan beberapa tes lain untuk menentukan diagnosis yang tepat. Tes yang digunakan sebagai pembanding untuk menentukan diagnosis yang tepat sering disebut sebagai standar emas (*gold standard test*) atau standar referensi (*reference standar test*).

Reference standard test lebih sering dilakukan dengan tes yang bersifat *invasive* atau lebih mahal dibandingkan dengan tes klinis yang dilakukan pada subjek. Sebagai contoh, studi yang membandingkan antara temuan pada tes klinis yang sederhana untuk lumbosacral nerve root compression pada pasien dengan sciatica (tes klinis meliputi pertanyaan-pertanyaan apakah nyeri lebih dirasakan pada tungkai dibandingkan punggung bawah, *straight leg raise test*, kelemahan, tidak ada tendon refleks dan lain sebagainya) dengan temuan pada magnetic resonance imaging (MRI). Pada studi tersebut MRI digunakan sebagai *reference standard*.

Kadang-kadang *reference standard* ditinjau berdasarkan pada hal-hal yang sudah terjadi (*hindsight*) karena diagnosis yang tepat sering terlihat seiring dengan perjalanan waktu. Sebagai contoh studi terkait *red flag* yang digunakan pada asesmen *low back pain*. Tanda-tanda *red flag* seperti penurunan berat badan secara tiba-tiba mungkin mengindikasikan adanya kanker. Tetapi ada ketidakpuasan menggunakan *reference standard* untuk diagnosis kanker pada mereka yang mengalami nyeri punggung bawah. Hal tersebut dimungkinkan

karena beberapa jenis kanker tidak mudah dideteksi, walaupun dengan menggunakan alat diagnostik yang bersifat *invasive* atau mahal.

Diagnosis mungkin dapat ditegakkan beberapa waktu kemudian ketika tahapan penyakit sudah lanjut. Dalam kasus tersebut, *reference standard* mungkin meliputi monitor pada pasien dalam jangka waktu lama. Diagnosis yang tepat pada waktu tes di awal mungkin dipertimbangkan sebagai kanker jika *follow-up* dalam jangka waktu lama secara sekuensis mendeteksi kanker, dan diagnosis dipertimbangkan sesuatu selain kanker jika *follow-up* dalam jangka waktu lama

Dua jenis *cross-sectional study* dapat dibedakan berdasarkan bagaimana peneliti melakukan rekrutmen (*sampling*) subjek pada studi. Jenis studi yang paling bermanfaat untuk mencari sampel dari populasi subjek yang diduga sesuai dengan diagnosis. Sebagai contoh, studi tentang tes klinis *lumbosacral nerve root compression* seperti yang dijelaskan pada pembahasan sebelumnya. Studi tersebut merekrut subjek dengan nyeri punggung bawah yang mengalami nyeri menjalar sampai ke tungkai karena populasi yang diduga terdiagnosis dengan *lumbosacral nerve root compression*. Seperti yang diharapkan, MRI digunakan untuk mengkonfirmasi adanya *lumbosacral nerve root compression* dalam beberapa subjek. Hal tersebut mengindikasikan bahwa tidak semua positif dengan *lumbosacral nerve root compression* ketika menggunakan MRI. Diagnosis yang tepat pada setiap subjek tidak diketahui sampai semua subjek terlibat langsung dalam studi.

Studi lain dikenal sebagai *cohort studies*. Dalam suatu pendekatan alternatif, satu grup terdiri dari subjek yang diduga terdiagnosis suatu penyakit dan grup lain terdiri dari subjek yang tidak terduga suatu penyakit. Sebagai contoh, investigasi akurasi *Phalen's test* untuk mendiagnosis patologi *carpal tunnel syndrome* (CTS) dengan merekrut 2 grup subjek. Satu grup yang secara klinis dan terkonfirmasi dengan *electromyography* positif terdiagnosis dengan CTS, dan grup lain adalah sukarelawan yang tidak mengalami keluhan di tangan. Tujuan studi oleh para peneliti adalah untuk membedakan secara akurat antara dua grup subjek. Studi tersebut sering juga disebut sebagai *case-control studies* karena disain dengan membandingkan antara mereka yang positif mengalami suatu kondisi berdasarkan diagnosis (*case*), dan kelompok lain sebagai pembanding (*control*). Walaupun kedua istilah *cohort studies* dan *case-control*

studies pada contoh di atas sering membingungkan untuk memasukkan pada kategori yang benar-benar ideal karena adanya tumpang tindih.

Randomized trials

Secara teori kita dapat menggunakan *randomized trials* untuk menginformasikan kepada kita tentang manfaat tes diagnostik. *Trials* tidak memberikan penjelasan terkait akurasi diagnostik tes, tetapi lebih kepada bagaimana efek penggunaan suatu tes diagnostik terhadap perubahan *outcomes* pada subjek/pasien. Prinsip penggunaan *randomized trials* adalah untuk menginvestigasi efek tes diagnostik secara sederhana. Subjek dialokasikan secara acak ke grup, apakah menerima atau tidak suatu tes diagnostik dan kemudian *oucomes* dari kedua grup tersebut dibandingkan.

Jika tes memberikan informasi tentang akurasi diagnostik yang mendukung keputusan yang lebih baik dalam penanganan, ini akan merefleksikan *outcomes* yang lebih baik pada grup yang diberikan tes tersebut. Dengan kata lain, jika tes diagnostik tidak memberikan informasi akurat, atau jika memberikan informasi akurat tetapi informasi tersebut tidak berkontribusi terhadap penanganan yang lebih baik grup yang menerima tes tidak akan memberikan *outcomes* yang lebih baik. Beberapa randomized trial telah dilakukan untuk menentukan nilai x-ray pada primary care terhadap pasien dengan keluhan nyeri punggung bawa. Pasien dibagi dalam dua grup yaitu grup pertama mereka yang rutin dirujuk untuk melakukan x-ray dan kelompok lain tidak dilakukan hal yang sama. Kemudian dilakukan perbandingan *outcomes* pada kedua grup tersebut. *Outcomes* yang diukur meliputi tingkat disabilitas, konsultasi medis secara berkelanjutan dan biaya perawatan

Screening

Kita dapat mendifferensiasikan dua jenis tes diagnostik. Jenis tes pertama adalah tes dimana kita mempertimbangkan jika suatu tes diaplikasikan ketika seseorang mengalami suatu masalah dan menggunakan tes tersebut untuk menegakkan diagnosis yang dapat menjelaskan masalah pada orang tersebut. Jenis tes kedua adalah *screening test*. *Screening tests* adalah tes yang diaplikasikan pada orang yang tidak punya alasan khusus terkait dugaan adanya suatu diagnosis.

Screening mungkin dapat bersifat *practice-based*, misalnya pada mereka yang mengalami nyeri punggung bawah yang kemudian dilakukan *screening* terkait depresi yang mungkin terjadi. *Screening* dapat juga bersifat *community-based programme* misalnya di beberapa negara dilakukan *screening scoliosis* pada anak-anak sekolah usia muda yang merupakan bagian dari *school-based screening programmes*.

Nilai potensial dari suatu *screening* adalah kemungkinan dapat mendeteksi adanya penyakit. Dan untuk beberapa penyakit, deteksi dini dapat mengarahkan kepada penanganan yang lebih efektif. Evaluasi terbaik pada program *screening* dilakukan dengan *randomized trial* karena dapat memberikan keuntungan. Tes *screening* akan memberikan hasil yang menunjukkan efek manfaat jika kapabel dan secara akurat mendeteksi kondisi. Dan jika deteksi dilakukan pada tahap awal, intervensi akan lebih efektif.

Umumnya *randomized trial* dari prosedur diagnostik telah dilakukan pada tes *screening* medis. Beberapa contoh adalah efek *mammogram screening* untuk kanker payudara serta PAP smears untuk kanker. Clinical trials dari tes *screening* biasanya dilakukan dalam suatu studi dengan jumlah sampel yang sangat besar sehingga sering berbiaya sangat mahal. Hal ini berbeda dengan yang dilakukan dalam lingkup fisioterapi. Hanya ada sedikit studi *randomized trials* terkait diagnostik atau tes *screening* yang dilakukan. Ketika *randomized trials* telah dilakukan, banyak fisioterapis melanjutkan dengan *screening* untuk mengetahui ada tidaknya evidence terkait efek manfaat. Misalnya di beberapa negara pada anak-anak sekolah dilakukan *screening* untuk mengetahui adanya disfungsi cerebral.

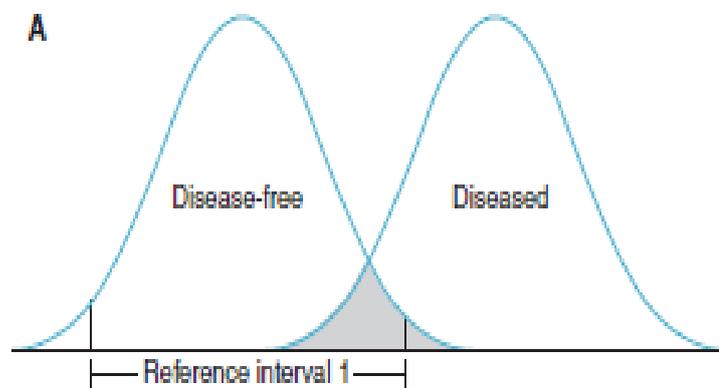
Systematic reviews

Dalam beberapa tahun terakhir, banyak studi *systematic reviews* telah dipublikasikan, misalnya *systematic review* tentang *anterior cruciate ligament injury*, *Ottawa ankle rules* dan *carpal tunnel syndrome*. Seperti halnya *systematic reviews* terkait intervensi atau prognosis, *systematic reviews* tentang akurasi tes diagnostik potensial memberikan asesmen yang transparan dan tidak bias tentang akurasi tes, secara potensial memberikan presisi estimasi tes akurasi, sehingga secara potensial memberikan sumber informasi terbaik tentang akurasi tes diagnostik tersebut. Hal tersebut juga dijumpai dalam *Cochrane Collaboration*

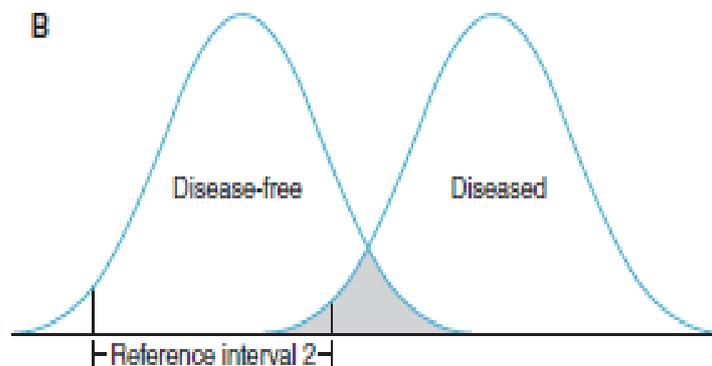
yang telah mempublikasikan *systematic reviews* terkait akurasi tes diagnostik dalam *Cochrane Library*. Misalnya meliputi *review* tes klinis untuk *radiculopathy* (van der Windt et al 2010) dan *red flags* untuk *vertebral fracture* pada mereka yang mengalami nyeri punggung bawah (Henschke et al 2010).

Reference intervals

Tes diagnostik mempunyai tipikal apa yang dikatakan oleh Riegelman sebagai *reference intervals*. *Reference intervals* adalah rentang skor yang menggambarkan individu tanpa masalah gerak. Individu dengan masalah gerak akan berada di luar *reference intervals* tersebut. Tentu saja, tidak semua tes diagnostik berada dalam rentang distribusi normal seperti yang digambarkan pada gambar 3 dan 3 berikut ini:



Gambar 2. *Reference Interval 1*



Gambar 3. *Reference Interval 2*

Aspek penting terkait gambar 2 dan 3 di atas adalah adanya tumpang tindih dalam distribusi pada mereka yang mengalami dan tidak mengalami

persoalan gerak. Area abu-abu menggambarkan variasi yang diharapkan pada mereka yang mengalami persoalan gerak atau tidak. Daerah ini akan mempengaruhi tes diagnostik dan manfaatnya dalam proses diagnostik. Jika kita gagal mengkategorikan seseorang dengan masalah gerak secara spesifik, kemudian perencanaan mengikuti kategorisasi tersebut maka diagnostik mungkin tidak membantu pasien.

Jika kondisi pasien tidak meningkat akibat pemilihan intervensi, kedua komponen baik intervensi dan diagnosis awal harus dievaluasi. Riegelman mencatat bahwa mengetahui bagaimana hidup dengan ketidakpastian adalah gambaran sentral dari justifikasi klinis yang baik. Hal ini akan memberikan pedoman bahwa diagnosis adalah suatu proses dimana fisioterapis menggunakan laporan-laporan pasien, tes dan pengukuran yang valid dan reliabel, pengalaman klinis, dan evaluasi yang konstan terkait progres pasien.

Pertanyaan-Pertanyaan Diagnostik Di Klinis

Penegakan suatu diagnosis dalam fisioterapi berasal dari integrasi proses eksaminasi dan evaluasi. Diagnosis yang tepat akan membentuk pola sebagai dasar pemilihan intervensi dan potensial prognosis terhadap pasien. Namun demikian, banyak tes yang digunakan dalam praktik fisioterapi mungkin tidak dievaluasi sepenuhnya terkait kualitas dan reliabilitas tes tersebut.

Kualitas suatu tes ditentukan oleh tujuan spesifik dan sering merupakan pertimbangan dari masalah spesifik pada pasien. Untuk setiap tes diagnostik yang digunakan dalam pemeriksaan, kita harus mengetahui jawaban dari pertanyaan-pertanyaan sebagai berikut:

1. Apa informasi diagnostik yang kita cari?
2. Apakah suatu tes membantu kita mengkonfirmasi atau menolak hipotesis klinis terkait problem gerak pada pasien?
3. Bagaimana hasil suatu tes mempengaruhi rekomendasi treatment?
4. Apa yang kita ketahui tentang karakteristik tes yang kita gunakan?
5. Apa yang harus kita ketahui tentang karakteristik tes yang kita gunakan?

Terkait perbandingan dengan reference standard, beberapa hal yang perlu diperhatikan adalah perbandingan reference standard dengan tes lain

untuk menilai akurasi sebagai pembanding dengan objek studi. Serta yang tidak kalah pentingnya adalah penilaian akurasi didasarkan pada validitas tes tersebut.

Salah satu cara untuk meningkatkan validitas adalah dengan melakukan *blinding*. Hal ini karena hasil dapat menjadi bias akibat adanya harapan/ekspektasi asesor. Ekspektasi ini mengarah kepada kecenderungan untuk menginterpretasikan hasil sesuai dengan *reference standard* yang ada. Sehingga pertimbangan akurasi tes diagnostik harus ditentukan apakah tes klinis dan *reference standard* dilakukan secara *blinding*.

Terkait relevansi *evidence*, interpretasi akurasi dari suatu tes diagnostik mempunyai kesamaan dengan interpretasi studi terkait efek terapi dan prognosis. Dengan demikian perlu pertimbangan subjek studi terkait kriteria untuk penarikan kesimpulan studi. Di sisi lain hal ini akan membutuhkan kemampuan skill tester terkait implementasi dan pengalaman klinis untuk interpretasi hasil. Tester harus mempunyai kemampuan level training dan keahlian (sertifikasi) tertentu terkait studi yang akan dilakukan. Pertimbangan lain adalah *setting* pengambilan data apakah dilakukan dalam skala kecil misalnya di rumah sakit atau klinis atau dalam skala besar misalnya di suatu komunitas). Perlu diperhatikan bahwa proses aplikasi dan interpretasi tes diagnostik bersifat probabilitas.

Pencarian Literatur

Disain Penelitian Pada Tes Diagnostik Spesifik

Cohort design adalah studi disain yang paling sering digunakan untuk mengembangkan suatu tes diagnostik. Prinsip yang sama terkait pencarian literatur berlaku sama untuk pencarian literatur diagnostik. Namun demikian, jenis disain yang digunakan dalam studi tes diagnostik dan pengukuran adalah *cohort*, bukan *randomized clinical trials (RCTs)*.

Prospective cohort design lebih disarankan daripada *retrospective design* untuk mengevaluasi tes diagnostik. Meskipun mungkin untuk menggunakan *retrospective design*, disain tersebut mengalami banyak masalah khususnya terkait dengan validitas informasi. Suatu contoh hipotesis *retrospective design* akan membantu menjelaskan lebih jauh terkait kebutuhan untuk menggunakan *prospective design*.

Sebagai contoh peneliti ingin mengetahui hasil *Lachman test* yang secara umum digunakan di klinik untuk mendiagnosis kecurigaan adanya *anterior cruciate ligament (ACL) tear* yang hasilnya sama jika dibandingkan dengan penggunaan MRI pada pasien. Kita sering menggunakan rekam medis pasien untuk menentukan hasil kedua tes tersebut. Namun demikian, tidak semua mempunyai catatan tentang hasil pemeriksaan dengan *Lachman test*, hasil tes tidak tercatat, atau ketidakjelasan pada pencatatan tersebut. Sebagai tambahan, konsistensi antar klinisi yang melakukan tes baik pada *Lachman test* atau MRI, membuat hal tersebut menjadi tidak mungkin untuk menentukan penanganan yang tepat dan reliabel.

Prospective studies pada tes diagnostik direncanakan untuk mengumpulkan data baik tes klinis (*Lachman test*) atau tes pembanding sebagai *gold/reference standard* (MRI). Pengembangan suatu tes diagnostik yang valid dan reliabel memerlukan dua hal yaitu *index test* dan *gold standard test* pada setiap subjek. Dalam suatu penelitian, prosedur pengumpulan data harus dilengkapi untuk memenuhi kebutuhan desain studi yang valid, misalnya terkait reliabilitas tester pada kedua tes tersebut. Peserta studi setuju untuk berpartisipasi pada kedua prosedur sebagai bagian dari penelitian. Gambaran hal tersebut dapat tercapai melalui *prospective study*. Subjek dapat diukur pada satu titik waktu (*prospective and cross-sectional*) dan data diperoleh dari satu waktu yang singkat atau pengukuran pada satu waktu diikuti dengan studi jangka panjang untuk menentukan target kondisi yang terkonfirmasi (*prospective and longitudinal*).

Aspek Statistik Studi Diagnostik

Beberapa hal terkait aplikasi statistik pada studi diagnostik antara lain:

1. Sensitivity: proporsi pasien dgn kondisi kedua tes baik tes diagnostik maupun *reference standard* positif.
2. Specificity: proporsi pasien dgn kondisi kedua tes baik tes diagnostik maupun *reference standard* negatif.

Kedua komponen di atas merupakan parameter statistik penggambaran validitas tes diagnostik terkait temuan tes terhadap *gold/reference standard*. Dengan demikian dapat digunakan untuk menentukan diagnosis yang tepat. Pada *prognostic clinical prediction rules*

dan *intervention clinical prediction rules (PCPRs/ICPRs)*, *reference standard* adalah *outcome* pasien setelah pemberian treatment yang dikaitkan dengan periode waktu.

Ada empat kemungkinan hasil pada aspek *sensitivity* dan *specificity* sebagai berikut:

1. *True positive* jika hasil tes diagnostik positif dan *reference standard* juga positif.
2. *True negative* jika tes diagnostik negative dan *reference standard* juga negative.
3. *False positive* jika hasil tes diagnostik positif, sebaliknya *reference standard* negatif.
4. *False negative* jika tes diagnostik negative, sebaliknya *reference standard* positif.

Secara matematik, keempat komponen di atas dibuat dalam model 2x2 *contingency table* pada tabel 1 sebagai berikut:

	Reference Standard Positive	Reference Standard Negative
Diagnostic Test Positive	true positive	false positive
Diagnostic Test Negative	false negative	true negative

Sensitivity	Specificity
$\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$	$\frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$

Tabel 1. 2x2 Contingency Table

Dalam contoh perbandingan Lachman test dengan MRI sebagai gold standard dapat dilihat pada tabel 2 berikut ini:

TABLE 5.1 Table Comparing Two Test Results

	GOLD STANDARD POSITIVE	GOLD STANDARD NEGATIVE
Clinical test Positive	A True positive	B False-positive
Clinical test Negative	C False-negative	D True negative

Table cell A = Patients with positive Lachman and positive MRI tests are true positives
 Table cell B = Patients with positive Lachman and negative MRI tests are false-positives
 Table cell C = Patients with negative Lachman and positive MRI tests are false-negatives
 Table cell D = Patients with negative Lachman and negative MRI tests are true negatives

Tabel 2. Perbandingan *Lachman Test* dan MRI

Untuk perhitungan dapat diberikan contoh sebagai berikut pada tabel 3 di bawah ini:

TABLE 5.3 Subject Values Summed for Each Test

	MRI POSITIVE	MRI NEGATIVE
Lachman test Positive	A True positives n = 39	B False-positives n = 10
Lachman test Negative	C False-negatives n = 3	D True negatives n = 8
	<i>Sensitivity</i> A + A/C 39/39 + 3 = 0.93	<i>Specificity</i> D/D + B 8/8 + 10 = 0.44

Tabel 3. Perhitungan Nilai *Sensitivity* dan *Specificity*

Satuan untuk nilai *sensitivity* dan *specificity* adalah dalam bentuk persen atau desimal. Misalnya *sensitivity & specificity Canadian cervical spine rules* = 100% & 43% (1 & 0.43). Nilai sensitivitas tes yang tinggi mengindikasikan bahwa tes tersebut ideal untuk *screening* pasien pada suatu patologi/kondisi yang memerlukan referral medis. Contoh interpretasi *sensitivity Canadian cervical spine rule* 100% menginformasikan adanya probabilitas fraktur cervical spine.

Perhitungan lain adalah dengan *likelihood ratio*. Perhitungan ini menunjukkan kuantifikasi arah dan *magnitude* kecenderungan probabilitas individu baik positif atau negatif terkait *reference standard*. Ada dua jenis *likelihood ratio* yaitu:

1. *Positive likelihood ratio*, jika kedua tes baik tes diagnostik maupun *reference standard* hasilnya positif.
2. *Negative likelihood ratio*, jika kedua tes baik tes diagnostik maupun *reference standard* hasilnya negatif.

Untuk perhitungan ditunjukkan pada gambar 4 dan 5 berikut ini:

$$LR+ = \frac{\text{Sensitivity (true positive rate)}}{1 - \text{Specificity (true negative rate)}}$$

Gambar 4. Rumus Perhitungan *Positive Likelihood Ratio*

$$LR- = \frac{1 - \text{Sensitivity (false-negative rate)}}{\text{Specificity (true negative rate)}}$$

Gambar 5. Rumus Perhitungan *Negative Likelihood Ratio*

Berikut ini salah satu contoh artikel studi terkait komponen-komponen *sensitivity*, *specificity*, dan *positive/negative likelihood* pada tabel 4 berikut ini:

TABLE 5. Sensitivity, specificity, and likelihood ratios (LR) of the McMurray's test with confidence intervals (CI).

Study	Sensitivity (%) CI (%)			Specificity (%) CI (%)			LR+ (CI 95%)	LR- (CI 95%)
	Medial & Lateral combined = meniscal tear	Medial meniscus	Lateral meniscus	Medial & Lateral combined = meniscal tear	Medial meniscus	Lateral meniscus		
Akseki et al ³	63 (55-71)	67 (59-75)	53 (45-61)	83 (77-89)	69 (62-76)	88 (83-93)	3.71 (3.19-4.13)	0.45 (0.39-0.52)
Anderson & Lipscomb ⁵	58* (48-68)			29* (20-38)			0.82 (0.5-1.3)**	1.45 (0.4-4.9)**
Boeree & Ackroyd ¹⁹	27* (21-33)	29.3 (23-36)	25 (19-31)	89* (85-93)	87.3 (83-92)	89.8 (86-94)	2.31*	0.81*
Corea et al ⁴	58.5 (48-69)	65 (55-74)	52 (41-62)	93.4 (88-98)	93 (88-98)	94 (88-99)	8.86 (7.17-10.91)	0.44 (0.36-0.54)
Evans et al ²³	33* (24-42)	16 (9-23)	50 (40-60)	96* (92-100)	98 (95-100)	94 (89-99)	8.33*	0.70*
Fowler & Lubliner ²²	29* (22-36)			96* (93-99)			7.25*	0.74*
Karachalios et al ²¹		48	65		94	86	8.00 4.64	0.553 0.40
Kurosaka et al ⁶	37 (30-44)			77 (70-84)			1.61*0.82*	
Manzotti et al ²⁰		88	79		50	20	1.76 0.98	0.24 1.05
Noble & Erat ²⁵	63 (56-70)			57 (50-64)			1.50 (1.1-2.1)**	0.60 (0.5-0.9)**
Sae-Jung et al ²⁴	70.6 (55-82)	70	68.2	82.49 (55-95)	60.7	47.8	4.00 (1.4-11.3)	0.358 (0.22-0.56)

When confidence intervals are not present, the CIs were incalculable due to absence of raw data.

** No raw data available to calculate CIs values taken from Solomon et al²¹.

*CIs calculated using formulas from Sackett et al¹³.

Tabel 4. Aplikasi *Sensitivity, Specificity, Positive and Negative Likelihood Ratio*

Dalam penentuan nilai terkait pengambilan keputusan berdasarkan hasil perhitungan *positive and negative likelihood ratio* digambarkan pada tabel 5 berikut ini:

TABLE 5.5 Likelihood Ratios Impact on Post-Test Probabilities

POSITIVE LIKELIHOOD RATIOS		NEGATIVE LIKELIHOOD RATIOS	
Large	>10	Large	<0.1
Moderate	5 to 10	Moderate	0.1 to 0.2
Small	2 to 5	Small	0.2 to 0.5
Negligible	<2	Negligible	0.5 to 1.0

Data from: Cleland J. *Orthopedic Clinical Examination: An Evidence-Based Approach for Physical Therapists*. Carlstadt, NJ: Icon Learning Systems; 2005; and Jaeschke T, Guyatt GH, Sackett DL. How to use an article about a diagnostic test: B. what are the results and will they help me in caring for my patients? *JAMA*. 1994;271:703-707.

Tabel 5. Standar Nilai Dan Dampak *Positive and Negative Likelihood Ratio*

Lebih jelas hal tersebut dapat dilihat perbandingan nilai dalam pengambilan keputusan pada tabel 6 berikut ini:

TABLE 7. Comparison of likelihood ratio's for McMurray's test with modified tests.

Study	LR	McMurray's			Modified		
		General	Medial meniscus	Lateral meniscus	General	Medial meniscus	Lateral meniscus
Akseki et al ¹	LR+	3.71 (3.19-4.13)			5.15 (4.48-5.93)		
	LR-	0.45 (0.39-0.52)			0.38 (0.33-0.44)		
Anderson & Lipscomb ²	LR+	0.82 (0.5-1.3)**			4.8 (0.8-30.0)		
	LR-	1.45 (0.4-4.9)**			0.4 (0.2-0.6)		
Karachalios et al ³	LR+		8.0	4.64		26.8 (14-51)	23.0 (13-37)
	LR-		0.55	0.40		0.11 (0.06-0.18)	0.08 (0.02-0.2)
Kurosaka et al ⁴	LR+	1.61*			4.18		
	LR-	0.82*			0.35		
Sae-Jung et al ⁴	LR+		1.63	1.64		1.77	1.95
	LR-		0.69	0		0.33	0

* CIs in calculable due to absence of raw data.

** No raw data available to calculate CIs figures taken from Solomon et al 2001⁵.

Tabel 6. Perbandingan Nilai *Positive and Negative Likelihood Ratio*

Selain perhitungan-perhitungan di atas, dapat juga menggunakan nilai *predictive value*. Nilai ini terdiri dari:

1. *Positive predictive value*: proporsi dalam satuan persen (%) dengan hasil positif pada studi tes yang mempunyai problem seperti yang ditentukan dengan *reference standard*.
2. *Negative predictive value*: proporsi dalam satuan persen (%) dengan hasil negatif pada studi tes yang mempunyai problem seperti yang ditentukan dengan *reference standard*.

Tabel 7 berikut ini adalah rumus cara perhitungan nilai *predictive value*:

TABLE 5.4 Values for Positive Predictive Value (PPV) and Negative Predictive Value (NPV)

	MRI POSITIVE	MRI NEGATIVE	
Lachman test Positive	A True positives n = 39	B False positives n = 10	PPV A/A + B 39/39 + 10 = 0.80
Lachman test Negative	C False negatives n = 8	D True negatives n = 3	NPV D/C + D 3/8 + 3

Tabel 7. Perhitungan *Positive Predictive Value* dan *Negative Predictive Value*

Selanjutnya adalah metode lain terkait penentuan nilai diagnostik yaitu dengan melakukan perbandingan akurasi diagnostik antara 2 atau lebih tes diagnostik. Metode ini disebut sebagai *diagnostic odd ratio* (DOR). Perhitungan metode ini adalah sebagai berikut:

$$DOR = (True\ Positive/False\ Negative)/(False\ Positive/True\ Negative)$$

Berikut ini adalah salah satu evidence terkait studi tes diagnostik menggunakan DOR pada tabel 8 berikut ini:

Table 2 Diagnostic measures of single PETS ranked by DOR

PETS category	Single PETS	No. of studies	Pooled ^a results (95% CI)		DOR	Accuracy ^b	Accuracy rank	Likelihood ratios ^c positive/negative
			Sensitivity	Specificity				
SLAP	Compression-rotation	2	0.43 (0.31, 0.56)	0.89 (0.67, 0.97)	6.36 (1.41, 28.59)	0.68 (0.59, 0.77)	1	3.91/0.64
	Yergason	3	0.20 (0.13, 0.30)	0.92 (0.81, 0.97)	2.91 (0.94, 9.08)	0.59 (0.54, 0.65)	3	2.50/0.87
	Anterior apprehension	2	0.74 (0.61, 0.84)	0.45 (0.35, 0.55)	2.29 (1.12, 4.69)	0.51 (0.51, 0.66)	4	1.35/0.58
	Crank	2	0.46 (0.33, 0.60)	0.72 (0.54, 0.85)	2.18 (0.82, 5.78)	0.60 (0.49, 0.71)	2	1.64/0.75
	Speed	3	0.20 (0.11, 0.32)	0.88 (0.73, 0.95)	1.73 (0.53, 5.65)	0.57 (0.49, 0.64)	5	1.67/0.91
	Relocation	2	0.61 (0.48, 0.72)	0.47 (0.37, 0.57)	1.36 (0.69, 2.66)	0.53 (0.45, 0.61)	6	1.15/0.83
	O'Brien	3	0.66 (0.55, 0.75)	0.36 (0.21, 0.55)	1.10 (0.46, 2.60)	0.50 (0.39, 0.60)	9	1.03/0.94
	Bicipital groove tenderness	2	0.26 (0.17, 0.37)	0.74 (0.63, 0.82)	0.98 (0.47, 2.05)	0.52 (0.45, 0.59)	7	1.00/1.00
	Kibler/anterior slide	2	0.10 (0.04, 0.23)	0.85 (0.73, 0.93)	0.61 (0.17, 2.23)	0.51 (0.44, 0.57)	8	0.67/1.06
SIS	Hawkins-Kennedy	2	0.58 (0.50, 0.66)	0.67 (0.47, 0.83)	2.86 (1.14, 7.17)	0.60 (0.53, 0.68)	1	1.76/0.63
	Neer	2	0.59 (0.52, 0.67)	0.60 (0.40, 0.77)	2.17 (0.91, 5.19)	0.59 (0.52, 0.67)	2	1.48/0.68
RCT ^d	Any full thickness RCT	2	0.74 (0.39, 0.92)	0.77 (0.69, 0.83)	9.24 (1.99, 42.84)	0.76 (0.63, 0.88)	NA	3.22/0.63
	Supraspinatus full thickness only	2	0.60 (0.46, 0.72)	0.70 (0.61, 0.78)	3.50 (1.74, 7.02)	0.66 (0.58, 0.73)	NA	2.00/0.57
	Any RCT	3	0.60 (0.52, 0.68)	0.63 (0.55, 0.71)	2.63 (1.62, 4.27)	0.62 (0.56, 0.68)	NA	1.62/0.63

PETS physical examination tests of the shoulder, DOR diagnostic odds ratio, SLAP superior labrum anterior superior, SIS subacromial impingement, RCT rotator cuff tear

^aCalculated by a fixed effect model

^bBased on average prevalences: 0.456 for SLAP lesions tests, 0.767 for SIS tests and 0.422 for the RCT test

^cCalculated from pooled sensitivity and specificity

^dPooling was only possible for the Supraspinatus test (= Empty can test = Jobe's test); Weakness indicates positive test

Tabel 8. Pengukuran Studi Diagnostik

E. Pertanyaan

1. Jelaskan proses *appraisal* pada studi terkait diagnostik?
2. Jelaskan apa yang dimaksud *sensitivity* dan *specificity*?
3. Jelaskan kemungkinan-kemungkinan hasil dari perhitungan *sensitivity* dan *specificity*?
4. Jelaskan apa yang dimaksud *positive likelihood ratio* dan *negative likelihood ratio*?
5. Jelaskan apa yang dimaksud *positive predictive value* dan *negative predictive value*?

F. Daftar Pustaka

Rob Herbert, Gro Jamtvedt, Kare Birger Hagen, Judy Mead, Practical Evidence-Based Physiotherapy 2nd Ed, (Elsevier Churchill Livingstone, London, 2011), pp. 32-35

Linda Fethers, Julie Tilson, Evidence Based Physical Therapy, F.A Davis Company, Philadelphia, 2012), pp 13-27