



Modul 8

FEB 326-Evidence-Based Practice Fisioterapi

Materi 8

Critical Appraisal Terkait Kualitas Studi Intervensi

Disusun Oleh

Wahyuddin

UNIVERSITAS ESA UNGGUL

2020

A. Pendahuluan

Aplikasi *evidence-based practice* fisioterapi dimulai dengan pertanyaan-pertanyaan klinis yang lama proses selanjutnya pertanyaan-pertanyaan tersebut harus dapat terjawab.

Pertanyaan terkait efek intervensi dibagi dalam beberapa komponen yang sering disingkat PICO yaitu:

- Patient or problem.
- Intervention or management strategy
- Comparative intervention
- Outcome.

Beberapa penulis lain menambahkan komponen time (T).

B. Kompetensi Dasar

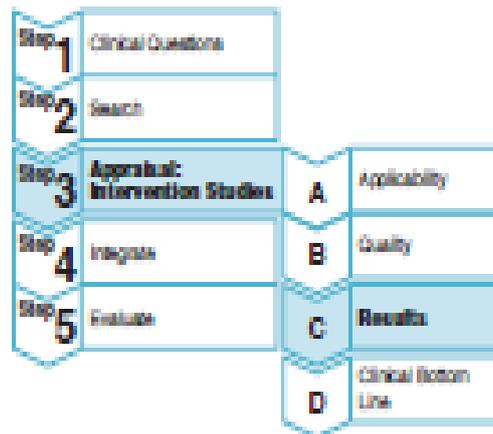
Kompetensi dasar yang harus dimiliki oleh mahasiswa dan mahasiswi pada materi ini adalah pemahaman tentang pencarian referensi terkait efek suatu intervensi pada kondisi yang relevan.

C. Kemampuan Akhir Yang Diharapkan

Setelah mempelajari materi ini, diharapkan semua mahasiswa dan mahasiswi mampu menganalisis referensi terkait studi intervensi dan memahami pentingnya hal tersebut dalam aspek praktik klinis

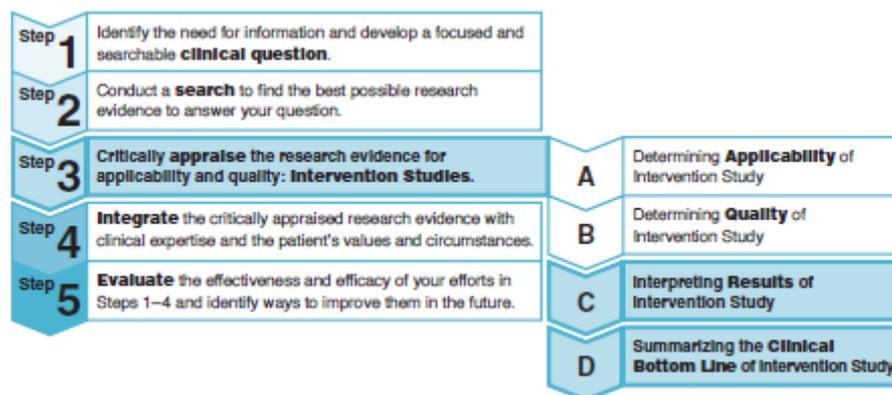
D. Uraian Materi

Membaca dan menafsirkan hasil penelitian merupakan suatu aktifitas menantang. Jika kita mengembangkan pendekatan sistematis untuk membaca melalui hasil studi, akan menjadi lebih mudah dan lebih cepat untuk menginterpretasikan temuan-temuan studi tersebut. Materi ini akan menjelaskan proses untuk menilai hasil dan memandu kita untuk melakukan penilaian, termasuk menginterpretasikan hasil interpretasi pengujian statistik. Secara umum dapat terlihat pada gambar 1 berikut ini:



Gambar 2: Lokasi dan Appraisal Hasil Studi

Secara spesifik dapat dilihat pada gambar 2 berikut ini:



Gambar 2: Appraisal Hasil Studi dan Clinical Bottom Line

Informasi dari tabel akan memberikan banyak hal pada pembaca. Hasil penelitian intervensi diringkas dalam tabel. Salah satu kebiasaan yang berguna untuk mengembangkan adalah untuk membaca tabel yang mencerminkan informasi yang terkandung di dalamnya sebelum mencoba untuk mengerti teks di bagian hasil.

Judul tabel menunjukkan apa yang akan dimasukkan dalam bagian utama dari tabel. Informasi pada tabel dapat menggambarkan karakteristik demografis dan klinis dari sampel studi untuk setiap group. Data dirangkum menggunakan statistik deskriptif. Statistik deskriptif akan memberi kesan keseluruhan nilai khas untuk kelompok serta variabilitas dalam dan diantara group. Nilai angka dalam tabel dilaporkan sebagai rerata dan standar deviasi

untuk karakteristik masing-masing dan setiap kelompok. Rincian satuan unit untuk karakteristik masing-masing ditambahkan di bawah tabel sebagai catatan kaki.

Contoh tersebut dapat diperlihatkan pada tabel 1 berikut ini:

Variable	Control Group (n = 60)	Training Groups	
		Endurance (n = 59)	Strength (n = 60)
Demographics			
Age, y	46 (5)	46 (6)	45 (6)
Height, cm	164 (5)	165 (6)	165 (5)
Weight, kg	69 (12)	68 (10)	67 (11)
Body mass index†	26 (4)	25 (3)	25 (3)
Clinical			
Duration of neck pain, y	8 (5)	9 (6)	8 (6)
Short depression inventory score‡	6 (3)	6 (4)	5 (3)
Grip strength, right hand, N§	293 (54)	299 (50)	299 (54)
Grip strength, left hand, N§	266 (46)	270 (53)	286 (52)
Maximum oxygen uptaken, mL/kg per min	31 (5)	32 (4)	33 (5)
Smoking, No. (%)	10 (17)	12 (20)	10 (17)

Note 4.2
Units of measure and comments

Note 4.1
Means and standard deviations for age are listed by group.

Hypothetical Ages:
Control – 52 (6)
Endurance – 46 (6)
Strength – 42 (3)

Note 4.3
Grip strength has a wide variability among groups

Abbreviation: N, Newton, which is a measure of force.
*Data are presented as mean (SD) unless otherwise indicated.
†Body mass index is calculated as weight in kilograms divided by the square of height in meters.
‡Mood is assessed on a theoretical range of 1 to 21, with a lower score indicating a better mood.
§Grip strength was measured using a hand-held Jamar grip-strength device while the participant was in a seated position and the elbow supported in a right angle.

Tabel 1: Contoh Tabel Data Demografi dan Klinis (Ylinen J, Takala EP, Nykanen M, et al. 2003)

Interpretasi Hasil Studi Intervensi

Pertanyaan Kritis

Pada bagian penjelasan yang dibahas setelah tabel, termasuk didalamnya adalah pertanyaan-pertanyaan yang dipertimbangkan untuk menilai kualitas studi intervensi. Beberapa pertanyaan-pertanyaan tersebut adalah sebagai berikut:

Pertanyaan 1: Apakah terdapat similaritas antar grup yang dibandingkan pada baseline?

Karakteristik kelompok harus relatif sehomogen mungkin pada awal studi intervensi, yaitu sebelum pemberian intervensi. Pengacakan subyek harus memastikan bahwa kelompok-kelompok tersebut setidaknya serupa jika tidak persis sama. Tapi bagaimana seharusnya kemiripan tersebut? Bagaimana jika mereka "agak" berbeda? Seberapa berbedanya sampel yang masih dapat diterima sebelum studi dimulai?

Misalnya pada suatu studi, terdapat perbedaan usia rata-rata 10 tahun antara satu kelompok dengan kelompok lain. Apakah ini usia yang "cukup" sama, atau apakah perbedaan usia berkontribusi terhadap respon intervensi yang pada akhirnya akan berdampak pada hasil? Hal ini dapat mengarah kepada untuk sumber-sumber penyebab terjadinya bias atau kadang disebut sebagai kesalahan yang dapat mempengaruhi interpretasi terhadap hasil dari faktor-faktor tersebut.

Perbedaan kelompok pada awal studi yang berpotensi sebagai sumber kesalahan yang dapat berkontribusi terhadap hasil penelitian dan dengan demikian mengurangi kepastian bahwa kesimpulan hasil sebagai dampak dari intervensi. Kelompok mungkin memiliki nilai-nilai yang berbeda pada karakteristik pra-intervensi karena:

- kelompok benar-benar berbeda dan pengacakan tidak dapat dilakukan secara tepat,
- telah dilakukan pengukuran sebelumnya oleh orang dan instrumen yang berbeda,
- instrumen atau tes yang digunakan menjadi sumber terjadinya kesalahan dalam pengukuran.

Misalnya pada suatu studi terkait kekuatan menggenggam dengan menggunakan alat ukur dynamometer. Data yang didapatkan menunjukkan pada kelompok intervensi adalah nilai rata-rata 20 newton lebih besar dibandingkan dengan kelompok kontrol. Apakah ini menunjukkan perbedaan "nyata" atau hasil bias karena penggunaan dynamometer atau orang-orang yang mengukur? Pertama, periksa kesalahan yang mungkin terjadi. Jika kita melakukan pengukuran kekuatan dua kali pada seorang pasien, hasilnya akan memiliki kemungkinan beberapa variabilitas pengukuran. Variabilitas ini ini dapat disebut kesalahan karena tidak persis sama secara tepat.

Menerima bahwa pengukuran-pengukuran tidak persis sama, tapi seberapa banyak variabilitas (kesalahan) yang dapat diterima? Jika harapan kita bahwa intervensi akan meningkatkan kekuatan, tetapi pengulangan pengukuran yang tidak sama, maka bagaimana kita dapat mengetahui apakah pengukuran kesalahan (variabilitas) dan apa efek dari treatment yang diberikan? Jika kita mengharapkan manfaat dari intervensi untuk menjadi kecil, maka kesalahan dapat mempengaruhi hasil pengukuran. Reliabilitas pengukuran adalah

kemampuan seseorang dan instrumen untuk menghasilkan nilai-nilai yang konsisten dari waktu ke waktu. Tes dan perangkat yang digunakan oleh fisioterapis harus merupakan rangkaian pengukuran reliabel untuk mereproduksi pengukuran.

Reliabilitas Pengukur dan Partisipan

Pertanyaan 2: Apakah outcome pengukuran reliabel dan valid?

Agar data menghasilkan data berkualitas tinggi hasil dari sebuah penelitian, data harus valid dan reliabel. Reliabilitas data direproduksi oleh satu atau lebih orang (fisioterapi terapis) dan stabil dalam satu periode waktu. Intra-rater reliability adalah pengulangan pengukuran oleh fisioterapis yang sama pada pasien sama di dua atau lebih titik waktu.

Misalnya, jika kita melakukan pengukuran gerakan fleksi lutut dengan rentang gerak 120 derajat, kita harus dapat memperoleh skor yang sama pada pengukuran kedua dan ketiga. Inter-rater reliability adalah pengulangan antara dua atau lebih fisioterapis yang melakukan pengukuran pada pasien yang sama. Kedua fisioterapis harus mendapatkan skor 120 derajat dari contoh di atas. Fisioterapis yang mengukur partisipan studi, alat-alat dan instrument yang mereka gunakan harus reliabel.

Faktor-faktor yang terdapat pada pasien juga dapat menyebabkan variabilitas pengukuran. Faktor-faktor tersebut dapat berupa fluktuasi alami penyakit atau proses-proses fisiologik yang normal, perubahan pada siang hari/malam, keadaan emosi, kewaspadaan dan berbagai faktor lain yang mempengaruhi kinerja manusia. Pengulangan variabel pengukuran, disebut intra subject variability (misalnya, sebelum intervensi dimulai) memberikan metode kuantitatif untuk menentukan variabel yang tepat.

Satu pengukuran pre-intervensi, serupa dengan satu nilai rata-rata, memiliki keterbatasan nilai tanpa pemahaman tentang variabilitas normal dari apa yang diukur. Jika intra-subjek variability tinggi, mungkin berarti bahwa apa yang kita ukur memiliki variabilitas tinggi atau teknik pengukuran perlu ditingkatkan sehingga menjadi lebih reliabel.

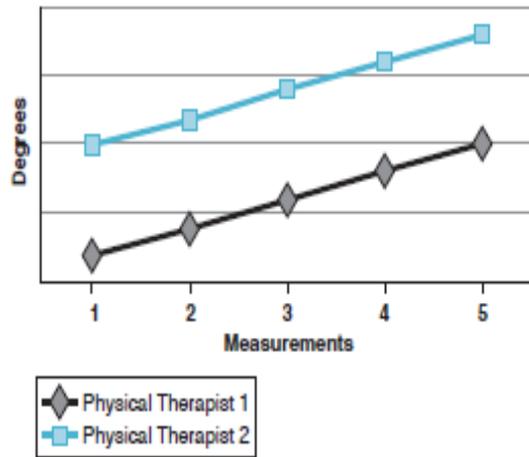
Reliabilitas Pengukuran Ditentukan oleh Instrumen

Dynamometer adalah salah contoh instrumen akan perlu diuji untuk mendapatkan tingkat akurasi pengukuran dan reproduksibilitas dari waktu ke waktu. Suatu instrumen yang valid akan mengukur apa yang seharusnya diukur. Instrumen yang reliabel akan menghasilkan hal yang sama pada pengukuran berulang apabila dikelola dengan benar oleh penguji yang reliabel. Reliabilitas dalam dan di antara orang-orang biasanya dinyatakan dengan asosiasi (korelasi) dari nilai-nilai yang diperoleh dalam pengukuran berulang. Baik intra-rater and inter-rater reliability harus memiliki nilai asosiasi yang tinggi, yaitu nilai-nilai tersebut relatif sama.

Jenis korelasi yang dilakukan tergantung pada skala pengukuran yang digunakan, karena skala akan menentukan bentuk data. Reliabilitas harus ditetapkan dalam penelitian, dan seharusnya dengan nilai yang tinggi. Studi memiliki validitas dan reliabilitas tinggi ditentukan antara penilai sebelum melakukan studi. Penulis mungkin dapat mengutip penelitian lainnya dimana telah terdapat nilai reliabilitas, tetapi hal ini kurang valid saat benar-benar melakukan pengukuran pada kondisi riil. Fakta bahwa dua atau lebih penilai dapat mencapai reliabilitas yang tinggi tidak mendukung fakta bahwa penguji dalam suatu studi memiliki reliabilitas yang sama.

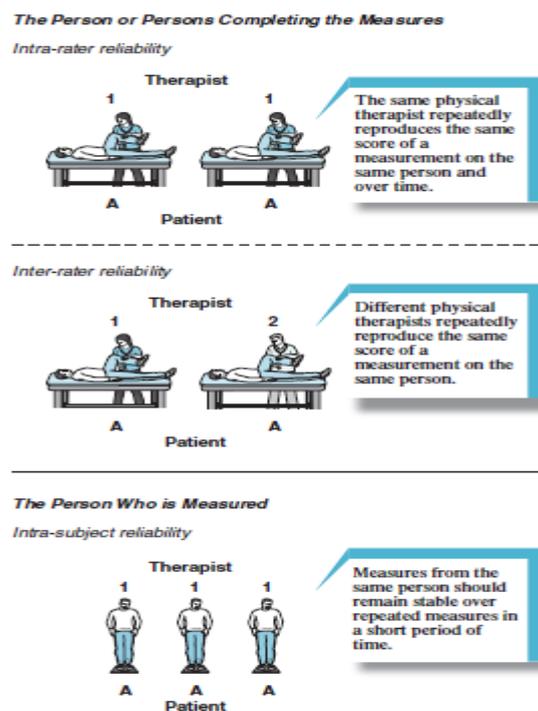
Metode Analisis Reliabilitas

Pearson product-moment coefficient of correlation kadang-kadang digunakan untuk melaporkan nilai reliabilitas. Nilai korelasi Pearson antara dua penguji dikatakan sempurna jika bernilai 1,0. Namun, secara aktual mungkin akan berbeda. Nilai rentang gerak yang diukur oleh seorang fisioterapis mungkin akan berbeda dengan pengukuran yang dilakukan oleh fisioterapis lain, tetapi nilai-nilai mutlak dalam dua seri tidak sama. Data tersebut tergambar pada gambar 3 berikut ini:



Gambar 3: Data Hasil Pengukuran Rentang Getak Sendi

Ukuran seperti intraclass correlation coefficient (ICC) memperhitungkan kedua sifat perubahan dan nilai-nilai mutlak dan dengan demikian secara statistik lebih dianjurkan untuk mengevaluasi keandalan data yang kontinu (rasio atau interval). Ada berbagai metode statistik untuk menghitung keandalan untuk instrumen dan pengukur). Pilihan jenis statistik yang digunakan mencerminkan jenis data yang dianalisis. Berbagai jenis reliabilitas digambarkan pada gambar 4 di bawah ini:



Gambar 4. Jenis-Jenis Reliabilitas

Penilaian Hasil Statistik

Tabel juga digunakan untuk menampilkan hasil dari studi intervensi. Membaca melalui tabel sebelum membaca bagian hasil dapat memberikan kesan keseluruhan hasil studi. Tabel 2 berikut ini adalah suatu studi oleh Santamato et al yang membandingkan dua intervensi untuk kasus subacromial impingement syndrome. Tiga kolom pertama pada tabel daftar tes, nilai baseline (sebelum intervensi) dan setelah intervensi yang digunakan dalam penelitian untuk masing-masing kelompok. Hasil dibuat dalam bentuk rerata dan standar deviasi pada masing-masing kelompok (High Intensity Laser Therapy dan Ultrasound).

Variable	Both Groups (n=70)	HILT Group (n=35)	US Therapy Group (n=35)	Difference in Means (95% CI)	Actual P Value	Bonferroni-Corrected P Value ^b
VAS score						
\bar{X} (SD)	-3.01 (1.46)	-3.86 (1.53)	-2.17 (0.75)	-1.69 (-2.27 to -1.12)	<.001	<.01
%	-48.89	-61.36	-33.04			
CMS score						
\bar{X} (SD)	10.86 (3.68)	12.69 (3.64)	9.03 (2.70)	3.66 (2.13 to 5.19)	<.001	<.01
%	17.19	20.06	14.31			
SST score						
\bar{X} (SD)	2.14 (0.98)	2.46 (1.17)	1.83 (0.61)	0.63 (0.18 to 1.08)	.006	<.05
%	30.30	33.99	26.45			

^a CI-confidence interval, VAS-visual analog scale, CMS-Constant-Murley Scale, SST-Simple Shoulder Test.

^b The statistical inferences were adjusted according to Bonferroni inequality (0.05/6-0.008 and 0.01/6-0.002).

Note 4.5

Is this a difference that we should consider in our clinical decisions?

Note 4.6

The P value is .006 ($p < .05$) for the difference between the HILT and US Therapy groups.

Note 4.4 Confidence Interval

We can be 95% confident that the true mean lies between -2.27 to -1.12

Tabel 2. Nilai Pengukuran dan Hasil Analisis Pada Kelompok High Intensity Laser Therapy dan Ultrasound

Pertanyaan 3: Apakah Studi Melaporkan Nilai Confidence Intervals?

Kolom kelima pada tabel 2 di atas menunjukkan perbedaan nilai antara kedua kelompok intervensi. Nilai yang tercantum dalam kurung adalah confidence intervals dari nilai rata-rata. Confidence intervals adalah serangkaian nilai rata-rata. Sementara nilai rata-rata adalah estimasi nilai riil yang didapatkan dari populasi orang yang diberikan intervensi dan kemudian dilakukan pengukuran. Dalam semua penelitian studi, hanya sampel dari jumlah total populasi yang terlibat yang diukur, sampel ini memberikan perkiraan apa yang dapat diharapkan dari kelompok populasi yang lebih besar. Jika kita ingin suatu gagasan tentang bagaimana akurat nilai rata-rata mewakili populasi karena nilai rata-rata hanya bersifat perkiraan. Confidence intervals dapat memberikan

informasi ini, dan biasanya penyampaian pada 90 atau 95 percentile. Guyatt et al menawarkan contoh lemparan koin untuk menggambarkan konsep confidence intervals. Semakin besar jumlah sampel (lemparan koin), semakin baik estimasi terkait nilai riil. Ini berarti akan mengarahkan kepada nilai confidence intervals yang semakin kecil.

Interaksi Statistik Deskriptif dan Inferensial

Pertanyaan 4: Apakah analisis statistik deskriptif dan inferensial diaplikasikan terhadap hasil?

Statistik deskriptif membantu merangkum informasi tentang kelompok sebelum dan sesudah pemberian intervensi. Namun demikian, ringkasan angka harus diinterpretasikan agar lebih mudah dipahami terkait implikasi pada pengambilan keputusan klinis. Misalnya apakah kita fokus pada perbedaan nilai rata-rata terkait karakteristik demografi dan klinis?

Statistik deskriptif, pada contoh nilai rata-rata mengindikasikan bahwa terdapat perbedaan pada sampel. Namun demikian, jenis statistik ini tidak membantu memberikan informasi apakah perbedaan tersebut berdampak terhadap validitas studi yang mempunyai konsekuensi terhadap pengambilan keputusan klinis. Apakah perbedaan karakteristik subjek mempengaruhi outcome studi, ataukah apakah outcome tersebut disebabkan oleh intervensi yang dilakukan?

Sebagai tambahan kita tidak dapat menginterpretasikan perubahan rata-rata hanya melihat dari skor sebelum dan sesudah intervensi. Seperti pada data di tabel 2, pada kolom 3 dan 4 kedua kelompok intervensi terdapat perubahan positif pada tiga komponen pengukuran (VAS, CMS, dan SST). Tetapi kita juga melihat bahwa terdapat perubahan positif yang lebih besar pada kelompok High Intensity Laser Therapy. Dari data tersebut, apakah kita dapat menyimpulkan bahwa kelompok High Intensity Laser Therapy lebih efektif dibanding kelompok Ultrasound?

Statistik deskriptif bermanfaat, tetapi tidak cukup untuk membuat konklusi tentang perbedaan antar kelompok. Dalam hal ini maka digunakan statistik inferensial. Statistik inferensial adalah tool yang menggunakan matematika probabilitas untuk membantu membuat konklusi untuk menginterpretasikan perbedaan yang diobservasi dalam riset studi. Pendekatan

statistik ini fokus pada pertanyaan apakah outcome yang dihasilkan merupakan dampak dari intervensi atau hanya terjadi secara kebetulan?

Peran “Kejadian” Terkait Hasil

Statistik inferensial didasarkan pada matematika probabilitas yang digambarkan dengan probabilitas atau p values. Kolom keenam pada tabel 2 meliputi p values. P value menggambarkan probabilitas terkait perbedaan yang teridentifikasi karena suatu kejadian. Jika terdapat probabilitas yang tinggi (p value besar), kita tidak dapat membuat konklusi bahwa intervensi sebagai alasan terdapat perbedaan antara 2 kelompok. Apa nilai yang merepresentasikan bahwa terdapat nilai probabilitas yang tinggi atau rendah?

Konsensus antara peneliti pada umumnya menginginkan bahwa nilai probabilitas serendah mungkin. Kesepakatan ini yang dikenal sebagai level alpha (α) yang umumnya pada kisaran 5%. Hasil analisis menggunakan statistik inferensial yang digambarkan sebagai p value, dimana nilai ini akan dievaluasi apakah level alpha $<p=0.05$ atau $>p = 0.05$.

Jika kita membandingkan antar grup, kita dapat melihat bahwa terdapat perbedaan yang signifikan secara statistik dimana grup High Intensity Laser Therapy lebih besar dibanding kelompok dengan Ultrasound, walaupun kedua kelompok mengalami peningkatan setelah mendapatkan intervensi. Pertanyaan yang muncul adalah apakah kita konsen terkait perbedaan distribusi demografi dan karakteristik klinis pada awal sebelum pengaplikasian intervensi atau yang juga dikenal sebagai data baseline?

Tabel 1 memberikan informasi data baseline terkait karakteristik partisipan sebelum pemberian intervensi. Data tersebut berupa nilai rata-rata, standar deviasi, rentang nilai usia, timbulnya nyeri dan tahapan proses diagnostik. Bagaimana kita dapat menginterpretasikan nilai-nilai tersebut. Semua nilai-nilai di atas (p-value) berada pada nilai $> \alpha (0,05)$. Ini menunjang pernyataan bahwa tidak ada perbedaan signifikan antar grup. Artinya bahwa interpretasi nilai rata-rata pada kedua kelompok relatif sama pada baseline dan karakteristik subjek pada pengukuran baseline tidak menyebabkan bias pada hasil studi. Hal ini akan meningkatkan keyakinan bahwa treatment yang diberikan, bukan karena ketidakhomogenan antar grup pada baseline merupakan faktor yang mempengaruhi hasil.

Apa arti bahwa nilai rata-rata terhadap signifikansi hasil? Pertimbangkan bahwa jika alpha ditentukan pada nilai 0,05 ini berarti tingkat signifikansi didasarkan pada nilai tersebut. Bagaimana jika rentang p value berada di antara 0,05 ke 0,1 atau bahkan lebih tinggi? Apakah secara otomatis kita asumsikan bahwa perbedaan antara grup terjadi karena adanya peluang? Apakah nilai 0,06 mengindikasikan bahwa terdapat peluang 6% terjadinya perbedaan diakibatkan oleh peluang. Apakah justifikasi harus diberikan terhadap temuan-temuan seperti itu?

Kita tidak dapat menolak temuan penting hanya karena p value lebih besar dari nilai alpha. Pada kejadian seperti kita harus mereview semua p value dalam studi dan pertimbangkan relevansi nilai antara 0,15 dan 0,05. Mungkin ada suatu pola terkait hasil yang dapat dipertimbangkan jika p value lebih besar daripada nilai alpha. Ada beberapa perangkat statistik inferens atau hal lain yang dapat digunakan. Kemampuan pemahaman terhadap aspek statistik akan sangat membantu menginterpretasikan hasil suatu studi. Secara spesifik khususnya pada studi yang menganalisis dampak dari suatu intervensi.

Statistik untuk Evaluasi Relevansi Klinis

Pertanyaan 5: Apakah ada efek treatment?

Jika ada apa relevansi secara klinis? Statistik deskriptif dan inferensial adalah dua kategori perangkat statistik yang digunakan untuk menganalisis data dari suatu penelitian. Pembaca sering menyimpulkan (dan penulis membuat implikasi) bahwa temuan signifikansi secara statistik berasal dari suatu studi yang valid (misalnya hasil p value $<0,05$ pada suatu studi terkait intervensi yang didesain secara baik) berarti bahwa hasil-hasil studi harus penting. Namun demikian, dalam beberapa hal hasil signifikansi secara statistik mungkin tidak penting secara klinis. Misalnya, jika kita mencoba suatu intervensi baru pada pasien dengan hambatan gerak yang besar pada beberapa jari di tangan dominan. Setelah diberikan penanganan dalam beberapa lama, pasien mengalami peningkatan kemampuan untuk menggerakkan jari-jari sekitar seperempat inci. Jika kita mereplikasi intervensi tersebut pada jumlah sampel yang besar dengan capaian outcome yang sama, analisis statistik mungkin menunjukkan signifikansi perubahan tersebut pada p value $<0,05$. Namun

demikian, perubahan tersebut tidak berdampak besar terhadap peningkatan penggunaan tangan secara fungsional.

Agar dapat menunjukkan nilai klinis yang penting, hasil penelitian harus mampu:

- menunjukkan perubahan dalam suatu pengukuran menunjukkan nilai kepada pasien terkait kehidupan sehari-hari. Dalam contoh di atas, pengukuran penggunaan tangan secara fungsional dan juga rentang gerak sendi mungkin membantu mengevaluasi nilai perubahan pada pasien.
- menunjukkan perubahan dalam konteks magnitude yang akan membuat mereka berbeda dalam kehidupan sehari-hari (misalnya fungsi, kepuasan, kenyamanan dan lain sebagainya)

Karena itu, kita membutuhkan beberapa pengukuran selain aspek p value untuk menjawab pertanyaan apakah hasil dari suatu studi yang dilakukan secara valid akan bermanfaat?

Effect Size

Dari penjelasan di atas, p value membantu kita untuk menentukan peran kejadian terkait hasil dari suatu studi. Namun demikian, kita juga perlu mengetahui seberapa besar perbedaan yang terjadi antar treatment. Satu pendekatan untuk mengevaluasi magnitude dari perbedaan tersebut adalah mengkalkulasi effect size. Hal paling sering didasarkan pada pedoman oleh Cohen. Effect size yang diberi tanda d memberikan suatu pengukuran seberapa besar perbedaan antara sampel, seberapa jauh nilai rata-rata dapat diterima. Hal ini menyangkut variabilitas yang terlihat pada rumus pada gambar 5 berikut ini:

$$d = \frac{\text{Mean}_{\text{group 1}} - \text{Mean}_{\text{control group}}}{\text{SD}_{\text{control group}}}$$

Gambar 5. Rumus Effect Size

Nilai d yang besar mengindikasikan besar perbedaan antara dua kelompok. Artikel tidak selalu mencantumkan nilai effect size, tetapi statistik deskriptif berupa nilai rata-rata dan standar deviasi selalu dicantumkan pada

tabel. Kita dapat menghitung nilai d untuk mengevaluasi magnitudo perbedaan yang dilaporkan. Interpretasi nilai d secara statistik sesuai pedoman yang diberikan oleh Cohen ditampilkan pada tabel 3 berikut ini:

TABLE 4.2 Interpretation of Effect Size Values

Large	>0.8
Medium	0.5-0.8
Small	0.2-0.5

From: Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum; 1988.

Tabel 3. Interpretasi Nilai Effect Size (Cohen)

Number Needed to Treat

Banyak diskusi-diskusi dalam literature evidence-based practice yang mengeksaminasi pentingnya hasil yang terkait dengan dichotomous outcomes. Dichotomous outcome berupa dua kategori, misalnya sehat atau sakit. Effect size dapat dihitung berdasarkan hasil yang bersifat kontinyu. Misalnya pengukuran rentang gerak sendi, skor pengukuran fungsional atau pengukuran kekuatan dengan dynamometer. Ada beberapa contoh penting dalam pengukuran fisioterapi dimana outcome bersifat dichotomous. Misalnya kesuksesan intervensi yang mungkin diukur dengan membandingkan jumlah pasien yang kembali atau tidak ke aktifitas kerja atau olahraga, berjalan secara mandiri atau tidak, atau yang mungkin diklasifikasikan dalam beberapa kriteria terdapat peningkatan outcome atau tidak.

Kita tidak dapat mengatakan bahwa suatu intervensi baru akan membantu setiap orang, tetapi hasil statistik yang signifikan mengindikasikan bahwa terdapat proporsi yang lebih besar pada pasien yang menerima intervensi baru mendapatkan outcome yang positif dibandingkan yang tidak. Apa jenis statistik yang diperlukan yang dapat menjawab pertanyaan berikut: jika saya menggunakan suatu intervensi baru pada semua jenis pasien baru, berapa banyak treatment yang akan dilakukan untuk melihat manfaat dibandingkan dengan treatment lama? Pertanyaan ini akan dijawab dengan statistik

menggunakan number needed to treat. Number needed to treat adalah rasio atau perbandingan outcome yang diharapkan pada kelompok intervensi dan kelompok kontrol. Rasio tersebut secara statistik dapat dipalिकासikan pada praktik klinis. Secara formula dapat dilihat pada gambar 6 berikut ini:

$$\frac{\text{Desired outcome in the experimental group}}{\text{Desired outcome in the control group}} = \text{NNT}$$

Gambar 6: Formula Number Needed to Treat

Diharapkan bahwa semua penulis artikel-artikel pada suatu jurnal memberikan informasi yang bersifat detail pada laporan-laporan mereka seperti persentase peningkatan pada setiap kelompok untuk menghitung number needed to treat sesederhana mungkin. Tabel 4 berikut ini memberikan informasi perhitungan number needed to treat:

TABLE 4.4 Computing Number Needed to Treat

	CER	EER	ARR	RRR	NNT
Definition of term	Control event rate	Experimental event rate	Absolute risk reduction (risk difference)	Relative risk reduction	Number needed to treat
Calculation of statistic (letters refer to Table 4.2)	c/c + d	a/a + b	CER - EER	$\frac{\text{CER} - \text{EER}}{\text{CER}}$	1/ARR
Answers	85/100 = 0.85	95/100 = 0.95	-0.10, or the intervention reduced the rate of hospitalizations by 10%	0.85 - 0.95/0.85 = -0.12 or the new intervention caused a 12% relative reduction in hospitalizations compared to the control	10

Tabel 4: Perhitungan number needed to treat

Pada contoh di atas, outcome yang diharapkan adalah tanpa hospitalisasi, sehingga kita berharap bahwa event rate pada kelompok intervensi lebih tinggi dibandingkan kelompok control. Karena itu, negative relative risk reduction dan absolute risk reduction mengindikasikan secara langsung pada

hasil. Pada kasus tersebut, tanda negatif tidak relevan ketika menginterpretasikan number needed to treat. Namun demikian, jika outcome yang diharapkan adalah penurunan rate, negative relative risk ratio dan absolute risk reduction akan mengindikasikan bahwa intervensi yang diberikan memberikan hasil yang berlawanan. Karena itu kondisi pada kelompok control lebih berhasil dibandingkan kelompok intervensi. Karena itu perlu dipikirkan melalui outcome yang diharapkan pada saat menghitung number needed to treat dan menginterpretasikan hasil secara simultan. Meskipun suatu intervensi yang lebih baik tidak akan menghasilkan outcome yang lebih baik 100 persen pada pasien (misalnya nilai number needed to treat lebih besar dari 1).

Intervensi yang lebih baik akan menghasilkan tingkat kesuksesan yang lebih baik pula. Karena itu tidak ada aturan yang mendefinisikan number needed to treat yang baik, tergantung pada konteks. Kita harus menggunakan alasan klinis dan pengalaman untuk menentukan jawaban. Umumnya number needed to treat digunakan untuk mengevaluasi suatu program intervensi dibanding membuat suatu keputusan intervensi pada seorang pasien. Hal itu mungkin digunakan untuk menentukan jika suatu perubahan dalam treatment harus terjadi pada masalah yang bersifat khusus.

Sebagai contoh, jika suatu program intervensi untuk mencegah terjadinya jatuh pada orang tua mempunyai nilai number needed to treat sebanyak 5, maka untuk setiap 5 orang tua yang menerima intervensi baru, pencegahan jatuh mungkin dilakukan pada tambahan satu orang. Tabel 5 berikut ini dapat digunakan untuk menginterpretasikan hasil studi intervensi dan menentukan clinical bottom line.

TABLE 4.5 Key Questions to Interpret the Results and Determine the Clinical Bottom Line for an Intervention Research Study

QUESTIONS	YES/NO	WHERE TO FIND INFORMATION	COMMENTS AND WHAT TO LOOK FOR
1. Were the groups similar at baseline?	<input type="checkbox"/> Yes <input type="checkbox"/> No	Results (occasionally listed in Methods)	The study should provide information about the baseline characteristics of participants. Ideally, the randomization process should result in equivalent baseline characteristics between groups. The authors should present a statistical method to account for baseline differences.
2. Were outcome measures reliable and valid?	<input type="checkbox"/> Yes <input type="checkbox"/> No	Methods will describe the outcome measures used. Results may report evaluators' inter- and/or intra-rater reliability.	The study should describe how reliability and validity of outcome measures were established. The study may refer to previous studies. Ideally, the study will also provide an analysis of rater reliability for the study. Reliability is often reported with a kappa or intraclass correlation coefficient score—values >0.80 are generally considered to represent high reliability (Table 4.1).
3. Were CIs reported?	<input type="checkbox"/> Yes <input type="checkbox"/> No	Results	The study should report a 95% CI for the reported results. The size of the CI helps the clinician understand the precision of reported findings.
4. Were descriptive and inferential statistics applied to the results?	<input type="checkbox"/> Yes <input type="checkbox"/> No	Methods and Results	Descriptive statistics will be used to describe the study subjects and to report the results. Were the groups tested with inferential statistics at baseline? Were groups equivalent on important characteristics before treatment began?
5. Was there a treatment effect? If so, was it clinically relevant?	<input type="checkbox"/> Yes <input type="checkbox"/> No	Results	The study should report whether or not there was a statistically significant difference between groups and if so, the magnitude of the difference (effect sizes). Ideally, the study will report whether or not any differences are clinically meaningful.

Tabel 5. Pertanyaan-Pertanyaan Kunci Terkait Interpretasi Hasil dan Penentuan Clinical Bottom Line Pada Studi Efek Intervensi

Ringkasan

Menjadi suatu hal yang bermanfaat jika kita mengikuti tahapan-tahapan dalam menginterpretasikan studi terkait hasil atau dampak dari suatu intervensi pada suatu kondisi. Latihan-latihan pada setiap tahapan-tahapan tersebut akan menjamin keterampilan kritisi semakin meningkat dan memberikan implikasi pada kemaknaan secara klinis yang menjamin hasil studi tersebut menjadi bagian dari praktik klinis kita.

Tahap pertama adalah membaca tabel dan gambar pada bagian hasil studi tersebut sebelum membaca konklusi penulis atau metode statistik yang digunakan untuk melakukan analisis data. Ini akan memberikan kesantentang

bagian hasil dan bentuk hasil itu sendiri. Idealnya terdapat suatu tabel yang meliputi karakteristik subjek pada kelompok intervensi. Umumnya karakteristik tersebut direpresentasikan dengan penggunaan statistik deskriptif. Kelompok yang dibandingkan seharusnya mempunyai karakteristik yang sama yang mungkin mempengaruhi respon terhadap treatment.

Selain hal di atas, juga dapat terdapat penggambaran p value yang dapat dibandingkan dengan nilai alpha untuk menentukan tingkat signifikansi atau tidak berdasarkan analisis statistik. Kemudian yang berikut perhatikan nilai confidence intervals. Ini akan membantu menginterpretasikan p value, besar sampel dan apa konklusi hasil studi. Perhatikan statistik yang mengarah kepada kemaknaan klinis seperti effect size dan number needed to treat.

Setelah melalui pengamatan secara akurat pada tabel dan grafik, kita harus membaca keseluruhan teks pada bagian hasil. Penulis seharusnya memberikan penjelasan secara bertahap pada setiap tabel dan grafik terkait proses yang digunakan untuk analisis dan konklusi. Refleksikan apa yang kita percaya berdasarkan hasil studi dan baca keseluruhan konklusi penulis. Buat asesmen apakah kita setuju dengan konklusi tersebut dan bagaimana konklusi itu dapat diimplementasikan di praktik klinis kita.

E. Pertanyaan

1. Jelaskan mengapa karakteristik subjek pada kelompok-kelompok penelitian harus pada kondisi sehomogen mungkin!
2. Jelaskan mengapa pengukuran outcome pada studi terkait efek intervensi harus valid dan reliabel!
3. Jelaskan apa yang dimaksud dengan effect size!
4. Jelaskan apa yang dimaksud dengan number needed to treat!
5. Apa yang dimaksud dengan pernyataan bahwa hasil penelitian terkait efek intervensi tidak hanya menekankan pada aspek signifikansi secara statistik, tetapi juga relevan secara klinis!

F. Daftar Pustaka

Linda Fethers, Julie Tilson, Evidence Based Physical Therapy, F.A Davis Company, Philadelphia, 2012), pp 45-54.

Ylinen J, Takala EP, Nykanen M, et al. Active neck muscle training in the treatment of chronic neck pain in women: a randomized controlled trial. *JAMA*. 2003;289:2509–2516.

Guyatt G, Jaeschke R, Heddle N, et al. Basic statistics for clinicians. 2. Interpreting study results: confidence intervals. *CMAJ*. 1995;152:169–173.

Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum; 1988.

Straus SE, Richardson WS, Glasziou P, et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. 3rd ed. Edinburgh, Scotland: Elsevier; 2005.

Glasziou P, Del Mar C, Salisbury J. *Evidence-Based Practice Workbook*. 2nd ed. Malden, MA: Blackwell Publishing; 2007.