



Modul 14

SFS510-Methodologi Penelitian Fisioterapi II

Materi 14

Pengukuran Reliabilitas

Disusun Oleh

Wahyuddin

UNIVERSITAS ESA UNGGUL

2020

Pendahuluan

Reliabilitas merupakan hal mendasar atau fundamental untuk semua aspek pengukuran. Jika perilaku pasien dapat bersifat reliabel, kita dapat mengharapkan tanggapan yang konsisten dalam kondisi tertentu. Pemeriksa yang reliabel adalah orang yang memberikan nilai konsisten untuk perilaku yang tidak berubah dari pasien. Sebuah instrumen yang reliabel akan dilakukan dengan prediksi yang akurat pada kondisi yang stabil. Sifat realitas diperlukan mengingat sangat jarang hasil pengukuran yang sempurna. Kesalahan suatu instrumen akan berdampak terhadap tingkat inkonsistensi.

Konsep ini sangat penting dalam penafsiran koefisien reliabilitas dalam desain studi reliabilitas. Misalnya kita tertarik untuk mempelajari reliabilitas sebuah perangkat baru untuk mengukur jarak gerak ekstensi lumbal. Kita mengumpulkan sampel dalam jumlah besar yang dikategorikan sebagai individu normal. Hasil pengukuran berada pada rentang 20 sampai 25 derajat. Meskipun jika cukup konsisten terhadap beberapa percobaan secara berturut-turut, koefisien reliabilitas mungkin akan rendah karena varian secara total kecil.

Koefisien reliabilitas yang rendah dapat menyesatkan dalam beberapa kondisi. Solusi untuk masalah ini adalah menyertakan subjek yang memiliki rentang skor yang lebih luas. Kita harus mempelajari individu normal serta pasien dengan hipermobilitas dan hipomobilitas gerakan ekstensi lumbal. Peneliti harus selalu mempertimbangkan rentang skor yang digunakan untuk memperkirakan reliabilitas dalam menginterpretasikan koefisien reliabilitas.

Intraclass correlation coefficient (ICC)

Pendekatan historis untuk menguji reliabilitas melibatkan penggunaan koefisien korelasi. Korelasi juga terbatas sebagai koefisien reliabilitas karena bersifat bivariate yaitu hanya dua peringkat atau penilai dapat berkorelasi pada satu waktu. Hal ini tidak mungkin untuk menilai reliabilitas secara simultan lebih dari dua penilai atau hubungan antara aspek yang berbeda dari reliabilitas pemeriksa, bentuk tes, dan kesempatan pengujian. Karena elemen ini penting dalam pengujian reliabilitas, korelasi tidak memberikan mekanisme yang efisien untuk mengevaluasi cakupan reliabilitas secara

utuh. Keberatan lain terhadap penggunaan korelasi sebagai ukuran reliabilitas didasarkan pada definisi statistik reliabilitas yaitu korelasi tidak dapat memisahkan komponen varian karena kesalahan atau perbedaan dalam kumpulan data. Oleh karena itu, koefisien korelasi bukan koefisien reliabilitas yang benar. Hal ini sebenarnya lebih akurat untuk menggunakan kuadrat dari koefisien korelasi (koefisien penentuan/ r^2) untuk tujuan ini karena mencerminkan berapa banyak varian dalam satu pengukuran dihitung dengan varian dalam pengukuran kedua.

Koefisien relatif reliabilitas mencerminkan varian yang benar dalam satu set skor kontinyu sebagai proporsi dari total varian. Salah satu indeks reliabilitas relatif yang paling umum digunakan adalah intraclass correlation coefficient (ICC). ICC terutama digunakan untuk menilai test-retest atau rater reliability pada pengukuran kuantitatif (interval atau rasio). Nilai ICC berkisar dari 0,00 hingga 1,00, dengan nilai yang lebih tinggi menunjukkan reliabilitas yang lebih besar. Karena ICC adalah indeks unit, kita dapat membandingkan ICC sebagai metode pengujian alternatif untuk menentukan tes mana yang lebih reliabel. Hal ini dimungkinkan bahkan ketika menggunakan skor dengan unit pengukuran yang berbeda.

ICC memiliki keuntungan untuk dapat menilai reliabilitas di dua, tiga, atau lebih set skor dan memberikan penerapan yang luas untuk penilaian test-retest reliability pada pengujian berulang, interrater reliability pada beberapa penilai, atau intrarater reliability beberapa kali pengujian. Penerapan ini adalah salah satu dari beberapa keuntungan yang membedakan ICC dari koefisien korelasi lainnya, seperti Pearson r , yang hanya dapat dilakukan pada korelasi antara dua set skor, dan tidak menilai agreement.

Hal ini analog untuk menanyakan berapa banyak varian total dalam satu set data dibagi dengan dua pengukuran ("true" varians) dan berapa banyak tidak dibagi (varian kesalahan). Jika kita bisa mengkorelasikan nilai sebenarnya dengan nilai yang diamati dalam satu set data, kuadrat dari koefisien korelasi akan menjadi koefisien reliabilitas.

Strategi ini membahas interpretasi agreement, tetapi tidak berguna karena tidak memberikan indeks tunggal untuk menggambarkan reliabilitas. Skor mungkin berkorelasi tetapi secara signifikan berbeda, atau kurang berkorelasi tetapi tidak secara

signifikan berbeda. Bagaimana seharusnya hasil ini diinterpretasikan? Hal ini memerlukan penggunaan satu indeks yang dapat menjawab pertanyaan ini.

ICC adalah salah satu indeks tersebut. Seperti koefisien reliabilitas lainnya, ICC berkisar dari 0,00 sampai 1,00. Dihitung dengan menggunakan perkiraan varian yang diperoleh melalui analisis varian. Oleh karena itu, ini mencerminkan kedua tingkat korespondensi dan kesepakatan antara peringkat.

ICC dihitung dengan menggunakan perkiraan varian yang diperoleh dari repeated measures analysis of variance (ANOVA) yang membandingkan skor di seluruh pemeriksa. Dari perkiraan varian ini, ada persamaan yang berbeda yang dapat digunakan untuk menghitung nilai ICC tergantung pada jenis pengukuran yang diperoleh (ICC form), dan tujuan yang dimaksudkan pada desain studi reliabilitas (ICC model).

Secara statistik ICC memiliki beberapa keunggulan. Pertama, dapat digunakan untuk menilai reliabilitas diantara dua atau lebih peringkat dan memberikan penerapan klinis yang luas. Kedua, ICC tidak memerlukan jumlah yang pemeriksa yang sama untuk setiap subjek yang memungkinkan untuk fleksibilitas dalam studi klinis. Ketiga, meskipun dirancang terutama untuk digunakan dengan data interval/rasio, ICC dapat diterapkan tanpa distorsi data pada skala ordinal ketika interval antara pengukuran tersebut diasumsikan setara.

Selain itu, dengan data yang dinilai sebagai dikotomi (ada atau tidak adanya sifat), ICC telah terbukti setara dengan pengukuran nominal agreement, menyederhanakan perhitungan dalam kasus dimana lebih dari dua pemeriksa yang terlibat. Oleh karena itu, ICC menyediakan indeks yang berguna dalam berbagai situasi analisis.

Generalizability

Keuntungan utama lainnya dari ICC adalah mendukung model generalisasi yang diusulkan oleh Cronbach sebagai perkiraan komprehensif reliabilitas. Konsep teori generalisasi didasarkan pada gagasan bahwa perbedaan antara nilai yang diamati disebabkan oleh berbagai faktor, bukan hanya varian skor dan kesalahan acak. Perbedaan terjadi karena variasi dalam sistem pengukuran, seperti karakteristik penilai

atau subjek, kondisi pengujian, bentuk alternatif tes, administrasi tes pada kesempatan yang berbeda dan sebagainya. Faktor ini disebut aspek generalizability.

Esensi dari teori generalisasi adalah bahwa aspek berkontribusi pada kesalahan pengukuran sebagai komponen terpisah dari varian yang dibedakan dari kesalahan acak. Dalam teori reliabilitas klasik, kesalahan varian tidak dibedakan dan menggabungkan semua sumber kesalahan pengukuran. Dalam teori generalisasi, kesalahan varian adalah multivarian, lebih lanjut dipartisi untuk memperhitungkan pengaruh aspek tertentu pada kesalahan pengukuran. Oleh karena itu, koefisien generalisasi (ICC) adalah perpanjangan dari koefisien reliabilitas:

Klasifikasi ICC

Ada enam persamaan yang berbeda untuk menghitung ICC yang dibedakan dengan tujuan studi reliabilitas, desain studi, dan jenis pengukuran yang diambil. Hal ini diperlukan untuk membedakan antara pendekatan tersebut.

ICC Form

Koefisien ICC dapat dinyatakan dalam salah satu dari dua bentuk, tergantung apakah nilai yang berasal dari peringkat tunggal atau dari rata-rata peringkat di beberapa pengujian. Umumnya studi reliabilitas dirancang sehingga peringkat tunggal diperoleh dari setiap subjek, ini yang dikenal sebagai form 1. Dalam beberapa studi, skor masing-masing subjek dihitung sebagai nilai mean di beberapa pengujian yang disebut form k. Jika diketahui sebelumnya bahwa pengukuran cenderung tidak stabil, mungkin perlu untuk merekam beberapa uji untuk memastikan reliabilitas yang lebih baik.

Pencatatan rata-rata skor memiliki efek meningkatkan reliabilitas karena kesalahan acak berkurang setiap terjadi apakah over estimasi atau under estimasi. Perhatikan bahwa koefisien reliabilitas yang dihitung berdasarkan nilai rata-rata berarti hanya boleh digeneralisasikan ke pengaturan lain dimana skor juga dicatat. Ini tidak boleh digunakan untuk memperkirakan reliabilitas dalam memperoleh skor pengukuran tunggal.

Model ICC: Random and Fixed Effects

Model ICC tergantung pada apakah hasil analisis reliabilitas harus digeneralisasikan di luar situasi pengujian studi tertentu. Asumsikan, misalnya, bahwa kita ingin menetapkan inter-rater reliability pada empat klinisi yang mengukur skor 10 minutes walking test (10mWT). Pertama-tama kita harus memutuskan apakah klinisi ini dapat dianggap sebagai "dipilih secara acak" dari antara populasi klinisi yang jauh lebih besar dengan latar belakang dan pelatihan yang sama. Jika demikian, maka tujuan utama kita dalam menghitung reliabilitas para peneliti ini adalah untuk menggeneralisasi temuan untuk klinisi lain yang terdiri dari populasi penilai yang lebih besar. Karena akan sangat jarang bagi penyelidik untuk memiliki akses ke semua anggota dari target populasi untuk membuat pilihan yang benar-benar acak, penggunaan istilah "acak" dalam konteks ini adalah teoritis. Titik penting adalah bahwa kita bermaksud untuk menggeneralisasi hasil yang serupa pada penilai lainnya. Oleh karena itu, penilai ini akan mewakili efek acak.

Alternatif lain adalah empat penilai dalam studi dapat merepresentasikan efek tetap (fixed effect). Mungkin kita merencanakan penyelidikan masa depan di mana kita akan melatih para penilai ini dalam mengelola 10mWT untuk mengukur efek dari intervensi eksperimental. Dalam hal ini, satu-satunya tujuan dalam sebuah studi pendahuluan reliabilitas adalah untuk menetapkan bahwa para peneliti khusus ini dapat secara reliabel mampu mengelola 10mWT sesuai dengan protokol studi yang akan datang.

ICC model

Kita tidak akan peduli tentang generalisasi dalam temuan reliabilitas untuk setiap populasi klinisi yang lebih besar. Selain mengidentifikasi penilai, baik sebagai "random" atau "fixed" effect, kita juga harus memperjelas peran subjek. Umumnya, subjek dalam penelitian klinis dianggap sebagai random effect karena penyelidik biasanya bermaksud untuk menggeneralisasi temuan dari studi mereka ke populasi yang lebih besar dari individu yang sama kepada siapa tes dapat diberikan. Perbedaan antara random dan fixed effect membantu kita memilih model ICC yang sesuai.

Shrout dan Fleiss menggambarkan tiga model ICC. Mereka membedakan model ini sesuai dengan bagaimana penilai dipilih dan ditugaskan untuk subjek.

Model 1. Dalam model 1, setiap subjek yang dinilai oleh set yang berbeda (k penilai). Para penilai dianggap dipilih secara acak dari populasi yang lebih besar penilai, sehingga penilai adalah random effect. Namun demikian, penilai untuk satu subjek belum tentu merupakan penilai yang sama yang melakukan pengukuran pada subjek lain. Oleh karena itu, dalam desain ini tidak ada cara untuk mengasosiasikan penilai tertentu dengan variabel yang diukur. Satu-satunya varian yang sebenarnya dapat dinilai adalah perbedaan diantara subjek. Sumber lain varian kesalahan, termasuk penilai atau kesalahan pengukuran, tidak dapat dipisahkan.

Model 2. Model 2 adalah model yang paling sering diterapkan dari ICC untuk menilai inter-rater reliability. Dalam desain ini, setiap subjek dinilai oleh satu set penilai yang sama. Penilai dipilih secara acak yaitu mereka diharapkan untuk mewakili populasi, dan hasil dapat digeneralisasikan kepada yang lain dengan karakteristik yang sama. Subjek juga dianggap dipilih secara acak dari populasi individu yang akan menerima pengukuran. Oleh karena itu, subjek dan penilai keduanya merupakan random effects. Hal ini mungkin hanya bersifat teoritis dalam praktik. Dan maksud dari penelitian ini adalah untuk menunjukkan bahwa reliabilitas pengukuran dapat diterapkan kepada orang lain.

Model 3. Pada model 3, setiap subjek dinilai oleh satu set penilai yang sama. Dalam hal ini, tidak ada niat untuk menggeneralisasi temuan di luar yang terlibat. Dalam desain ini, penilai dianggap sebagai fixed effect karena diseleksi secara purposive (bertujuan), tidak secara acak. Subjek dianggap sebagai random effect. Oleh karena itu, model 3 adalah model campuran. Model ini digunakan ketika seorang peneliti ingin menetapkan bahwa peneliti tertentu dapat diandalkan dalam pengumpulan data mereka, tetapi reliabilitas orang lain tidak relevan. Model 3 juga merupakan statistik yang sesuai untuk mengukur intrarater reliability karena pengukuran peniali tunggal tidak dapat digeneralisasikan kepada penilai lain.

Form ICC: Penilaian tunggal dan rata-rata

Setiap model ICC dapat dinyatakan dalam dua bentuk, tergantung pada apakah skor adalah peringkat tunggal atau rata-rata. Paling sering, studi reliabilitas didasarkan pada perbandingan skor dari individu penilai. Akan tetapi, ada kalanya, jika mean

beberapa penilai atau peringkat dapat digunakan sebagai unit reliabilitas. Misalnya ketika pengukuran tidak stabil, mungkin perlu menggunakan mean dari beberapa pengukuran sebagai skor individu untuk mendapatkan reliabilitas yang memuaskan. Menggunakan skor rata-rata memiliki efek meningkatkan perkiraan reliabilitas dimana mean dianggap lebih baik dalam memperkirakan skor yang secara teoritis mengurangi kesalahan varian.

Enam jenis ICC diklasifikasikan menggunakan dua angka dalam tanda kurung. Nomor pertama menunjuk model (1, 2, atau 3), dan angka kedua menandakan bentuk, menggunakan salah satu pengukuran (1) atau mean dari beberapa pengukuran (k) sebagai unit analisis. Sebagai contoh, ketika menggunakan pengukuran tunggal dalam sebuah studi generalisasi, kita akan menentukan penggunaan ICC (2, 1). Jenis ICC yang digunakan harus selalu diindikasikan.

Dengan demikian, terdapat enam jenis ICC:

- model 1: ICC (1, 1) atau ICC (1, k)
- model 2: ICC (2, 1) atau ICC (2, k)
- model 3: ICC (3, 1) atau ICC (3, k)

k mewakili jumlah nilai yang digunakan untuk mendapatkan mean, dan dapat digantikan dengan jumlah aktual skor ketika menotasi ICC. Untuk contoh di atas:

- dalam studi inter-rater reliability, ICC (2, 3) akan digunakan untuk menunjukkan random model dengan mean dari tiga percobaan yang dicatat sebagai skor.
- dalam studi intra-rater reliability, penilai mengambil satu pengukuran di setiap percobaan. Oleh karena itu, ICC (3, 1) akan digunakan untuk menunjukkan model campuran dengan nilai tunggal.

Analisis varians

ICC didasarkan pada ukuran varians yang diperoleh dari ANOVA. Untuk sebuah studi interrater reliability, penilai adalah variabel bebas, sementara untuk studi intrarater, percobaan adalah variabel bebas.

Analysis of Variance

ICC didasarkan pada varians perkiraan yang diperoleh dari repeated measures ANOVA. Hasil untuk data 10mWT mencakup nilai yang Diperoleh dari ANOVA bersama dengan output ICC untuk model 2 dan model 3. Di ANOVA, antara varian item mencerminkan perbedaan skor antar pemeriksa.

Jika rasio F signifikan, hal ini menunjukkan bahwa varian antar skor yang diberikan oleh penilai yang berbeda adalah besar dibandingkan dengan kesalahan residual yang menunjukkan bahwa skor penilai sangat berbeda antara satu dengan yang lain dan mungkin tidak reliabel. Dalam contoh ini, rasio F tidak signifikan secara statistik ($p = .554$), menunjukkan interrater reliability yang tinggi.

Varian antar orang mencerminkan perbedaan antara skor peserta studi. Heterogenitas diantara peserta adalah karakteristik penting dari sampel untuk pengujian reliabilitas, dan heterogenitas ini tercermin dalam mean square value yang tinggi antara satu dengan yang lain. Dalam contoh ini, efek antara satu dengan yang lain signifikan ($p < .001$). Ini adalah kondisi yang diharapkan dan diperlukan untuk pengujian reliabilitas kecuali jika tes F ini tidak signifikan.

Mode/ 1: One-Way ANOVA

Untuk model 1, dimana subjek sebagai variabel bebas. Efek antara subjek memberikan informasi kepada kita jika skor subjek berbeda satu sama lain sesuai yang kita harapkan. Komponen kesalahan mewakili variasi subjek di seluruh penilaian. Beberapa kesalahan ini akan terjadi karena perubahan dari satu percobaan ke percobaan lain.

Model 2 dan 3: Repeated Measures ANOVA

Untuk model 2, ANOVA dilakukan dalam model two-way random effect dimana subjek dan penilai dianggap dipilih secara acak dari populasi yang lebih besar. Karena itu, hasil penelitian dapat digeneralisasikan kepada subjek dan penilai lainnya. Untuk model 3, dilakukan, dengan penilai sebagai efek tetap (tidak dipilih secara acak) dan subjek sebagai efek acak. Hasil analisis numerik sebenarnya akan sama untuk jenis ANOVA acak dan campuran. Satu-satunya perbedaan akan terletak pada interpretasi data.

Output SPSS untuk ICC

Paket perangkat lunak yang umum digunakan akan menghasilkan berbagai bentuk ICC sebagai bagian dari analisis reliabilitas (dalam skala).

ANOVA ①						
		Sum of Squares	df	② Mean Square	F	Sig.
Between People		99.375	5	BMS 19.875	2.203	④ .130
Within People	Between Items	21.792	3	RMS 7.264		
	Residual	49.458	15	EMS 3.297		
	Total	71.250	18	3.958		
Total		170.625	23	7.418		

$$ICC(2, 1) = \frac{BMS - EMS}{BMS + (k - 1) + \frac{k(RMS - EMS)}{n}} = \frac{19.88 - 3.30}{19.88 + (4 - 1)3.30 + \frac{4(7.26 - 3.30)}{6}} = .51$$
$$ICC(2, k) = \frac{BMS - EMS}{BMS + \frac{(RMS - EMS)}{n}} = \frac{19.88 - 3.30}{19.88 + \frac{(7.26 - 3.30)}{6}} = .81$$
$$ICC(3, 1) = \frac{BMS - EMS}{BMS + (k - 1)EMS} = \frac{19.88 - 3.30}{19.88 + (4 - 1)3.30} = .56$$
$$ICC(3, k) = \frac{BMS - EMS}{BMS} = \frac{19.88 - 3.30}{19.88} = .83$$

The values for BMS, RMS and EMS are taken from the ANOVA as shown above; *k* is the number of raters (or ratings), which will be the data in columns; *n* is the number of subjects or rows (see Table 26.2).

Tabel 1. Output ICC

Confidence intervals (CI)

Karena nilai ICC diperoleh dengan menggunakan data dari sampel subjek studi, CI dapat menunjukkan berbagai nilai dimana ICC dalam populasi target. Seperti yang diharapkan, dengan melihat skor rata-rata dibanding skor individual, random error menurun dan reliabilitas meningkat saat menghitung ICC (2, k). CI memberikan konteks untuk menafsirkan nilai ICC yang dihitung menggunakan data sampel studi .

Interpretasi ICC

Seperti bentuk lain dari reliabilitas, tidak ada nilai standar reliabilitas yang dapat diterima dengan menggunakan ICC. Klinisi yang melakukan praktik berbasis bukti harus menentukan "berapa banyak" reliabilitas yang diperlukan untuk membenarkan penggunaan alat tertentu untuk suatu tujuan tertentu.

Penjelasan nilai ICC yang rendah

Ada dua alasan utama mengapa nilai ICC rendah. Penjelasan pertama adalah para penilai atau skor test-retest tidak setuju. Ini bukan interpretasi langsung namun sering terjadi jika terdapat lebih dari dua penilai atau pengukuran. Sebagai contoh, ICC adalah rata-rata berdasarkan varian di semua penilai dan ketidaksetujuan mungkin melibatkan semua, beberapa, atau hanya satu penilai. Kita dapat melihat cara skor atau grafik setiap rater untuk dapat membantu dalam menentukan ketidaksetujuan diantara mereka.

Alasan kedua berkaitan dengan varian yang tidak mencukupi diantara subjek studi. Kurangnya variabilitas dapat terjadi ketika subjek yang homogen pada atribut diukur, atau ketika sistem rating itu sendiri membatasi kemungkinan skor hanya pada rentang yang sangat sempit. Nilai ICC menunjukkan proporsi dari total varian di satu set skor yang disebabkan oleh heterogenitas di antara peserta studi (setelah sumber varian lain telah diperhitungkan).

Jika nilai subjek sangat mirip satu sama lain, kontribusi relatif dari sumber varian lainnya termasuk varian di seluruh penilai untuk varian total yang diamati dalam dataset akan meningkat, sehingga nilai ICC deflasi. Efek ini dapat diperiksa dengan melihat besarnya dan signifikansi dari antara orang sumber varian dalam tabel ANOVA . Jika nilai subjek terlalu homogen, nilai ini tidak akan signifikan, dan nilai ICC akan dideflasi.

Magnitude ICC

Seperti bentuk lain dari reliabilitas, tidak ada nilai standar untuk reliabilitas yang dapat diterima menggunakan ICC. ICC berkisar antara 0,00 dan 1,00, dengan nilai lebih dekat dengan 1,00 mewakili reliabilitas yang lebih kuat. Tetapi karena reliabilitas adalah karakteristik pengukuran yang diperoleh untuk berbagai derajat meskipun jarang yang sempurna, peneliti harus menentukan berapa banyak reliabilitas yang diperlukan untuk membenarkan penggunaan alat tertentu. Sifat variabel terukur akan menjadi faktor, dalam hal stabilitas dan ketepatan yang diperlukan untuk membuat penilaian klinis suara tentang hal itu.

Sebagai pedoman umum, nilai di atas 0,75 adalah indikasi reliabilitas yang baik, dan yang di bawah 0,75 adalah nilai reliabilitas poor ke moderat. dalam banyak

pengukuran klinis, reliabilitas harus lebih dari 0,90 untuk memastikan validitas yang wajar. Ini hanya pedoman, dan tidak boleh digunakan sebagai standar mutlak. Para peneliti dan klinisi harus mempertahankan penilaian mereka dalam konteks skor spesifik yang dinilai dan tingkat presisi yang dapat diterima dalam pengukuran.

Kesalahan penilai

Pada nilai ICC yang tinggi kita dapat mengatakan bahwa reliabilitas yang baik, dan untuk mengekspresikan keyakinan dalam pengukuran. Ketika reliabilitas kurang dari memuaskan, peneliti berkewajiban untuk menyortir melalui penjelasan alternatif untuk menentukan sumber kontribusi kesalahan.

Agreement

Ketika unit pengukuran berada pada skala kategori, reliabilitas dinilai dengan tepat dengan pengukuran agreement. Indeks yang paling sederhana adalah dengan menggunakan persen agreement. Ini adalah ukuran seberapa sering penilai setuju terhadap skor yang diberikan kepada individu subjek atau seberapa sering setuju terhadap test-retest. Untuk menentukan reliabilitas sejati dari data kategori, kita harus mempertimbangkan kemungkinan bahwa beberapa bagian dari hasil bisa terjadi secara kebetulan; yaitu, jika dua penilai adalah untuk menetapkan subjek untuk kategori sepenuhnya secara acak, beberapa tingkat kesepakatan masih akan diharapkan.

Kesimpulan

Karena masalah reliabilitas sangat penting untuk validitas ilmu klinis, basis statistik untuk menafsirkan reliabilitas harus dipahami oleh mereka yang melakukan penelitian dan mereka yang membaca laporan penelitian. Apa yang kita pelajari dari melihat melalui literatur profesional adalah bahwa metode yang digunakan untuk menganalisis reliabilitas tampaknya bervariasi dengan peneliti yang berbeda dan dalam disiplin ilmu yang berbeda. Meskipun ahli statistik telah mengatasi masalah reliabilitas untuk waktu yang lama, tidak ada konsensus tentang bagaimana data reliabilitas ditangani.

Memilih pendekatan tertentu untuk reliabilitas pengujian harus didasarkan pada pemahaman tentang sifat variabel respon, apa jenis interpretasi yang diinginkan, dan apa masalah pengukuran terbesar yang menjadi fokus peneliti. Pertimbangan harus diberikan pada skala pengukuran, jumlah variabilitas yang dapat diharapkan dalam skor sampel, dan unit pengukuran apa yang digunakan. Kita harus menyadari aplikasi yang dimaksudkan data dan tingkat presisi yang diperlukan untuk membuat keputusan klinis yang aman dan bermakna.

Referensi

Keppel G. *Design and Analysis: A Researcher's Handbook* (4th ed.). Englewood Cliffs, NJ:Prentice Hall, 2004.

Leslie G. Portney, Mary P. Watkins. *Foundations of Clinical Research, Applications to Practice* 2nd ed. F.A Davis Company, Philadelphia 2015.

Dennis Howitt, Duncan Cramer, *Introduction to SPSS statistics in psychology : for version 19 and earlier* 5th ed, Person Education limited Edinburgh Gate Harlow England 2011.

Leiyu Shi, *Health Services Research Methods* 2nd ed, Thomson Delmar Learning United States 2009.

Hilla Brink, Christa van der Walt, Gisela van Rensburg. *Fundamentals of Research Methodology for Health Care Professionals*. Juta & Co. (Pty) Ltd Lansdowne Cape Town, 2009.

Leslie G. Portney. *Foundations of clinical research : applications to evidence-based practice*. F.A Davis Company, Philadelphia 2020