



Modul 12

FEB 326-Evidence-Based Practice Fisioterapi

Materi 12

Critical Appraisal Terkait Pengukuran Outcome

Disusun Oleh

Wahyuddin

UNIVERSITAS ESA UNGGUL

2019

## Pendahuluan

Pengukuran outcome merupakan komponen penting dalam penelitian untuk mendapatkan hasil atau kesimpulan studi yang akurat. Hal ini juga menjamin data yang dihasilkan valid dan berdampak kepada validitas hasil. Selain itu pengukuran outcome dapat digunakan untuk mengevaluasi dampak dari suatu intervensi.

Pengukuran outcome memberikan informasi tentang pasien dan kemajuan mereka. Memilih pengukuran outcome yang tepat pada pasien penting untuk memahami status dan kemajuan mereka dari waktu ke waktu. Idealnya, psychometric properties suatu pengukuran outcome yang digunakan dalam praktek telah dikembangkan dan diuji melalui serangkaian studi penelitian. Psychometric properties berupa reliability, validity, dan kebermaknaan klinis. Memahami penelitian yang digunakan untuk mengembangkan dan menilai hasil psychometric properties mengarahkan kita untuk memilih dan menggunakan outcome pengukuran yang efektif dalam praktik klinis.

## Jenis dan Utilitas Outcome Pengukuran

Pengukuran outcome adalah karakteristik atau kualitas diukur untuk menilai status pasien. Pengukuran outcome umumnya dilakukan pada awal, tengah, dan akhir penanganan. Pengukuran outcome dapat dibagi menjadi 2 kategori:

- (1) kuesioner
- (2) performance-based measures

Kuesioner dilakukan dengan cara melengkapi apakah dilakukan oleh fisioterapis yang melakukan wawancara pada pasien atau pasien secara independen menyelesaikan kuesioner. Kuesioner umumnya diberikan skor pada setiap jawaban pasien. Performance-based measures memerlukan pasien untuk melakukan serangkaian gerakan atau tugas. Penilaian dapat didasarkan pada tujuan pengukuran (misalnya, waktu untuk menyelesaikan suatu tugas) atau penilaian kualitatif yang ditetapkan (misalnya, melakukan mekanika normal atau abnormal pada untuk suatu tugas).

Dua outcome pengukuran yang umum digunakan modified Oswestry Disability Questionnaire dengan menggunakan serangkaian pertanyaan untuk mengukur disabilitas yang berhubungan dengan nyeri punggung bawah pada gambar 1 berikut ini.

### Modified Low Back Pain Disability Questionnaire

This questionnaire has been designed to give your therapist information as to how your back pain has affected your ability to manage in everyday life. Please answer every question by placing a mark in the **one** box that best describes your condition today. We realize you may feel that two of the statements may describe your condition, but **please mark only the box that most closely describes your current condition.**

#### Pain Intensity

- I can tolerate the pain I have without having to use pain medication.
- The pain is bad, but I can manage without having to take pain medication.
- Pain medication provides me with complete relief from pain.
- Pain medication provides me with moderate relief from pain.
- Pain medication provides me with little relief from pain.
- Pain medication has no effect on my pain.

#### Personal Care (e.g. Washing, Dressing)

- I can take care of myself normally without causing increased pain.
- I can take care of myself normally, but it increases my pain.
- It is painful to take care of myself, and I am slow and careful.
- I need help, but I am able to manage most of my personal care.
- I need help every day in most aspects of my care.
- I do not get dressed, I wash with difficulty, and I stay in bed.

#### Lifting

- I can lift heavy weights without increased pain.
- I can lift heavy weights, but it causes increased pain.
- Pain prevents me from lifting heavy weights off the floor, but I can manage if the weights are conveniently positioned (e.g., on a table).
- Pain prevents me from lifting heavy weights, but I can manage light to medium weights if they are conveniently positioned.
- I can lift only very light weights.
- I cannot lift or carry anything at all.

#### Walking

- Pain does not prevent me from walking any distance.
- Pain prevents me from walking more than 1 mile.
- Pain prevents me from walking more than 1/2 mile.
- Pain prevents me from walking more than 1/4 mile.
- I can walk only with crutches or a cane.
- I am in bed most of the time and have to crawl to the toilet.

#### Sitting

- I can sit in any chair as long as I like.
- I can only sit in my favorite chair as long as I like.
- Pain prevents me from sitting for more than 1 hour.
- Pain prevents me from sitting for more than 1/2 hour.
- Pain prevents me from sitting for more than 10 minutes.
- Pain prevents me from sitting at all.

#### Standing

- I can stand as long as I want without increased pain.
- I can stand as long as I want, but it increases my pain.
- Pain prevents me from standing for more than 1 hour.
- Pain prevents me from standing for more than 1/2 hour.
- Pain prevents me from standing for more than 10 minutes.
- Pain prevents me from standing at all.

#### Sleeping

- Pain does not prevent me from sleeping well.
- I can sleep well only by using pain medication.
- Even when I take medication, I sleep less than 6 hours.
- Even when I take medication, I sleep less than 4 hours.
- Even when I take medication, I sleep less than 2 hours.
- Pain prevents me from sleeping at all.

#### Social Life

- My social life is normal and does not increase my pain.
- My social life is normal, but it increases my level of pain.
- Pain prevents me from participating in more energetic activities (e.g., sports, dancing).
- Pain prevents me from going out very often.
- Pain has restricted my social life to my home.
- I have hardly any social life because of my pain.

#### Traveling

- I can travel anywhere without increased pain.
- I can travel anywhere, but it increases my pain.
- My pain restricts my travel over 2 hours.
- My pain restricts my travel over 1 hour.
- My pain restricts my travel to short necessary journeys under 1/2 hour.
- My pain prevents all travel except for visits to the physician/therapist or hospital.

#### Employment/Homemaking

- My normal homemaking/job activities do not cause pain.
- My normal homemaking/job activities increase my pain, but I can still perform all that is required of me.
- I can perform most of my homemaking/job duties, but pain prevents me from performing more physically stressful activities (e.g., lifting, vacuuming).
- Pain prevents me from doing anything but light duties.
- Pain prevents me from doing even light duties.
- Pain prevents me from performing any job or homemaking chores.

Copyright 2001 and 2007 American Physical Therapy Association

FIGURE 10.2 (A) Modified Oswestry Disability Questionnaire. Modified by Fritz & Irrgang The Chartered Society of Physiotherapy, from Fairbanks JCT, Couper J, Davies JB, et al. The Oswestry Low Back Pain Disability Questionnaire. *Physiotherapy*. 1980;66:271-273; with permission.

## Gambar 1. Modified Low Back Pain Disability Questionnaire

Berg Balance Scale adalah suatu bentuk performance-based measure untuk mengukur keseimbangan pada gambar 2. Perhatikan bahwa Oswestry Disability Questionnaire merupakan serangkaian pertanyaan yang dijawab oleh pasien, sedangkan Berg Balance Scale memerlukan pasien untuk secara fisik melakukan serangkaian tugas yang kemudian dicirikan oleh fisioterapis.

### Berg Balance Test

INSTRUCTIONS: Demonstrate each task and/or give instructions as written. When scoring, record the lowest response category for each item. Patient cannot use an assistive device for this test.

1. **Sitting to Standing: Please stand up**  
*Instructions: Try not to use your hands for support.*
  - 4 Able to stand independently, does not use hands and is stable
  - 3 Able to stand independently using hands
  - 2 Able to stand using hands after several tries
  - 1 Needs min assistance to stand or stabilize
  - 0 Needs mod-max assistance to stabilize
2. **Standing Unsupported**  
*Instructions: Stand for 2 minutes without holding onto anything.*
  - 4 Stands safely for 2 minutes
  - 3 Stands for 2 minutes with supervision
  - 2 Stands unsupported for < 2 minutes
  - 1 Needs several tries to stand 30 seconds
  - 0 Unable to stand 30 seconds unassisted
3. **Standing to Sitting: Please sit down**  
*Instructions: Try to use your hands as little as possible.*
  - 4 Sits with minimal use of hands
  - 3 Controls descent with hands
  - 2 Uses back of legs against chair to control descent
  - 1 Sits independently but has uncontrolled descent
  - 0 Needs assistance to sit
4. **Sitting Unsupported with Feet on Floor**  
*Instructions: Sit with your arms folded across chest for 2 minutes.*
  - 4 Sits safely for 2 minutes
  - 3 Sits 2 minutes with supervision
  - 2 Sits 30 seconds
  - 1 Sits 10 seconds
  - 0 Unable to sit for 10 seconds unsupervised
5. **Transfers**  
*Instructions: Please move from the chair you are sitting in to the other chair (place chairs at 45° angle, one chair with arm rests, one without).*
  - 4 Transfers safely with minimal use of hands
  - 3 Transfers safely with definite need of hands
  - 2 Transfers with verbal cues or supervision
  - 1 Needs 1 person assist or supervision for safety
  - 0 Needs 2 person assist
6. **Standing Unsupported with Eyes Closed**  
*Instructions: Stand with feet shoulder-width apart. Close your eyes and stand still for 10 seconds.*
  - 4 Stands for 10 seconds safely
  - 3 Stands for 10 seconds with supervision
  - 2 Stands 3 seconds
  - 1 Unable to keep eyes closed for 3 seconds but stays steady
  - 0 Needs help to keep from falling
7. **Standing Unsupported with Feet Together**  
*Instructions: Place your feet together and stand without holding. Hold for 1 minute.*
  - 4 Able to place feet together independently and stand 1 minute safely
  - 3 Places feet together independently and stands for 1 min with supervision
  - 2 Places feet together independently but unable to hold for 30 seconds
  - 1 Needs help to attain position but able to stand for 15 seconds
  - 0 Needs help to attain position and unable to hold for 15 seconds
8. **Reaching Forward with Outstretched Arm**  
*Instructions: Raise your arm to shoulder height and reach forward with your fist as far as you can.*
  - 4 Greater than 10 inches
  - 3 Greater than 5 inches
  - 2 Greater than 2 inches
  - 1 Reaches forward but needs supervision
  - 0 Needs help to keep from falling
9. **Pick Up Object from Floor**  
*Instructions: Please pick up plastic cup. (Place on floor 6-12 inches in front of feet.)*
  - 4 Able to pick up safely and easily
  - 3 Able to pick up but needs supervision
  - 2 Unable to pick up but reaches 1-2 inches from object and keeps balance independently
  - 1 Unable to pick up and needs supervision while trying
  - 0 Unable to try or tries and needs assist to keep from falling
10. **Turn to Look Behind over Left and Right Shoulders**  
*Instructions: Turn to look behind you over toward the left shoulder without moving your feet. Repeat with the right.*
  - 4 Looks behind from both sides and weight shifts well
  - 3 Looks behind one side only, other side shows less weight shift
  - 2 Turns sideways only but maintains balance
  - 1 Needs supervision while turning
  - 0 Needs assistance to keep from falling
11. **Turning 360°**  
*Turn completely around in full circle. Pause. Now turn in the other direction.*
  - 4 Turns full circle in < 4 seconds each side.
  - 3 Turns safely to only one side in < 4 seconds
  - 2 Turns safely but slowly
  - 1 Needs close supervision or verbal cuing
  - 0 Needs assist while turning
12. **Dynamic Weight Shifting while Standing Unsupported**  
*Instructions: Place each foot alternately on the stool for a count of 8 (demonstrate with foot flat; count touches).*
  - 4 Completes 8 steps in 20 seconds safely
  - 3 Completes 8 steps in > 20 seconds
  - 2 Completes 4 steps without aid and with supervision
  - 1 Able to complete > 2 steps, needs min assist
  - 0 Needs assist to keep from falling/unable to try
13. **Standing Unsupported One Foot in Front**  
*Instructions: Place one foot directly in front of other as close as possible and hold for 30 sec.*
  - 4 Places foot in tandem independently and holds 30 seconds
  - 3 Places one foot in front of other independently and holds for 30 seconds
  - 2 Takes small step independently and holds for 30 seconds
  - 1 Needs help to place foot, holds 15 seconds
  - 0 Loses balance while stepping or standing
14. **Standing on One Leg**  
*Instructions: Stand on one leg for as long as you can without holding.*
  - 4 Lifts leg independently and holds > 10 seconds
  - 3 Lifts leg independently and holds 5-10 seconds
  - 2 Lifts leg independently and holds ≈ 3 seconds
  - 1 Tries to lift leg, unable to hold 3s but remains standing independently
  - 0 Unable to try or needs assist to prevent fall

Total Score: \_\_\_\_\_ (maximum = 56)

FIGURE 10.2 cont'd (B) The Berg Balance Scale. From: Berg KQ, Wood-Douphinee SL, Williams JL, et al. Measuring balance in the elderly: validation of an instrument. Can J Public Health. 1992;83(suppl 2):7-11; with permission.

### Gambar 2. Berg Balance Scale

Terkait konsep ICF pada berbagai level, dapat dilihat pada tabel 1 berikut ini:

OUTCOME MEASURE	QUESTIONNAIRE OR PERFORMANCE BASED	WHAT IT MEASURES	ICF LEVEL MEASURED
Geriatric Depression Scale <sup>8</sup>	Questionnaire	Screens for depression in older adults	Body function
Mini-Mental State Exam <sup>26</sup>	Performance-based	Cognitive function	Body function
Activities-Specific Balance Confidence Scale (ABC Scale) <sup>27</sup>	Questionnaire	Balance confidence	Activity
10-Meter Walk Test <sup>28</sup>	Performance-based	Walking speed	Activity
Pediatric Evaluation of Disability Inventory (PEDi) <sup>29</sup>	Questionnaire	Children's functional capabilities	Activity
Neck Disability Index <sup>4</sup>	Questionnaire	Impact of neck pain on life participation	Participation
Oswestry Disability Questionnaire <sup>14</sup>	Questionnaire	Impact of low back pain on life participation	Participation

Tabel 1. Level Pengukuran Outcome

#### Apa Yang Membuat Suatu Pengukuran Bermanfaat di Klinis?

Agar dapat bermanfaat secara klinis, harus dapat menunjukkan psychometric properties berkualitas tinggi. Sebagai contoh terkait aspek reliability, validity, and kemaknaan klinis sebagai berikut:

- Reliability merujuk kepada konsistensi hasil pengukuran. Misalnya, jika kita menggunakan Berg Balance Scale bagi pasien sama pada hari yang sama, kita haru mendapatkan skor yang serupa. Ini adalah contoh inter tester reliability.
- Validity merujuk kepada hasil mengukur kemampuan untuk mengukur karakteristik atau fitur yang akan diukur. Misalnya, jika kita memilih neck disability index untuk mengukur dampak sakit leher pada kemampuan pasien untuk berpartisipasi dalam peran kehidupan peran, sangat penting untuk mengetahui apakah neck disability index benar-benar mengukur dampak sakit leher pada partisipasi dalam peran kehidupan. Ini adalah contoh dari content validity.
- Kemaknaan klinis mengacu pada kemampuan hasil pengukuran untuk memberikan informasi kepada klinisi dan pasien dengan informasi konsekuensial. Misalnya, jika kita menggunakan berjalan tes 10 meter untuk mengukur kecepatan berjalan pasien sebelum dan setelah terapi, kita mungkin mempertanyakan hubungan antara peningkatan berjalan kecepatan dan kemampuan pasien terhadap fungsi dalam kehidupan sehari-hari.

#### Studi Asesmen Reliabilitas Pengukuran Outcome

Berbagai studi desain digunakan untuk mengeksplorasi keandalan mengukur hasil. Kami membahas empat jenis keandalan dan desain studi yang digunakan untuk menilai mereka. Pertama, internal consistency dan test-retes reliability, mengukur konsistensi ukuran itu sendiri, sedangkan kedua dua, mengukur konsistensi penilai yang melakukan test.

### Internal Consistency

Internal consistency menetapkan sejauh mana beberapa item dalam mengukur hasil (misalnya, item 14 Berg balance scale) mencerminkan konstruksi yang sama. Sebagai contoh, Berg balance scale ini bertujuan untuk mengukur keseimbangan. Untuk menentukan internal konsistensi skala, barang antar korelasi umumnya dinilai menggunakan uji statistik Cronbach alpha yang berkisar dari 0 ke 1. Nilai mendekati 0 menunjukkan bahwa setiap item ukuran tidak berkorelasi dengan satu sama lain. Sementara nilai mendekati 1 menunjukkan bahwa setiap item memiliki korelasi yang kuat.

Secara praktis nilai Cronbach alpha harus lebih besar dari 0,70 dan kurang 0,90. Nilai-nilai yang lebih besar daripada 0,90 menyarankan pengulangan (korelasi berlebihan) antara item dalam pengukuran outcome. Berg Balance Scale menunjukkan internal konsistensi tinggi antara orang dengan nilai alpha Cronbach 0.83 dan untuk orang dengan stroke dengan nilai Cronbach alpha 0.97.7 nilai ini menunjukkan bahwa 14 item mempunyai konstruksi yang sama.

### Test-Retest Reliability

Test-retest reliability menetapkan sejauh yang mengukur hasil menghasilkan hasil yang sama ketika berulang kali diterapkan pada pasien yang tidak mengalami perubahan dalam ciri-ciri yang diukur. Misalnya, jika keseimbangan pasien tidak berubah antara hari Senin dan Jumat minggu yang sama, skor pasien pada ukuran keseimbangan harus pada dasarnya sama pada hari Senin dan pada hari Jumat. Dalam tes tes ulang studi, rating satu ukuran peserta sama dua atau lebih kali, biasanya selama beberapa hari. Peserta dipilih untuk studi harus diharapkan untuk tetap sama atau serupa dalam ciri-ciri yang diukur.

### Intra-Rater Reliability

Intra-rater reliability merujuk kepada konsistensi yang mengukur hasil pengukuran ketika digunakan oleh fisioterapis sama pada pasien sama. Intra-rater reliability penting dalam praktek klinis untuk membandingkan hasil yang dilakukan oleh rater yang sama dari waktu ke waktu, misalnya, pada awal dan akhir terapi. Studi Intra-rater reliability melibatkan penilai (misalnya, fisioterapis) yang diminta untuk melakukan mengukur dua kali atau lebih pada individu sama.

Perbedaan antara dua sifat psikometrik adalah bahwa test-retest reliability mengacu pada hasil pengukuran itu sendiri, sedangkan intra-rater reliability mengacu pada keterampilan penilai melakukan tes.

### Inter-Rater Reliability

Inter-rater reliability merujuk kepada konsistensi penilai berbeda yang menghasilkan skor yang sama pada pasien sama. Hal ini penting dalam praktek klinis. Sebagai contoh, pertimbangkan situasi di mana pasien dengan total knee arthroplasty jarak gerak sendi untuk gerakan fleksi sebesar 90 derajat setelah 1 minggu terapi. Namun, ketika rekan lain mengukur pada pasien sama, mereka mendapatkan nilai 100 derajat. Selain meninjau inter-rater reliability, kita akan membandingkan teknik pengukuran untuk meningkatkan dalam praktek klinis.

Studi inter-rater reliability membutuhkan dua atau lebih penilai untuk menilai pasien sama. Studi harus mencakup pasien dengan nilai pada berbagai kemungkinan sesuai pengukuran. Konsistensi nilai antara penilai dihitung untuk menentukan reliability antara penilai. Ketika nilai inter-rater reliability rendah, ada

berbagai variabilitas dalam nilai. Ketika nilai inter-rater reliability tinggi, ada sedikit variabilitas.

#### Penilaian Kualitas Studi Reliabilitas Pengukuran Outcome

Bagian A: menentukan penerapan studi reliability sama dengan penentuan untuk jenis studi lain.

Bagian B: menentukan kualitas reliability yang mempelajari pertanyaan-pertanyaan penting untuk bertanya tentang kualitas untuk studi reliability.

Bagian C: Interpretasi hasil studi reliability menggunakan beberapa metode statistik umum yang dapat digunakan untuk membuat perbandingan test-retest reliability, intra-rater reliability dan interrater reliability (misalnya, intraclass korelasi koefisien [ICC],

#### Bagian D: Clinical Bottom Line

Clinical bottom line akan menginformasikan apakah ukuran cukup handal untuk digunakan dalam praktek klinis. Selain itu, penelitian tentang hasil tindakan harus memberikan rincian tentang metode yang digunakan untuk melakukan ukuran. Replikasi protokol yang digunakan dalam studi berkualitas tinggi dapat meningkatkan kehandalan Anda dengan ukuran. Itu tidak mungkin persis menduplikasi protokol studi dalam praktek klinis. Kita harus mengharapkan bahwa reliability ukuran yang dapat menurunkan hasil modifikasi dilaksanakan.

Pertanyaan-pertanyaan yang dapat diajukan untuk menilai studi terkait reliability dapat dilihat pada tabel 2 berikut ini:

QUESTION	ANSWER	WHY IT MATTERS
1. Is the study's purpose relevant to my clinical question?	Yes____ No____	If the purpose of the study is not relevant to your question, it is unlikely to be applicable.
2. Are the inclusion and exclusion criteria clearly defined and would my patient qualify for the study?	Yes____ No____	Ideally, the study inclusion and exclusion criteria for both study participants (e.g., patients) and testers (e.g., therapists) should be similar to your clinical environment. If they are not, you need to take this into consideration when integrating the study findings into your practice.
3. Are the outcome measures relevant to my clinical question and are they conducted in a clinically realistic manner?	Yes____ No____	Consider if the outcome measure was applied in a clinically replicable manner. This includes consideration of the amount of training provided to testers.
4. Is the study population (sample) sufficiently similar to my patient and to my clinical setting to justify the expectation that I would experience similar results?	Yes____ No____	Beyond the inclusion and exclusion criteria (question 2), it is important to consider the similarity of the study participants and therapists to your clinical environment.
5. Consider the number of participants being rated and the number of raters. Does the study have sufficient sample size?	Yes____ No____	Sufficient sample size is required to reduce the chance for random error to affect the results. A study that reports the method used to determine the minimum sample size (i.e., sample size calculation) has higher quality.
6. Do participants have sufficient diversity to assess the full range of the outcome measure?	Yes____ No____	A study of reliability should include the full spectrum of people that the outcome measure was designed to test. For example, consider a measure designed to assess a child's ability to engage in elementary school activities. Study participants should include all ages of elementary school children and children with a full range of participation abilities.
7. Are study participants stable in the characteristic of interest?	Yes____ No____	If the participants in a reliability study change substantially in the characteristic of interest, the reliability result will not be accurate. For example, in the case of an outcome measure designed to assess urinary incontinence, it is important that the participants not change any medication regimen that could change the severity of their incontinence between study assessments. Such a change would produce erroneous results for a study of reliability.
QUESTION	ANSWER	WHY IT MATTERS
8. Is there evidence that the outcome measure was conducted in a reasonably consistent manner between assessments?	Yes____ No____	If it is conducted differently between assessments, the reliability of the outcome measure will be inaccurate. Consider a questionnaire completed by patients regarding their satisfaction with therapy services. If the first assessment is conducted in a quiet, private room and the second assessment is conducted over the phone and without privacy, patients might be prone to give different answers. This scenario would give inaccurate estimates of the measure's test-retest reliability.
9. Were raters blinded to previous participant scores?	Yes____ No____	Raters should not have knowledge of previously collected scores. For example, for a study of PT students' inter-rater reliability in taking manual blood pressure measurements, students must be blinded to the measures collected by other students. Lack of blinding is likely to produce inaccurate results.
10. Is the time between assessments appropriate?		If assessments are conducted within too short a time interval, the reliability may be inaccurate because raters (or participants) are influenced by the initial test. Consider a study of the intra-rater reliability of manual muscle testing. If assessments are taken with only a minute between tests, the raters are likely to remember the strength scores from one test to another and participants may have fatigue that causes a decrease in their strength. In this case, rater recall could <i>overestimate</i> intra-rater reliability and participant fatigue could <i>underestimate</i> intra-rater reliability, making the results difficult to interpret.

Tabel 2. Pertanyaan-Pertanyaan Penilaian Kualitas Reliability



## Jenis Validity

Validitas ukuran hasil yang mengacu pada ukuran relevansi untuk membangun dasar yang dimaksudkan untuk mengukur. Misalnya, jika Anda memilih ukuran hasil untuk layar untuk depresi pada orang dewasa yang lebih tua, Anda mungkin mengidentifikasi geriatric depression scale. Sebelum menggunakan ukuran ini untuk menentukan jika pasien geriatri Anda berisiko untuk depresi, Anda perlu tahu jika ini ukuran adalah ukuran berlaku depresi. Apakah itu benar-benar mengidentifikasi depresi dalam older orang dewasa? Sebuah studi validasi skala depresi geriatrik akan membantu Anda menjawab pertanyaan ini.

Konten validitas menetapkan bahwa ukuran hasil yang mencakup semua karakteristik yang dimaksudkan untuk mengukur. Misalnya, jika ukuran hasil yang menilai keterbatasan aktivitas yang berhubungan dengan nyeri punggung rendah, studi tentang keabsahan konten akan berusaha untuk menetapkan bahwa semua aspek kegiatan yang mungkin terkena sakit punggung rendah disertakan. Sebuah panel ahli subjektif menetapkan validitas konten. Masing-masing ahli menentukan apakah ukuran hasil mencakup semua aspek karakteristik yang dimaksudkan untuk mengukur. Para ahli juga mencari elemen-elemen ukuran yang dimasukkan karakteristik yang menarik. Informasi ini kemudian digunakan untuk mengubah ukuran sampai panel dan pengembang percaya bahwa validitas konten memuaskan telah dicapai.

Evaluasi validitas konten yang kurang formal disebut sebagai wajah validitas. Wajah validitas didasarkan pada evaluasi informal oleh ahli yang ukuran muncul untuk mengukur apa yang dimaksudkan untuk mengukur. Jika Anda memutuskan untuk mengukur kepuasan pasien di klinik Anda, Anda mungkin mengembangkan kuesioner yang meminta pasien tentang kepuasan dengan terapi. Jika Anda dan orang lain yang bekerja di klinik Anda percaya bahwa survei alamat kepuasan pasien, kita dapat mempertimbangkan survei untuk memiliki wajah validitas. Anda bisa, bagaimanapun, meningkatkan keabsahannya wajah dengan meminta saat ini dan mantan pasien jika mereka percaya bahwa kuesioner cukup mengeksplorasi kepuasan pasien.

## Criterion validity

Kriteria validitas menetapkan hasil pengukuran validitas dengan membandingkan ke yang lain, lebih mapan meyakinkan. Ukuran yang digunakan untuk perbandingan, kriteria pengukuran saja, dapat ditetapkan sebagai standar emas atau standar referensi. Ukuran kriteria standar emas memiliki validitas hampir tak terbantahkan untuk mengukur karakteristik yang menarik. Sebagai contoh, visualisasi persendian lutut akan dianggap sebagai kriteria standar emas untuk penilaian untuk ACL tear. Demikian juga, Berg balance telah dipelajari jadi secara ekstensif sebagai ukuran keseimbangan dan ebagai kriteria standar emas.

Ketika kriteria gold standard tidak ada untuk karakteristik tertentu, kriteria standar referensi digunakan. Standar referensi kurang tak terbantahkan dari gold standard tetapi dianggap wajar perbandingan untuk pengukuran hasil.

## Construct Validity

Construct validity adalah kemampuan hasil pengukuran untuk menilai karakteristik abstrak atau konsep. Ini adalah bentuk paling kompleks validitas untuk membangun. Untuk menetapkan validitas, hasil ukuran pengembang harus terlebih dahulu memberikan model teoritis yang menjelaskan konstruksi yang dinilai. Setelah

itu, serangkaian studi dilakukan untuk menetapkan apakah ukuran sebenarnya menilai konstruksi tersebut.

Tiga metode umum untuk menetapkan validitas yang dikenal-kelompok, konvergen, dan validitas diskriminan. Validitas dikenal-kelompok menetapkan bahwa hasil pengukuran menghasilkan nilai berbeda untuk kelompok dengan perbedaan-perbedaan yang dikenal pada ciri yang diukur. Misalnya, ukuran hasil yang didirikan untuk menilai stabilitas lutut akan diharapkan untuk menghasilkan nilai yang berbeda untuk orang-orang dengan dan tanpa cedera ACL.

Known-groups validity akan ditetapkan jika ukuran jelas digambarkan antara lutut dengan dan tanpa cedera ACL. Validitas konvergen menetapkan bahwa ukuran baru berkorelasi dengan pikiran untuk mengukur serupa lain karakteristik atau konsep. Sebaliknya, keabsahan diskriminan menetapkan bahwa ukuran tidak berkorelasi dengan ukuran berpikir untuk menilai karakteristik yang berbeda atau konsep. Misalnya, ukuran kekuatan fisik tidak akan diharapkan untuk memiliki korelasi tinggi dengan ukuran ketahanan fisik. Validitas diskriminan akan ditetapkan jika ukuran kekuatan fisik yang baru didirikan rendah korelasi dengan ukuran ketahanan fisik.

#### **Abstract 1**

**Objectives:** To determine the convergent validity of the 12-item Short-Form Health Survey, version 2 (SF-12v2), with 36-Item Short-Form Health Survey, version 2 (SF-36v2), in patients with spinal disorders, and to determine other key factors that might further explain the variances between the 2 surveys.

**Design:** Cross-sectional study.

**Setting:** Orthopedic ambulatory care.

**Participants:** Eligible participants (N=98; 24 with cervical, 74 with lumbosacral disorders) who were aged 18 years and older, scheduled to undergo spinal surgery, and completed the SF-36v2.

**Interventions:** Not applicable.

**Main Outcome Measures:** SF-36v2 and SF-12v2 (extracted from the SF-36v2).

**Results:** The 2 summary scores, physical and mental component scores (r range, .88-.97), and most of the scale scores (r range, .81-.99) correlated strongly between the SF-12v2 and SF36-v2, except for the general health score (cervical group,  $r = .69$ ; lumbosacral group,  $r = .76$ ). Stepwise linear regression analyses showed the SF-12v2 general health scores (cervical:  $\beta = .61$ ,  $P < .001$ ; lumbosacral:  $\beta = .68$ ,  $P < .001$ ) and the level of comorbidities (cervical:  $\beta = -.37$ ,  $P = .014$ ; lumbosacral:  $\beta = -.18$ ,  $P = .039$ ) were significant predictors of the SF-36v2 general health score in both groups, whereas age ( $\beta = .32$ ,  $P < .001$ ) and smoking history ( $\beta = -.22$ ,  $P = .005$ ) were additional predictors in the lumbosacral group.

**Conclusions:** SF-12v2 is a practical and valid alternative for the SF-36v2 in measuring health of patients with cervical or lumbosacral spinal disorders. The validity of the SF-12v2 general health score interpretation is further improved when the level of comorbidities, age, and smoking history are taken into consideration.

**Type of Validity Studied and Why:** Abstract 1 describes a study that measures construct validity. The authors report that convergent validity was assessed by determining if the new test (SF-12v2) measures the same construct as the established test (SF-36v2).

Lee CE, Browell LM, Jones DL. Measuring health in patients with cervical and lumbosacral spinal disorders: is the 12-item short-form health survey a valid alternative for the 36-item short-form health survey? *Arch Phys Med Rehabil.* 2008 May;89(5):829-33.

## Penilaian penerapan studi hasil pengukuran validitas outcome

Bagian A: menentukan penerapan validitas studi ini termasuk pertanyaan yang sama dijelaskan sebelumnya studi reliability.

Bagian B: menentukan kualitas studi validitas pertanyaan yang paling penting untuk bertanya tentang kualitas tergantung pada jenis berlaku di bawah penyelidikan dan metode yang digunakan.

Bagian C: Interpretasi hasil validitas studi satu metode statistik yang paling umum digunakan dalam studi hasil mengukur validitas adalah rho correlation (r).

#### Bagian D: Ringkasan klinis Bottom Line

Pada akhirnya, garis bawah dari studi tentang reliability yang memberitahu apakah ukuran ini cukup berlaku untuk digunakan dalam praktek klinis. Dengan kata lain, studi mengatakan kepada kita Apakah kita harus memiliki keyakinan bahwa ukuran sebenarnya mengukur apa bertujuan untuk mengukur. Sebelum kita dapat mendasarkan penafsiran kita tentang status pasien atau kemajuan pada hasil mengukur hasil, kita perlu memiliki keyakinan dalam ukuran itu sendiri.

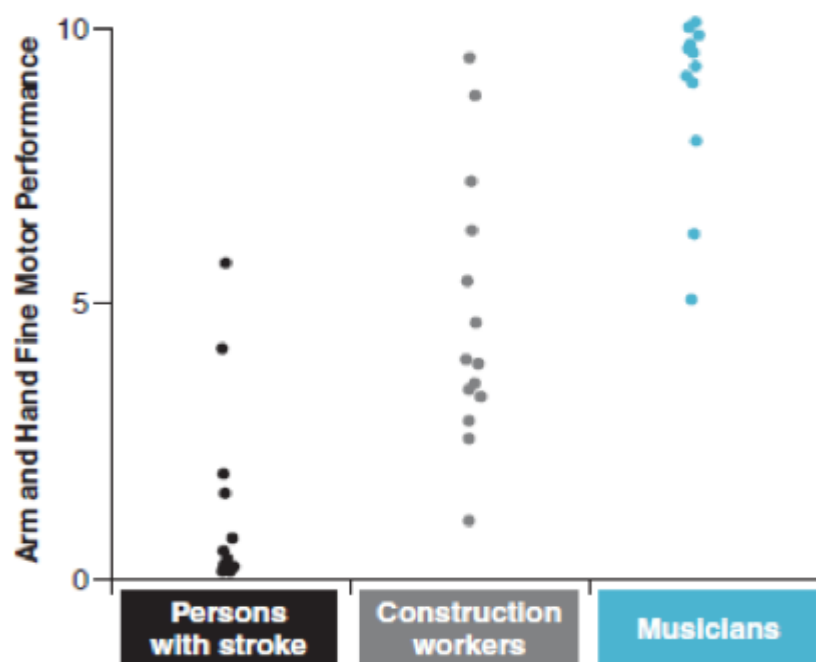
#### Studi Asesmen Pengukuran Kemaknaan Klinis Pengukuran Outcome

Studi yang menilai mengukur hasil klinis kebermaknaan jenis dari kebermaknaan hasil klinis langkah-langkah harus sah dan handal dan membantu kami dalam menafsirkan perubahan pasien kami memiliki makna klinis. Dalam bagian ini, kami menjelaskan konsep yang terkait dengan klinis kebermaknaan, termasuk floor and ceiling effect, minimal detectible change, responsiveness, dan minimal clinically important difference

#### Floor and Ceiling Effects

Floor and ceiling effects mencerminkan kurangnya cukup kisaran dalam ukuran untuk sepenuhnya ciri kelompok pasien. Sebagai contoh, mempertimbangkan ukuran hasil yang dirancang untuk mengukur dewasa kontrol motorik halus. Mungkin skor pada ini hipotetis ukuran berkisar dari 0 hingga 10, dengan nilai yang lebih tinggi menunjukkan kinerja yang lebih baik. Misalnya pada gambaran tiga kelompok berbeda:

1. Pasien stroke
2. Pekerja konstruksi
3. Musisi



Gambar 3. Floor and Ceiling Effects

Nilai pada pasien stroke tersusun di bagian bawah skala, dengan banyak mendapatkan skor minimum (0). Ini adalah contoh dari floor effect. Partitur untuk musisi berkumpul di bagian atas skala, dengan banyak mendapatkan skor maksimum (10). Ini adalah contoh ceiling effect. Nilai untuk pekerja konstruksi yang tersebar di berbagai, dan sangat sedikit mendapatkan maksimum (10) atau Skor minimum (0). Untuk populasi pekerja konstruksi, ukuran ini tidak memiliki Floor and ceiling effects. Oleh karena itu, ukuran sesuai untuk digunakan di kalangan pekerja konstruksi tetapi terbatas berarti untuk orang dengan stroke dan musisi.

Itu tidak biasa untuk melihat sebuah studi yang khusus dirancang untuk mempelajari floor and ceiling effects, tetapi itu umum untuk menghadapi studi yang menemukan efek ini sebagai bagian dari sebuah studi besar. Periksa untuk floor and ceiling effects ketika menilai studi oleh visual memeriksa distribusi nilai peserta pada hasil ukuran. Kegunaan dari ukuran berkurang dengan adanya floor and ceiling effects.

### Minimal Detectible Change

Minimal detectible change (MDC) adalah jumlah minimum perubahan yang diperlukan pada hasil pengukuran untuk mengantisipasi kesalahan dan variabilitas. Sebagai contoh, pengukuran ketahanan berjalan pasien menggunakan Six-Minute Walk Test setiap hari selama 5 hari, kita akan mengharapkan beberapa variabilitas dalam jarak yang pasien berjalan setiap hari (misalnya, hari 1 = 149 meter (m), hari 2 = 160 m, hari 3 = 155 m, hari 4 = 149 m, hari 5 = 152 m). Anda perlu tahu MDC untuk Six-Minute Walk Test dalam menentukan seberapa banyak perubahan yang diperlukan untuk melebihi variasi alami ini dan mewakili perubahan status pasien.

Minimal detectible change berasal dari penggunaan pengukuran outcome menggunakan ukuran hasil test-retest reliability dan dalam subjek deviasi standar untuk pengukuran yang berulang-ulang. Dasarnya untuk itu adalah studi test-retest reliability berkualitas tinggi. Karena test-retest reliability dan variabilitas yang berbeda untuk populasi pasien berbeda, MDC untuk mengukur hasil harus ditentukan untuk populasi yang berbeda. Sebagai contoh, MDC Six-Minute Walk Test telah dilaporkan 36 m untuk orang dengan penyakit Alzheimer, 1927 m untuk orang dengan penyakit Parkinson, 20 dan 53 m untuk orang dengan penyakit paru obstruktif kronik (PPOK).

Oleh karena itu, jika seorang pasien dengan penyakit Parkinson mengalami peningkatan 40-m, perubahan akan dianggap luar variabilitas alami (27 m) untuk penduduk. Sebaliknya, jika seorang pasien PPOK mengalami peningkatan 40-m sama, perubahan gagal melebihi MDC (53 m) untuk penduduk ini dan akan dianggap variasi alami daripada perubahan seharusnya.

### Responsiveness and Minimal Clinically Important Difference

Responsiveness mencerminkan kemampuan pengukuran outcome untuk mendeteksi perubahan dari waktu ke waktu. Jika terjadi peningkatan pasien dalam karakteristik tertentu, kita ingin tahu bahwa menggunakan ukuran akan mendeteksi, atau menanggapi, perubahan itu. Studi responsiveness bersifat longitudinal dan mengukur perubahan yang diharapkan pada subjek terkait dengan perubahan waktu. Salah satu ukuran responsif adalah minimal clinically important difference (MCID), juga disebut sebagai minimal important difference (MID). MCID mewakili jumlah minimum perubahan pada ukuran hasil yang pasien menganggap sebagai menguntungkan dan yang akan membenarkan perubahan. MCID akan menjawab

pertanyaan, "berapa banyak perubahan pada ukuran hasil yang cukup untuk menjadi bermakna?"

Studi MCID membandingkan gold standard seperti pada studi validitas. Pada studi MCID, gold standard adalah ukuran dikenal untuk mendeteksi perubahan bermakna, bila mungkin, dari sudut pandang pasien. Metode statistik yang digunakan untuk menentukan apa yang berubah pada ukuran hasil menarik terbaik berkorelasi dengan perspektif pasien. Kadang-kadang perspektif pasien tidak dapat dikumpulkan untuk studi MCID. Dalam kasus ini, sangat tepat, tapi kurang ideal, menggunakan perbandingan gold standard yang didasarkan pada sudut pandang fisioterapis, anggota keluarga atau pengasuh. Studi MCID perlu dipertimbangkan dalam konteks. Penerapan merupakan komponen penting dari penilaian untuk studi pelaporan MCID.

Untuk ukuran tertentu hasil, MCID khusus untuk populasi pasien tertentu, dengan mempertimbangkan, misalnya, diagnosis, usia, atau tingkat keparahan kondisi. Akhirnya, ketika menafsirkan studi MCID penting untuk mempertimbangkan MDC. MDC berdasarkan antisipasi kesalahan dan variabilitas, dan MCID yang lebih kecil daripada MDC penggunaannya lebih sedikit. Sebagai contoh, MCID Six-Minute Walk Test telah dilaporkan sebagai 54 m untuk orang-orang dengan orang COPD. Sangat penting bahwa nilai ini (54 m) menjadi lebih besar dari nilai MDC 53 m berasal dari pekerjaan oleh Brooks et al. MCID jika tidak lebih besar dari MDC, variabilitas alami ukuran (53 m) akan mengaburkan kemampuan untuk mendeteksi perubahan secara bermakna (54 m). Perbedaan signifikan secara statistik berbeda dari konsep MCID. Signifikansi statistik dipengaruhi oleh ukuran dan variabilitas (misalnya, standar deviasi) sampel dibandingkan.

Penilaian studi asesmen kemaknaan klinis pengukuran outcome

Bagian A: menentukan penerapan klinis kebermaknaan studi penerapan untuk studi klinis kebermaknaan memerlukan kesamaan kuat antara populasi pasien yang diteliti dan pasien kepada siapa kita ingin menerapkan hasil. Kemaknaan didefinisikan dari sudut pandang pasien. Dan pasien akan memiliki kemiripan perspektif.

Bagian B: menentukan kualitas klinis kebermaknaan belajar karena desain studi (selain analisis statistik) adalah sama untuk studi test-retest reliability dan studi MDC, pertanyaan untuk studi reliability.

Bagian C: Menafsirkan hasil studi kemaknaan klinis hasil dari studi yang mencakup penilaian dari ceiling and floor effects akan menyediakan proporsi (persentase) dari orang-orang yang mengalami floor effects (skor terendah) dan ceiling effects (nilai tertinggi). Studi yang menilai MDC akan melaporkan perubahan nilai yang mewakili MDC dan juga harus mencakup interval keyakinan untuk MDC. Sebagai contoh, sebuah studi mungkin laporan bahwa MDC pada ukuran hasil yang adalah 4 dari 100 poin dengan interval keyakinan 95% dalam 2 hingga 7. Ini akan memberitahu Anda bahwa perubahan hanya 4 atau lebih besar pada ukuran hasil (untuk pasien tertentu) harus dipertimbangkan perubahan sejati. Mengubah kurang dari 4 akan dianggap dalam variasi alami ukuran. Interval keyakinan mengatakan kepada kita bahwa kita adalah 95% yakin bahwa nilai MDC benar untuk populasi umum terletak di antara 2 dan 7.

Studi MCID juga akan melaporkan perubahan nilai untuk pengukuran hasil. Berbeda dengan MDC, nilai perubahan ini mewakili jumlah minimum perubahan pada ukuran hasil yang mungkin mewakili perubahan yang berarti bagi pasien. Interpretasi dari

nilai MCID harus menyertakan pertimbangan gold standard yang digunakan untuk menentukan perubahan yang berarti dalam populasi. Hasil MCID harus mencakup interval keyakinan

Bagian D: Ringkasan clinical bottom line studi kemaknaan klinis

Clinical bottom line berkaitan dengan apakah ukuran tertentu hasil memberikan informasi yang berguna dalam lingkungan klinis. Terkait ceiling and floor effects, kita melihat rentang untuk menangkap perubahan pada pasien berdasarkan ciri-ciri yang diukur. Studi MDC menginformasikan kita kapan harus mempertimbangkan perubahan pada ukuran hasil yang melampaui variasi alami. Mungkin paling berguna dari kategori ini sifat psikometrik, studi MCID dapat membantu kita menilai apakah terapi yang mengakibatkan perubahan yang berarti bagi pasien kami. Nilai-nilai MCID dapat digunakan untuk menginformasikan pasien tujuan dan untuk membuktikan nilai intervensi untuk semua stakeholder.

Ringkasan

Pengukuran outcome memungkinkan fisioterapis fisik mengukur dan menganalisa perubahan yang pasien mengalami selama penanganan. Outcome pengukuran juga alat utama yang digunakan dalam penelitian untuk mengukur perubahan dalam peserta dari waktu ke waktu. Oleh karena itu, kemampuan untuk secara akurat menilai kualitas mengukur hasil adalah keterampilan dasar untuk terapis berbasis bukti. Sifat psikometrik digunakan untuk menggambarkan kualitas dari mengukur hasil. Dalam bab ini, kami menggambarkan jenis studi yang paling umum digunakan untuk menilai sifat psikometrik hasil tindakan. Keandalan studi menilai konsistensi mengukur hasil dan internal consistency and test-retest, intra-rater, and inter-rater reliability

Studi validitas memberikan kita wawasan tentang apakah ukuran hasil yang benar-benar mengukur apa dimaksudkan untuk diukur. Validity dapat dibagi content, criterion, and context validity. Akhirnya, mengukur hasil hanya dapat diterapkan dalam praktek jika kita memahami bagaimana menafsirkan kebermaknaan klinis. Studi klinis kebermaknaan menginformasikan kepada kami tentang keberadaan ceiling and floor effect, kemungkinan bahwa ukuran hasil yang akan berubah saat pasien mengalami perubahan (respon), perubahan apa pada ukuran mewakili perubahan seharusnya (MDC), dan jumlah perubahan yang akan dianggap bermakna untuk pasien (MCID).

Daftar Pustaka

Linda Fetters, Julie Tilson, Evidence Based Physical Therapy, F.A Davis Company, Philadelphia, 2012), pp 125-144