

## Uji Inter-Rater Reliability Western Ontario and McMaster University (WOMAC) Osteoarthritis Index pada Pasien Osteoarthritis Knee

Suryo Saputra Perdana<sup>1</sup>, Amaliyah Hana Safitri<sup>2</sup>, Nabila<sup>3</sup>, Nur Agung Martopo<sup>4</sup>

<sup>1,2,3,4</sup> Program Studi Fisioterapi, Fakultas Ilmu Kesehatan, Universitas Muhammadiyah Surakarta, Jawa Tengah, Indonesia

Jl. Ahmad Yani, Kartasura, Surakarta 57169, Jawa Tengah, Indonesia

Email: <sup>1</sup>ssp741@ums.ac.id ; <sup>2</sup>amlyhsafi@gmail.com ; <sup>3</sup>aabilafajrihasan@gmail.com;

<sup>4</sup>martopoagung@gmail.com

Tanggal Submisi: 5 Juni 2020; Tanggal Penerimaan: 8 Desember 2020

### ABSTRAK

Penderita Osteoarthritis (OA) *Knee* tercatat sebanyak 5% dari total seluruh penderita OA knee di dunia sehingga diperlukan alat ukur yang relevan dan *reliable* untuk mengukur *activities daily living* (ADL) dan *body function* pada penderita OA. Selain menguji apakah alat ukur tersebut valid dan reliabel, diperlukan uji inter-rater reliability dari Indeks The Western Ontario and McMaster University Osteoarthritis Index (WOMAC) untuk digunakan di Indonesia. Penelitian ini adalah *psychometric study* dengan pendekatan cross-sectional untuk menguji inter-rater reliability pada instrumen pengukuran WOMAC yang diujikan kepada *fresh graduated* fisioterapi dengan seorang pasien osteoarthritis dihari yang sama. Analisis item-per-item dilakukan kepada rater sebelum dilakukannya pemeriksaan kepada pasien. Reliabilitas *inter-rater* diukur menggunakan fleiss kappa ( $\kappa$ ). Indek WOMAC yang terdiri dari 5 pertanyaan yang menanyakan keluhan nyeri, 2 pertanyaan tentang kekakuan dan 17 pertanyaan tentang kesulitan dalam melakukan aktifitas keseharian. Hasil uji statistic mengungkapkan bahwa uji *inter-rater reliability* untuk WOMAC adalah *fair agreement*, yang berarti klasifikasi kesepakatan antar rater dikatakan belum cukup adekuat untuk digunakan pada *fresh graduated* fisioterapi.

**Kata Kunci:** Inter-rater, Osteoarthritis Knee, Reliability, WOMAC

### ABSTRACT

In Indonesia, as much as 5% of the entire population suffering from knee osteoarthritis, relevant measurement tools are needed to measure daily living activities (ADL) and body functions in patients with OA. Besides being able to be used as a valid and reliable measurement tool, an inter-appraisal reliability test from the The Western Ontario and McMaster University Osteoarthritis Index (WOMAC) is needed for use in Indonesia. In this study, this type of psychometric study uses an inter-rater reliability test on the WOMAC assessment instrument to be tested for physiotherapy with osteoarthritis patients on the same day. Item-by-item analysis is performed to the rater before requesting acceptance from the patient. Inter-rater reliability uses kappa fleiss ( $\kappa$ ). The WOMAC index consists of 5 questions that raise complaints, 2 questions about rigidity and 17 questions about difficulties in carrying out daily activities. Statistically result reveal that inter-rater reliability of WOMAC index is inadequate for implementing in fresh graduated phsyiotherapists.

**Keywords:** Knee Osteoarthritis, WOMAC, Inter-rater, Reliability.

ISSN 1979-7621 (Print). ISSN 2620-7761 (Online)

DOI : 10.23917/jk.v13i2.11043

## PENDAHULUAN

Kekakuan pada pagi hari dan penurunan kinerja fisik men jadi keluhan utama yang dirasakan oleh pasien OA lutut. Ditambah lagi kurangnya mobilitas sendi, kekuatan otot, ketidakstabilan sendi, dan krepitasi menjadi faktor lain dalam penyebab terjadinya OA (Adhiputra, 2017). Di Indonesia, prevalensi terjadinya OA pada usia 61 tahun adalah sebanyak 5% (15.5% pada pria dan 12.7% pada wanita) dari keseluruhan total populasi penduduk sebanyak 225 juta (Koentjoro, S.L., 2010). Karena sifatnya yang progresif dan kronis menyebabkan sakit parah dan cacat pada pasien sehingga mengganggu kegiatan sehari-hari. Perhitungan mencatat 80% penderita OA memiliki keterbatasan dalam bergerak dan 25% diantaranya kesulitan dalam melakukan kegiatan sehari-hari secara mandiri (WHO, 2016).

Untuk mengukur *activities daily living* (ADL) dan *body function* dibutuhkan alat ukur yang relevan. Menurut: *Koninklijk Nederlands Genootschap voor Fysiotherapie* (KNGF) *guidline*, *Western Ontario and McMaster Universities osteoarthritis index* (WOMAC), *the Knee injury and Osteoarthritis Outcome Score* (KOOS), dan *the Algofunctional Index* (AFI) adalah alat ukur yang relevan untuk mengukur ADL dan *body function* (Peters, 2010). WOMAC adalah instrumen penilaian berupa kuesioner yang telah digunakan oleh tenaga kesehatan profesional untuk mengevaluasi kondisi pasien OA lutut dan pinggul yang berfokus pada penilaian nyeri, kekakuan, *activity limitations* pasien OA lutut maupun pinggul (Salaffi *et al.*, 2003). Komponen utama yang dibutuhkan dalam sebuah alat ukur yang direkomendasikan dalam praktek klinis adalah validitas dan reliabilitas (Mchugh, 2012). WOMAC mempunyai validitas dan reliabilitas yang baik dan penulis memandang bahwa WOMAC merupakan instrumen penilaian yang perlu diaplikasikan di Indonesia dalam

setiap pemeriksaan OA knee. Namun dibalik valid dan reliabilitas yang didapat dalam setiap penelitian WOMAC, penulis tidak menemukan adanya pembahasan *inter-rater* dalam penelitian yang dilakukan sebelumnya. Sehingga penulis merasa perlu untuk menjadikan ini topik yang harus diteliti.

## METODE PENELITIAN

Penelitian ini menggunakan rancangan penelitian *psychometric study* dengan metode pendekatan *cross-sectional* (*non-experimental* dengan pengulangan pengukuran) untuk mengukur *inter-rater reliability* yang bertujuan untuk merefleksikan variasi antara dua atau lebih rater (penilai) dengan melakukan pengukuran pada subjek yang sama (Gwet, 2014).

Teknik analisa pada penelitian ini menggunakan metode *interrater reliability* dengan pengujian menggunakan statistik *kappa* ( $k$ ) yang terdiri dari; *Cohen kappa* yang menyatakan konsistensi pengukuran yang dilakukan oleh dua orang penilai (*rater*) dan *Fleiss kappa* yang menyatakan konsistensi pengukuran yang dilakukan oleh lebih dari dua penilai (*rater*). Hal ini dapat diartikan sebagai instrumen yang menyatakan sejauh mana jumlah perjanjian yang diamati antara penilai melebihi apa yang diharapkan. Dan kemudian diinterpretasikan dalam *kappa* digunakan untuk menentukan apakah sebuah *outcome measurement* telah layak untuk digunakan dalam praktik klinis (Scott, 1955).

**Tabel 1. Interpretation Kappa**

$k$	<i>Interpretation</i>
< 0	<i>Poor agreement</i>
0.01 – 0.20	<i>Slight agreement</i>
0.21 – 0.40	<i>Fair agreement</i>
0.41 – 0.60	<i>Moderate agreement</i>
0.61 – 0.80	<i>Substantial agreement</i>
0.81 – 1.00	<i>Almost perfect agreement</i>

**Tabel 2. Intraclass Correlation Coefficients (ICC)**

Score	Interpretation
<0,5	Poor reliability
0,5-0,75	Moderate reliability
0,75-0,90	Good reliability
>0,90	Excelent reliability

Sebuah *outcome measurement* belum dapat dikatakan *reliable* apabila jumlah ICC yang didapat >0,75 (kurang dari 0,75) atau masuk dalam kategori *Level Good Reliability* (Gwet, 2014).

## HASIL DAN PEMBAHASAN

Responden berjumlah 13 orang yang merupakan mahasiswa transfer S1 Fisioterapi Universitas Muhammadiyah Surakarta yang pernah memiliki pengalaman praktik selama 6 bulan dan telah memiliki STR. Usia berkisar antara 21-23 dan dari seluruh responden yang bersedia, memiliki

aktifitas kuliah dan tidak bekerja pada klinik ataupun instansi Sebanyak 13 orang responden (assessor) yang akan menjadi memenuhi kriteria inklusi xxxx (setelah itu dijelaskan karakteristik responden dan dijelaskan juga 2 jenis responden (assessor dan OA patients) kesehata dan 30% diantaranya tidak terbiasa dalam menggunakan bahasa inggris dalam kesehariannya.

Subjek penelitian merupakan ibu rumah tangga yang berumur 57 tahun yang menderita osteoarthritis knee berdasarkan diagnosis klinis berupa usia lebih dari 45 tahun, pain related activities, dan *morning stiffness* selama kurang dari 30 menit (Tate, 2010). Subjek penelitian terlihat kooperatif, yaitu mampu memahami penjelasan dan instruksi pemeriksaan yang dilakukan oleh peneliti serta mampu berkomunikasi dengan baik kepada peneliti maupun kepada responden penelitian (assessor).

**Tabel 3. Over All Kappa**

Kappa	Asymptotic Standard Error	Z	P Value	Lower 95% asymptotic CI Bound	Upper 95% asymptotic CI Bound
Overall 0,396	0,015	25,794	0,000	0,366	0,426

Dari data diatas dapat dilihat hasil interrater kappa menunjukkan nilai 0,396. Kappa Fleiss dapat berkisar dari -1 hingga +1. Nilai negatif kappa ( $\kappa$ ) menunjukkan bahwa kesepakatan antara dua atau lebih penilai kurang dari perjanjian yang diharapkan, dengan -1 menunjukkan bahwa tidak ada perjanjian yang diamati, dengan artian penilai tidak menyetujui apa pun. Namun nilai negatif sangat jarang terjadi (Woalder, 2017). Nilai 1 menunjukkan hasil maksimal dari kesepakatan, yang menunjukkan perjanjian sempurna (Fleiss, 1971).

Skala klasifikasi ini dengan nilai Fleiss kappa ( $\kappa$ ) 0,396, dapat mewakili kekuatan perjanjian yang *fair agreement*. Namun, nilai kappa sangat tergantung pada distribusi marjinal, yang digunakan untuk

menghitung tingkat porposi dari *inter-rater* (Gwet, 2014). Dengan demikian, nilai kappa akan berbeda tergantung pada distribusi marjinal dan hal ini menjadi salah satu kelemahan terbesar dari Fleiss Kappa. Hal ini menunjukkan bahwa tidak dapat membandingkan satu Fleiss kappa dengan yang lain kecuali distribusi marjinalnya sama (González Alonso & Pazmiño Santacruz, 2015).

*Convidence interval* (CI) 95% untuk Fleiss Kappa digunakan dalam penelitian ini. Akan terdapat kolom "Lower 95% asymptotic CI Bound " dan " Upper 95% asymptotic CI Bound", dengan hasil 0,366 hingga 0,426. Untuk meyakinkan bahwa nilai populasi sebenarnya dari fleiss kappa adalah antara 0,366 dan 0,426.

Hasil analisis didapatkan bahwa bahwa nilai-p adalah 0,000 ( $p < 0.05$ ).  $p < 0.05$  bermakna hasil yang signifikan secara statistik dan koefisien Fleiss Kappa yang didapatkan secara statistik berbeda secara signifikan dari 0 (nol). Jika  $p > 0,05$  menunjukkan bahwa tidak adanya hasil yang

signifikan secara statistik dari koefisien Fleiss Kappa (Glass & Di Eugenio, 2004).

Oleh karena itu, dapat disimpulkan bahwa ada kesepakatan *fair agreement* antara penilaian pasien OA terhadap instrumen penelitian WOMAC, dengan nilai  $\kappa = 0,396$  (95% CI, antara 0,366 dan 0,426),  $p < 0,0005$ .

**Tabel 4. Kappa for Individual Categories**

Rating Category	Conditional Probability	Kappa	Asymtotic Standard Error	Z	p-value	Lower 95% asymptotic CI Bound	Upper 95% asymptotic CI Bound
0	0,823	0,700	0,700	0,023	0,000	0,654	0,746
1	0,537	0,292	0,292	0,023	0,000	0,247	0,337
2	0,347	0,192	0,192	0,023	0,000	0,146	0,237
3	0,067	0,020	0,020	0,023	0,398	-0,026	0,065
4	0,000	-0,003	-0,003	0,023	0,889	-0,049	0,042

**Tabel 5. Intraclass Correlation Coefficient**

	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	0,531	0,387	0,700	17,214	23	276	0,000
Average Measures	0,936	0,891	0,968	17,214	23	276	0,000

Hasil dari ICC menunjukkan bahwa *inter-rater* dalam *outcome measurement* WOMAC sebesar 0,531 (*Moderate Reliability*). Dengan kata lain WOMAC belum dapat dikatakan *inter-rater* reliabel untuk digunakan dalam praktik klinis.

## KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan mengenai uji *inter-rater* reliability WOMAC pada individu dengan osteoarthritis *knee* di Indonesia dapat disimpulkan sebagai berikut:

Berdasarkan uji statistik fleiss kappa yang digunakan menunjukkan hasil bahwa uji *inter-rater* reliability untuk WOMAC adalah *fair agreement*, yang berarti klasifikasi kesepakatan antar rater dikatakan belum cukup. Pada penelitian ini memiliki hasil *inter-rater* yang beragam pada fisioterapi *fresh graduade*. Nilai kappa

0,396 yang menunjukkan *fair agreement* dan ICC *moderate reliability* dengan nilai 0,531. Hal ini membuktikan bahwa instrumen penilaian WOMAC terbukti *reliable* namun dengan *intrepetition fair agreement* yang berarti diperlukan lagi kesepakatan terhadap pemeriksaan OA dengan penggunaan WOMAC.

WOMAC merupakan instrumen pengukuran OA dengan pemeriksaan yang lengkap untuk dilakukan kepada penderita OA lansia dengan aktifitas rendah hingga menengah. Berdasarkan hasil penelitian diketahui bahwa WOMAC belum tepat untuk diimplementasikan kepada penderita OA *knee* di Indonesia, ada beberapa faktor diantaranya: hasil uji *inter-rater* reliability dan ICC masuk kedalam kategori *fair agreement* dan *moderate*. WOMAC masih berbentuk form pemeriksaan dalam bahasa inggris dan belum disesuaikan dengan budaya dan gaya hidup masyarakat Indonesia.

Bagi fisioterapi *fresh graduate* dan yang belum memiliki pengalaman diperlukan pemahaman yang cukup baik untuk menerjemahkan WOMAC dengan bahasa yang mudah dipahami pasien.

Saran untuk penelitian selanjutnya adalah agar dapat melakukan penelitian *cross-clinical* dalam skala pengukuran

WOMAC yang relevan secara budaya dan gaya hidup masyarakat Indonesia. Sehingga saran untuk penelitian selanjutnya diperlukan adanya *cross-clinical* WOMAC versi Indonesia untuk menyesuaikan isi pemeriksaan dengan budaya yang ada dan gaya hidup masyarakat di Indonesia.

## DAFTAR PUSTAKA

- Fleiss, J. L. (1971). *Nominal Scale Among Many Rater*. 76(5), 378–382.
- Glass, M., & Di Eugenio, B. (2004). The Kappa statistic : a second look. *Computational Linguistics*, 30(1), 95–101.
- González Alonso, J., & Pazmiño Santacruz, M. (2015). Cálculo e interpretación del Alfa de Cronbach para el caso de validación de la consistencia interna de un cuestionario, con dos posibles escalas tipo Likert. *Revista Publicando*, 27(1), 62–67.
- Gwet, K. (2014). *Handbook of inter-rater reliability fourth edition*.
- Mchugh, M. L. (2012). Lessons in biostatistics Interrater reliability : the kappa statistic. *Biochemia Medica*, 22(3), 276–282.
- Peters. (2010). Royal Dutch society for physical therapy KNGF guideline osteoarthritis hip-knee. *Supplement to the Dutch Journal of Physical Therapy*, 120(1). Retrieved from [http://www.ipts.org.il/\\_Uploads/dbsAttachedFiles/osteoarthritis\\_of\\_the\\_hip\\_and\\_knee\\_practice\\_guidelines\\_2010.pdf](http://www.ipts.org.il/_Uploads/dbsAttachedFiles/osteoarthritis_of_the_hip_and_knee_practice_guidelines_2010.pdf)
- Salaffi, F., Leardini, G., Canesi, B., Mannoni, A., Fioravanti, A., Caporali, R., ... Padova, U. (2003). *Reliabilitas dan validitas dari Western Ontario dan Universitas McMaster ( WOMAC ) Osteoarthritis Index pada pasien Italia dengan osteoarthritis lutut*. 4584(3), 551–560.
- Scott, W. A. (1955). Reliability of Content Analysis: The Case of Nominal Scale Coding. *Public Opinion Quarterly*, 19(3), 321. <https://doi.org/10.1086/266577>
- Tate, K. (2010). Guideline for the management of knee and hip osteoarthritis Second edition. In *Architectural Digest* (Vol. 67).
- Woalder. (2017). 乳鼠心肌提取 HHS Public Access. In *Physiology & behavior* (Vol. 176). <https://doi.org/10.1016/j.physbeh.2017.03.040>
- Xhaferi, E., Prof, A., Backa, T., Imami, A., & Pano, I. (2017). *Osteoarthritis risk factors and clinical features*. (April), 12–17.