

Komentarz do artykułu A. Boszczyka i wsp. „Tłumaczenie, adaptacja kulturowa i walidacja polskojęzycznej wersji kwestionariusza Foot and Ankle Outcomes Questionnaire”

Comment on Boszczyk et al. “Translation, Cultural Adaptation and Validation of Polish Version of Foot and Ankle Outcomes Questionnaire”

Przemysław Tomasz Paradowski^{1,2}, Dariusz Witoński³

¹ Klinika Ortopedii, Szpital Sunderby, Luleå, Szwecja,

² Klinika Ortopedii i Traumatologii, Uniwersytet Medyczny w Łodzi, Uniwersytecki Szpital Kliniczny im. WAM, Łódź, Polska

³ Społeczna Akademia Nauk, Łódź, Polska

¹ Department of Orthopaedics, Sunderby Central Hospital of Norrbotten, Luleå, Sweden

² Department of Orthopaedics and Traumatology, Medical University of Łódź, WAM University Hospital, Łódź, Poland

³ Social Academy of Science, Łódź, Poland

Szanowny Panie Redaktorze,

Z zainteresowaniem przeczytaliśmy artykuł Andrzeja Boszczyka i wsp. „Tłumaczenie, adaptacja kulturowa i walidacja polskojęzycznej wersji kwestionariusza Foot and Ankle Outcomes Questionnaire” [1]. Autorzy przeprowadzili jako pierwsi w naszym kraju proces adaptacji kulturowej i walidacji wersji polskiej skali oceny stopy i stawu skokowo-goleniowego *Foot and Ankle Outcomes Questionnaire* (FAOQ).

W zgodnej opinii większości autorów analiza przebiegu choroby i wyników leczenia powinna być przeprowadzona z użyciem instrumentów pomiarowych przeznaczonych do oceny subiektywnej prowadzonej z punktu widzenia chorego (ang. *patient-reported outcome* bądź też *patient-relevant outcome*, PRO) [2]. Poza wszelkimi wątpliwościami pozostaje również celowość poddania PROs adaptacji lingwistycznej i kulturowej oraz walidacji w języku, którym posługują się badani chorzy. PRO powinien być standaryzowany, czuły na zmiany kliniczne w czasie, zwięzły i łatwy w użyciu zarówno dla ocenianego, jak i dla oceniającego [3].

Adaptacji językowej i kulturowej doczekały się w Polsce wcześniej jedynie skala oceny kończyny górnej *Disabilities of the Arm, Shoulder and Hand* (DASH) [4] oraz skale oceny stawu kolanowego *International Knee Documentation Committee* (IKDC) [5] i *Knee injury and Osteoarthritis Outcome Score*

Dear Editor,

Recently, we have read with deep interest the article by Boszczyk et al. [1] entitled “Translation, Cultural Adaptation and Validation of Polish Version of Foot and Ankle Outcomes Questionnaire”. The authors were the first to linguistically and cross-culturally translate and validate the Polish version of the *Foot and Ankle Outcomes Questionnaire* (FAOQ).

It has already been well-established that both the course of a disease and the treatment outcome can be better described by patients themselves than on the basis of a clinical examination made by a physician. Such an approach is possible with patient-reported outcomes (PROs) [2]. PROs should be standardized, sensitive to clinical change, concise, and convenient for both the patients evaluated and clinicians [3]. Questionnaires should first undergo a linguistic and cross-cultural translation processes and their psychometric properties should be assessed in clinical validation studies.

Up to date, only the *Disabilities of the Arm, Shoulder and Hand* (DASH) questionnaire [4] and two knee questionnaires: *International Knee Documentation Committee* (IKDC) [5] and *Knee injury and Osteoarthritis Outcome Score* (KOOS) [6,7] have been linguistically and culturally adapted in the Polish language for use in orthopaedic surgery. However, only the KOOS scale was internationally esta-

(KOOS) [6,7]. Pełny proces walidacyjny przeprowadzono jedynie w przypadku tego ostatniego instrumentu badawczego.

W opublikowanym przez dr. A. Boszczyka i wsp. artykule proces adaptacji językowej i kulturowej został przeprowadzony na podstawie ogólnie przyjętego protokołu upowszechnionego przez badaczkę z Uniwersytetu w Toronto w Kanadzie Dorcas Beaton [8]. Tę część zadania wykonano w sposób modelowy.

Drugi etap pracy, czyli proces walidacji skali FAOQ wymaga w naszej opinii komentarza. I tak:

1. Przeprowadzenie procesu walidacji odbywa się zgodnie z protokołem obejmującym ocenę poszczególnych własności instrumentu badawczego. Stosowane obecnie wytyczne, takie jak lista *Scientific Advisory Committee* (SAC) [9], kryteria zaproponowane przez Bombardier and Tugwella [10], kryteria Andersen i Meyersa [11] czy wreszcie *COnsensus-based Standards for the selection of health status Measurement INstruments* (COSMIN) [12] odnoszą się do informacji, które należy przedstawić przy ocenie rzetelności (powtarzalności) instrumentu badawczego (ang. *reliability, reproducibility*), jej precyzji (ang. *accuracy*), zdolności do spełniania założonej funkcji (ang. *validity*) czy też zdolności do uchwycenia zmian klinicznych zachodzących w czasie (ang. *responsiveness*). Nie jest jasne czy autorzy przeprowadzali walidację FAOQ zgodnie z wytycznymi COSMIN, czy też zastosowali jakąś inną metodę kontrolną. Informacja taka powinna znaleźć się w opisie metodyki badań.
2. Jednym z etapów walidacji PRO jest *walidacja konstruktów* (ang. *construct validity*), a nie *konstrukcji* skali, jak zostało błędnie przedstawione przez autorów artykułu. Pojęcie *konstruktów* zostało zaadoptowane z psychologii, z prac George'a Kelly'ego [13]. *Konstrukt* jest obiektem teoretycznym, nieistniejącym obiektywnie, jest sposobem postrzegania i interpretowania danego zjawiska z perspektywy własnej chorego. Do stworzenia *konstruktów* dochodzi w toku operacjonalizacji, czyli procesu definiowania miary zjawiska, które nie jest bezpośrednio mierzalne. *Construct validity* określa, czy kwestionariusz badania jest porównywalny z innymi, podobnymi badaniami w stopniu, jakiego się oczekuje. Pojęcie to zawiera przypuszczenie, jak silne są zależności pomiędzy różnymi metodami badawczymi. Zależności te bada się albo z użyciem instrumentu oceniającego ogólny stan zdrowia chorych (generycznego), albo skali swoistej (specyficznej) dla ocianego stawu czy kończyny. W założeniu domeny (subskale) oceniające podobne zjawiska po-

bliszed as well as formally cross-culturally adapted and validated in Poland.

In the article, the linguistic and cultural adaptation process followed the recommendations by Dorcas Beaton et al. from the University of Toronto, Canada [8]. This part of the paper is well written. The validation process, however, needs to be discussed. We would like to point out certain issues which we believe will contribute to the effectiveness of the study.

1. Since there is a consensus that measurement instruments must have adequate measurement properties, it has been suggested that standardized guidelines for assessing the quality of these instruments should be followed. Available standards, including the Scientific Advisory Committee (SAC) list [9], the criteria proposed by Bombardier and Tugwell [10], the protocol described by Andersen and Meyers [11] and, last but not least, the *COnsensus-based Standards for the selection of health status Measurement INstruments* (COSMIN) [12] concern information that authors should take into account referring to such issues as reliability, validity, accuracy and sensitivity to change (responsiveness), to name a few. It is not clear from the study which standard protocol the authors followed when reporting on the FAOQ validation study. Such information should be given in the Methods section.
2. One of the most important aspects of the validation process is construct validity. The term *construct* is adapted from psychology, from the works of G.A. Kelly [13] and should not be confused with *construction*, as the authors did. A construct is a theoretical variable, not directly observable, an ideal object through which each person perceives and interprets reality. The creation of constructs is an effect of operationalization, which is a process of defining the measurement of a phenomenon that is not directly measurable. Construct validity is defined as the degree to which a measurement instrument (or its domain) is consistent with hypotheses based on the assumption that the measurement instrument validly measures the construct (characteristic) to be measured. Construct validity used to be examined by testing an *a priori* set of hypotheses about the expected relationships between the selected domains of condition-specific (targeted specifically at the body region of interest) and generic (that measures capture health conditions as a whole) instruments representing similar and dissimilar constructs. Since the FAOQ questionnaire has no distinct domains, it seems justified to hypothesize that the correlation between the FAOQ taken as a whole and SF-36

- winy wykazywać współzależność (korelację) dodatnią a domeny oceniające zjawiska ze sobą niezwiązane brak takiej współzależności, lub nawet korelację ujemną. Ponieważ kwestionariusz FAOQ nie jest podzielony na domeny należałoby postawić *a priori* hipotezę, że współzależność pomiędzy FAOQ traktowanym całościowo a domenami SF-36 reprezentującymi Zdrowie Fizyczne (ang. *Physical Health*) będzie średnia lub silna a współzależność pomiędzy FAOQ i domenami SF-36 reprezentującymi Zdrowie Psychiczne (ang. *Mental Health*) słaba. W pracy dr. Boszczyka i wsp. nie przedstawiono hipotez *a priori* i nie podano dokładnych wartości korelacji.
3. W pracy dr. Boszczyka i wsp. podano, że badanie zostało przeprowadzone u chorych, którzy byli poddani „leczeniu ortopedycznych schorzeń stopy i stawu skokowego”. Przytoczone sformułowanie sugeruje, że grupa badana nie była jednorodna, zarówno pod względem rozpoznania, jak i wieku. W pracy nie ma także informacji, jakiej interwencji terapeutycznej zostali poddani oceniani pacjenci. Wcześniejsze badania wykazały, że *konstrukt* jest różny u chorych w różnym wieku leczonych z powodu różnych chorób różnymi metodami. Zarówno objawy, czynność stawu skokowo-goleniowego i stopy, jak i oczekiwania co do efektów leczenia są inaczej postrzegane przez chorych młodszych leczonych na przykład z powodu złamania kości śródstopia, inaczej zaś przez pacjentów starszych poddawanych na przykład usztywnieniu stawu skokowo-goleniowego z powodu choroby zwyrodnieniowej. Z tego względu przyjmuje się, że walidację PRO powinno przeprowadzać się na jednorodnej grupie chorych (choć obojga płci), czyli w określonej grupie wiekowej i z określoną chorobą, często także leczonych w ten sam sposób. Badanie wykonane u chorych z różnymi problemami stopy i stawu skokowo-goleniowego ma zatem znacznie mniejszą wiarygodność.
 4. Ważnym elementem procesu walidacji jest ocena rzetelności (ang. *reliability*) instrumentu badawczego. Metoda badawcza uznawana jest za rzetelną, jeśli posiada zdolność pomiaru danego konstrukt w sposób powtarzalny, czyli tak, że wielokrotne oceny tego samego zjawiska (w założeniu nie podlegającego zmianie) dają mniej więcej te same wyniki. Rzetelność jest zatem miarą spójności (ang. *consistency*) i stabilności metody badawczej. Ocena rzetelności instrumentu badawczego przeprowadza się u osób w stabilnym stanie klinicznym [14] na wiele sposobów, między innymi poprzez wykonanie badania z użyciem różnych zestawów tego samego testu (ang. *internal* subscales representing Physical Health should be at least moderate and in the same direction. On the other hand, one could hypothesize that the correlation between the FAOQ and SF-36 subscales representing Mental Health should be weak. The study does not accurately describe which SF-36 domains were tested and what correlation values were obtained.
 3. The authors wrote that the study was performed in “patients treated for foot and ankle conditions”. This suggests that the study group was not homogeneous with respect to either diagnosis or age. In addition, there is no information concerning what therapeutic intervention (if any) was performed in the subjects. In earlier studies, it has been observed that constructs are different in patients from different age groups and in patients treated with different methods for different conditions. Both symptoms and foot/ ankle function are perceived in a different way in younger subjects undergoing surgery of a metatarsal fracture and in older individuals treated with ankle arthrodesis due to osteoarthritis. Thus, it is suggested that validation be performed in patients with specific conditions undergoing the same type of surgery. A different approach decreases the credibility of the study.
 4. Another crucial issue in any validation process is to assess the reliability of the measurement tool. Reliability is the extent to which scores for patients who have not changed are the same in repeated measurements under several conditions e.g. using different sets of items from the same PRO (internal consistency), over time (test-retest reliability), by different persons on the same occasion (inter-rater reliability), or by the same persons (i.e. raters or responders) on different occasions (intra-rater reliability). Reliability is regarded as an estimate of the consistency and stability of a measure. It is assessed in patients in stable clinical conditions, e.g. several months after surgical intervention [14]. Since the article does not state if the study patients underwent a surgical intervention, one could not claim if the test-retest reliability was evaluated in clinically stable conditions.
 5. Responsiveness, defined as the ability of the measurement tool to detect meaningful clinical change over time in the construct to be measured is regarded to be a determinant of usefulness of the instrument. Assessment is usually performed through at least two administrations of a questionnaire in a longitudinal design. The scores from before and after recovery following therapeutic intervention are usually compared. In addition, the clinical change detected by one instrument

consistency), w dwóch kolejnych badaniach (ang. *test-retest reliability*) przeprowadzanych w niewielkim odstępie czasu, badania przez różne osoby przy tej samej okazji (ang. *inter-rater reliability*), czy też przez te same osoby przy różnych okazjach (ang. *intra-rater reliability*). Na podstawie wyników obu badań oblicza się współczynnik korelacji wewnątrz klasy (ang. *intra-class correlation coefficient*, ICC) lub też współczynnik kappa. W omawianym artykule niejasne jest w jakim czasie od ewentualnej interwencji terapeutycznej badanie test-retest zostało przeprowadzone i czy oceniani chorzy pozostawali w stabilnym stanie klinicznym.

5. Zdolność instrumentu pomiarowego do uchwycenia zmiany klinicznej zachodzącej w czasie (ang. *responsiveness*) w konstrukcie podlegającym ocenie/pomiarowi jest podstawowym wyznacznikiem użyteczności tego instrumentu. Ocenę zmiany przeprowadza się najczęściej porównując wyniki badania sprzed interwencji terapeutycznej i po przeprowadzonym leczeniu. Dodatkowo, zmianę wykazaną przez instrument pomiarowy porównuje się z kryterium zewnętrznym (ang. *external criterion*), stanowiącym odniesienie, tzw. „złoty standard”. Jako standard służą najczęściej instrumenty przeznaczone do oceny zmiany klinicznej, tzw. *Global Rating of Change*, z których najpopularniejszymi są: *Global Perceived Effect*, *Patient Global Impression of Change*, *Transition Ratings*, oraz *Global Scale* [15]. Szkoda, że autorzy walidacji FAOQ nie wykonali badania stanu klinicznego chorych przed leczeniem i nie ocenili zdolności FAOQ do uchwycenia zmian zachodzących w czasie.

Przedstawione powyżej uwagi nie umniejszają w żaden sposób wartości pracy wykonanej przez dr. Boszczyka i wsp. Przeprowadzona przez autorów walidacja wypełnia lukę w piśmiennictwie polskim. Dostępność skali FAOQ w naszym języku pozwoli na przeprowadzenie badań klinicznych u chorych leczonych z powodu chorób stopy i stawu skokowo-goleniowego.

can be compared with the change detected by comparison with an external criterion regarded as a gold standard. Variable instruments (Global Rating of Change instruments) designed to quantify a patient's improvement or deterioration over time are available, including Global Perceived Effect Scales, Patient Global Impression of Change, Transition Ratings and Global Scale [15]. It is somehow disappointing that the authors did not perform their clinical assessment in a longitudinal manner and did not assess the ability of the FAOQ to capture clinical change over time.

In conclusion, the study performed by the authors fills a serious gap in the Polish literature and provides an important instrument for assessing foot/ankle conditions. We really appreciate the authors' work and hope that our considerations will help Polish orthopaedic surgeons understand the processes of cultural adaptation and validation of PROs and encourage them to use such measures in clinical investigations.

PIŚMIENNICTWO / REFERENCES

1. Boszczyk A, Błoński M, Pomianowski S. Translation, Cultural Adaptation and Validation of Polish Version of Foot and Ankle Outcomes Questionnaire. *Ortop Traumatol Rehab* 2015; 17: 175-187.
2. Spertus JA. Key issues in outcomes research. evolving applications for patient-centered health status measures. *Circulation* 2008; 118: 2103–2110.
3. Pynsent PB. Choosing an outcome measure. *J Bone Joint Surg [Br]* 2001; 83: 792–794.
4. Golicki D, Krzysiak M, Strzelczyk P. Translation and Cultural Adaptation of the Polish Version of the Disabilities of the Arm, Shoulder and Hand (DASH) and QuickDASH Questionnaires. *Ortop Traumatol Rehabil* 2014; 16: 387-395.
5. Piontek T, Ciemniowska-Gorzela K, Naczek J, Cichy K, Szulc A. Linguistic and cultural adaptation into Polish of the IKDC 2000 subjective knee evaluation form and the Lysholm scale. *Polish Orthop Traumatol* 2012; 77: 115-119.
6. Paradowski PT, Witoński D, Kęska R, Roos EM. Cross-cultural translation and measurement properties of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) following anterior cruciate ligament reconstruction. *Health Qual Life Outcomes* 2013; 11: 107.

7. Paradowski PT, Kęska R, Witoński D. Validation of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in patients with osteoarthritis undergoing total knee replacement. *BMJ Open* 2015; 5: e006947.
8. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000; 25: 3186–3191.
9. Lohr KN, Aaronson NK, Alonso J et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clinical Therapeutics* 1996, 18:979-992.
10. Bombardier C, Tugwell P. Methodological considerations in functional assessment. *J Rheumatol Suppl* 1987; 14 Suppl 15: 6-10.
11. Andresen EM, Meyers AR. Health-related quality of life outcomes measures. *Arch Phys Med Rehabil* 2000; 81: S30-S45.
12. Mokkink L, Terwee C, Knol D et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010; 10: 22.
13. Kelly GA. *The Psychology of Personal Constructs. Vol. I and II.* Norton, New York. 1955.
14. Paiva CE, Barroso EM, Carnesecca EC et al. A critical analysis of test-retest reliability in instrument validation studies of cancer patients under palliative care: a systematic review. *BMC Medical Research Methodology* 2014, 14: 8.
15. Kamper SJ, Maher CG, Mackay G. Global Rating of Change Scales. A Review of Strengths and Weaknesses and Considerations for Design. *J Man Manip Ther* 2009; 17: 163–170.

Szanowny Panie Redaktorze,

Z zadowoleniem przeczytałem uwagi doktorów Paradowskiego i Witońskiego do artykułu opublikowanego w niedawnym numerze *Czasopisma*. W pełni zgadzam się ze zdaniem autorów, że w ocenie wyników leczenia konieczne jest stosowanie instrumentów badawczych oceniających wyniki z punktu widzenia chorego (patient related outcome measure – PROM). Nasza praca wypełniła lukę, jaką stanowił brak adaptowanej polskojęzycznej wersji PROM do badania schorzeń stopy i stawu skokowego.

Przeprowadzony przez nad proces walidacji skali był analogiczny do walidacji skal opublikowanych w ostatnich latach na łamach czasopism polskich [1,2] i zagranicznych [3,4]. Uwagi dr. Paradowskiego oraz dr. Witońskiego potwierdzają zasadność nadal prowadzonych przez nasz zespół prac mających na celu walidację opracowanej skali w poszczególnych jednostkach klinicznych.

Z poważaniem,
Andrzej Boszczyk

Dear Editor,

We were pleased to read the Dr Paradowski and Dr Witoński's comments regarding our article published in a recent issue of *OTR*. We fully agree with the opinion that the measurement of treatment outcomes requires the use of a patient-related outcome measure (PROM). Our work delivered the Polish version of a PROM applicable to the foot and ankle.

Our validation process was analogous to the validation of questionnaires published in recent Polish [1,2] and international publications [3,4]. Nevertheless, Dr Paradowski and Dr Witoński's comments reinforce our current efforts to validate the questionnaire in selected foot and ankle conditions.

Kind regards,
Andrzej Boszczyk

PIŚMIENNICTWO / REFERENCES

1. Czarnecki P, Wawrzyniak-Bielęda A, Romanowski L. Polish Adaptation of Wrist Evaluation Questionnaires. *Ortop Traumatol Rehabil* 2015;17:241–8.
2. Golicki D, Krzysiak M, Strzelczyk P. Translation and Cultural Adaptation of the Polish Version of the Disabilities of the Arm, Shoulder and Hand (DASH) and QuickDASH Questionnaires. *Ortop Traumatol Rehabil* 2014;16:387–95.
3. Paez-Moguer J, Budiman-Mak E, Cuesta-Vargas AI. Cross-cultural adaptation and validation of the Foot Function Index to Spanish. *Foot Ankle Surg* 2014;20:34–9.
4. Öksüz Ç, Akel BS, Oskay D, Leblebicioğlu G, Hayran KM. Cross-cultural adaptation, validation, and reliability process of the Michigan Hand Outcomes Questionnaire in a Turkish population. *J Hand Surg Am* 2011;36:486–92.