

Journal of Gazi University Health Sciences Institute

journal homepage: <https://dergipark.org.tr/tr/pub/guhes>

An Overview of Outcome Measures in Healthcare

Şeyda Özal¹, Nihan Kafa², Nevin Atalay Güzel²

¹ Ankara Medipol University, Faculty of Health Sciences, Ankara, Türkiye

² Gazi University, Faculty of Health Sciences, Ankara, Türkiye

Article info:

Received: 17.03.2023

Accepted: 17.05.2023

Keywords:

clinician-administered,
outcome measures,
self-reported,
terminology

Abstract

Consideration of indicators of an individual's perceptions and attitudes is valuable to reflect the impact of physical findings on quality of life. In this regard, practical outcome measures assist in concluding. Self-Reported Outcome Measures (SROMs) corroborate medical findings and allow for easy profiling (eliciting information about the patient's condition). Even if their frequent usage in clinical and scientific studies, there is confusion over taxonomy, terminology, and definitions of outcome measures. Sometimes, clinicians face some obscurity of the complexity of the features and concepts those measurements represent. This leads to misnaming or misidentification of measurement tools. Users also have difficulty choosing the right tool because of this confusion. This study aims to describe and standardize the terminology of outcome measures and to clarify the classifications with examples. Thus, we aimed to increase the knowledge and awareness of health professionals about the measurement tools, make them interrogate their qualities and features, and encourage them to choose more appropriate ones.

Şeyda Özal (**Corresponding author**); ORCID: 0000-0003-2524-2180, e-mail: seyda.ylmz@yandex.com
Nihan Kafa; ORCID: 0000-0003-2878-4778, e-mail: fztnehankafa@gmail.com
Nevin A. Güzel; ORCID: 0000-0003-0467-7310, e-mail: natalay@gazi.edu.tr

Citation: Özal, Ş., Kafa, N. & Atalay Güzel, N. (2023). An overview of outcome measures in healthcare. *Journal of Gazi University Health Sciences Institute*, 5(2), 68-76. <https://doi.org/10.59124/guhes.1266109>

1. Introduction

Clinicians try to get as much information about the patient as possible to solve the problem (Nelson et al., 2015). Clinical measurements are required for examination, diagnosis, making a clinical decision, choosing the best treatment method, checking the results, comparing the results with other patients, and determining the priority of the various treatment methods (Dawson et al., 2010; Francis et al., 2016; Grove et al., 2021; Nelson et al., 2015). Objective outcomes such as vital signs, laboratory tests, performance tests, and imaging methods maintain precious information about the patient (Fischbach & Dunning, 2009). However, a patient; may have subjective symptoms arising from their feelings, experiences, environment, and mood, independent of medical findings (Dawson et al., 2010; Francis et al., 2016; Guyatt & Schunemann, 2007). Objective outcomes do not reflect the patient's satisfaction with the treatment, perceptions, and feelings about the disease. After all, the impacts of these subjective signs on the healing process should not be underestimated. Understanding the patient's perception, experience, expectations, and feelings about the situation enhances interpreting all these measurements correctly (Guyatt & Schunemann, 2007; Snyder et al., 2007). In this regard, our best helpers are the disease, symptom, function, or population-specific scales, questionnaires, or indexes. In favor of these helper tools, we can imagine the situation of patients and empathize with them better (Dawson et al., 2010; Guyatt & Schunemann, 2007; Snyder et al., 2007).

Outcome measures are used in clinics to strengthen diagnosis, determine the severity of the pathology, and monitor status changes over time (Wyrwich &

Wolinsky, 2000). They can be also used for screening population health in epidemiological studies (Basch et al., 2015). Outcome measures are significant criteria for individual or group analyses in quality improvement initiatives, clinical trials, or observational studies (Black, 2013; Dawson et al., 2010; Zheng et al., 2014).

Although the use of measurement and evaluation tools in healthcare settings began later than in fields such as education, sociology, and marketing, it has gained momentum due to its convenience. Even if its frequent usage in clinical and scientific studies, there is confusion over taxonomy, terminology, and definitions of outcome measures (Mokkink et al., 2010). There is a complexity in features and concepts that measurement represents. This situation causes them to be misnamed or faulty in descriptions. Also, users have difficulty choosing the right tool. This study aims to describe and standardize the terminology and definitions of outcome measures and to clarify the classifications with examples. Thus, we aimed to increase the knowledge and awareness of health professionals about the measurement tools, encourage them to choose more suitable ones, and make them interrogate their qualities and features. We also think this study will increase the use of existing tools, facilitate the identification of new tool needs, encourage the creation of new tools, and guide researchers in scale development studies.

2. Classification of Outcome Measures

We can classify subjective measurements in healthcare as Clinician-Adminstrated (CAOM), Self-Reported Outcome Measures (SROM), or Mixed-type (hybrid) scales (Michener, 2011; Mokkink et al.,

2010; Snyder et al., 2007). The classification scheme of outcome measures is given in Figure 1. Visual Analogue Scale, Numeric Ratio Scale, Verbal Descriptive Scale, and Facial Expression Scale are the prevalent response indication types. Likert-type interval scales are another preferred response type for these health assessment tools (Dawson et al., 2010). The response type should be proper for the target group and intended feature.

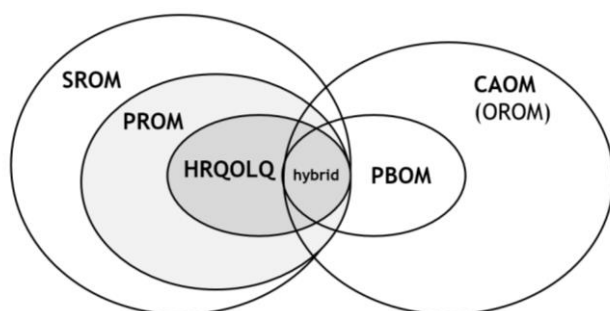
3. Clinician - Administrated Outcome Measures

In CAOMs, the patient is evaluated by a healthcare professional. The clinician scores the patient's condition by interpreting the results of self-observation and assessment. Therefore, they reflect the physical capability findings rather than the patient's perception. Scoring is dependent on the clinician's knowledge and experience. Accurate detection of the changes over time also depends on the practitioner. Inter-rater Consistency may be lower in such scales than in SROMs (Edwards et al., 2002). If it is based on the measurement result of performance without interpretation, it is called a Performance-based outcome measure. CAOMs may also be referred to as Observer-Rated Outcomes Measures (OROM). However, since it is preferred that the evaluator be a healthcare professional, the term Clinician-Administered is more appropriate. If the eventual point is based on performance without interpretation, it could be called a Performance-Based Outcome Measure (PBOM) (Figure 1).

4. Self-Reported Outcome Measures

These scales are measurement tools based on descriptions, declarations, expressions, or statements

directly reported by the individual, without interpretation by a clinician, evaluator, or another person (Dawson et al., 2010). General health conditions, wellness, symptoms, life expectations, thoughts, or decisions are questioned. The person evaluates the self-situation and tries to find the closest expression. These tools are called as Self-report Outcome Measures (SROMs), Patient-reported Outcome Measures (PROMs), or Health-Related Quality of Life Questionnaires (HRQOLQ). Although they mean close to each other, they have nuances (Francis et al., 2016). The Quality-of-Life Questionnaires are inclusive scales that investigate disease-related impacts. They ascertain various symptoms and reveal how much a person's quality of life is affected. Evaluation of a single indication will not be sufficient to make this inference. Thus, when a scale with a single or limited focus, such as symptom, function, or emotion, we prefer to call as PROMs. The focus of health research may not always be patients. Sometimes research is on healthy individuals or specific healthy groups such as athletes, women, pregnant, children, and the elderly. It would be more appropriate to call the scales developed for use in these populations SROM (Feeny et al., 2013; Francis et al., 2016) (Figure 1).



SROM Self-Reported Outcome Measure
PROM Patient-Reported Outcome Measure
HRQOLQ Health Related Quality of Life Questionnaire
CAOM (OROM) Clinician Administered Outcome Measure
OROM Observer-Rated Outcome Measure
PBOM Patient-Based Outcome Measure

Figure 1. Classification scheme of outcome measures

5. Prevalent Response Types in Health Outcome Measures

Times Since health-related purports and symptoms are based on individual experience, it is hard to measure and be perceived by another individual who has not experienced the same situation (Albrecht, 1996).

For example, it is not possible to exact define and detect even pain, which is a common symptom of many health problems, and experienced by almost everyone. Because it has such diverse components as that type, region, duration, and severity, and it depends on individual experience, perception, structural features, and environmental, social, and religious characteristics (Loeser & Melzack, 1999). Still, getting subjective information about pain and similar symptoms to determine the condition, prognosis, and treatment results is essential (Albrecht, 1996; Melzack & Raja, 2005). So far, various methods have been tried, and rating systems have been created to measure these symptoms. Visual Analog Scale (VAS), Numeric Ratio Scale (NRS), Verbal Descriptive Scale (VDS), and Facial Expression Scale (FES) are

Prevalent Response Types in healthcare (Figure 2). Studies have shown that all these response types are valid and reliable. A study comparing their validity for response types concluded that NRS, VAS, VDS, and FES in terms of sensitivity, respectively (Ferreira-Valente et al., 2011). However, it should be chosen considering the situation and the target audience. For example, although the sensitivity is lower, the FES method may give more accurate results than NOS or VAS in illiterate elderly or children (Garra et al., 2010). In addition to these types, Likert-type interval measurement tools adapted to various expressions are widely used in questioning attitudes, preferences, perceptions, and behavioral characteristics (Vagias, 2006).

Outcome measures in health care have been designed in various types according to focal points. The purposes of measurement tools vary widely from specific to generic. The specified scales are generally more sensitive to longitudinal changes than generic ones, but they may not capture the impacts of comorbidities. Some generic scales have situation-specific adaptations (Feeny et al., 2013). The principal scale types are compiled below and exemplified in Table 1.

- a. **Symptom Scales** investigate the presence and level of some symptoms (Garra et al., 2010; Melzack & Raja, 2005).
- b. **Symptom Indexes** allow us to make inferences about the disease it is associated with by questioning the existence and levels of various symptoms. The evaluation criteria of the sub-units of such tools are different from each other (Najafov et al., 2020).

- c. **Disease-Specific Scales** that question one or more of the parameters associated with any disease, such as symptoms, function, and mood (Özal et al., 2021).
- d. **Region-Specific Scales** explore the situation in the body area affected by the problem in terms of function, symptoms, and physical characteristics (Roach et al., 1991; van de Water et al., 2016).
- e. **Organ/System-Specific Scales** that examine organs or structures belonging to a particular system in the body (Hutchings et al., 2015).
- f. **Generic Scales** examine the general health level or quality of life regarding physical, mental, and psychosocial health (Hunt et al., 1981).
- g. **Health-Related Quality-of-Life Scales** investigate the consequences of specific pathology findings on the patient's quality of life (Mollaoğlu et al., 2015).
- h. **Emotional Scales** reveal a person's unconscious feelings or states, such as well-being, social isolation, or self-efficacy (Beck et al., 1987; Blanchard et al., 1995; Bradley, 1994).
- i. **Population-Specific Scales** assess the skill, performance, or wellness of a group of people with common characteristics (age, gender, disability, occupation, sport, or art branches) (Curtis et al., 1995; Washburn et al., 1993).
- j. **Attitude and Behavioral Scales** determine individuals' perspectives and perceptions toward any disease, phenomenon, condition, or object (Jaarsma et al., 2009; Merluzzi & Martinez Sanchez, 1997).

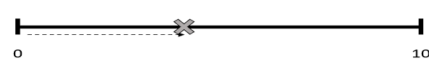
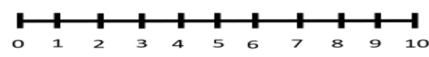





Visual Analog Scale					
Numeric Ratio Scale					
Verbal Descriptor Scale	No Pain	Mild Pain	Moderate Pain	Severe Pain	Unbearable Pain
Facial Expression Scale					
Likert-type (Equal Interval Scale)	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Figure 2. Prevalent Response Types in Health Outcome Measures

6. Conclusion

As in most fields, measurement has great importance in healthcare. Since it relates to clinical decision-making and checking outcomes, it is essential to evaluate patients from various dimensions to strengthen the evidence. In this regard, practical outcome measures assist in concluding. It is valuable to consider the indices of the individual's perceptions and attitudes to reflect the impacts of physical findings on the person's quality of life (Dawson et al., 2010). In other words, adding the scale results based on the person's expressions to the objective data provides the opportunity to evaluate the individual from a larger perspective and helps clinical decision-making. The symbolic, nominal, or numerical outputs obtained from these scales are also used to reveal the temporal changes in the medical condition or compare the differences between individuals or groups (Francis et al., 2016). Both CAOMs and SROMs enhance detailed information about the patient in a short time. They are cheap, easy to understand and apply. The SROMs improve patient-physician communication. Above all, they reflect sincere reports, real sensations, and feelings. Therefore, using these tools in clinical and academic studies is very advantageous. CROMs

are more appropriate than PROMS when patients cannot identify and assess their symptoms or report their health status due to mental or physical disabilities or age (e.g., newborns, elderly with dementia). However, PROMs are more valuable for findings that the clinician cannot detect externally (e.g., pain level). According to the situation, the appropriate one should be selected among the available tools. If possible, CAOMs and PROMs should be used simultaneously to strengthen the findings (Feeny et al., 2013; Powers III et al., 2017).

Despite the numerous advantages of SROMs, they have some disadvantages related to tool suitability or implementation errors. For example, comprehensibility and interpretation of questions or instructions may vary individually, rating methods may not be appropriate for assessment or sample-response matching may be incorrect (Lilienfeld & Fowler, 2006; Paulhus & Vazire, 2007).

Some subjects may not be able to accurately assess themselves or give more socially acceptable answers than true ones (Devaux & Sassi, 2016; McDonald, 2008). Precautions should be taken by taking these disadvantages into account. The most appropriate outcome measure should be selected for the problem, the injured area, and the condition. The person should be encouraged to give correct answers, and the findings should be compared with available more objective measurement parameters.

This article aims to raise awareness and knowledge about outcome measures among practitioners to minimize the disadvantages and maximize the accuracy of SROMs. Thus, clinicians may choose or develop more appropriate tools in this direction.

Table 1. Examples of Outcome Measures

Focal Type	Example Outcome Measure		Measurement Type	Response Type	Sub-sections
	Abbreviation	Full Name			
Symptom Scales	MPQ	McGill Pain Questionnaire	SROM	VDS	-
	WBS	Wong–Baker Faces Pain Rating Scale	SROM	FES	-
Symptom Indexes	LHB score	Long Head of Biceps Score	Mixed	NRS Numbered interval Ratio Scale	Pain/Cramps (SR & CA) Cosmesis (SR & CA) Elbow flexion strength (CA)
Disease-Specific Scales	WOOS	Western Ontario Osteoarthritis of Shoulder Index	SROM	VAS	Physical Symptoms, Sports/Recreation/Work Lifestyle Emotions
Region-Specific scales	SFINX	Shoulder Function Index	CA	Likert type	-
	SPADI	Shoulder Pain and Disability Index	SROM	VAS / NRS	Pain Disability
Organ/System-Specific Scales	GSRQ	Gastrointestinal Symptom Rating Questionnaire	SROM	Likert type Ordinal type Nominal type	-
Generic Scales	NHP	Nottingham Health Profile	SROM	Nominal type	Pain Emotional Reactions Energy Level Physical Mobility Social Isolation Sleep
Health-Related Quality-of-Life Scales	QOLIE-31	Quality of Life in Epilepsy-31 Inventory	SROM	Likert type	Seizure Worry Emotional Well-Being Energy/Fatigue Social Function Cognitive Function Medication Effects Overall Quality of Life
	BDI	Beck Depression Inventory	SROM	Likert type	-
Emotional Scales	WBQ	The Well-being Questionnaire	SROM	Likert type	Depression Anxiety Energy Positive wellbeing
	PTSDS	Post-Traumatic Stress Disorder Scale	SROM/CA	Likert type	-
Population-Specific Scales	PASE	Physical Activity Scale for the Elderly	SROM	Likert type Nominal type	Leisure Time Activity Household Activity
	WUSPI	Wheelchair Users’ Shoulder Pain Index	SROM	VAS	-
Attitude and Behavioral Scales	CBI	Cancer Behavior Inventory	SROM	Likert type	Maintaining Independence Participating in Medical Care Coping and Stress Management Managing Affect
	EHFScB scale	European Heart Failure Self-care Behaviour Scale	SROM	Likert type	-

SROM Self-Reported Outcome Measure, CAOM Clinician Administered Outcome Measure, VAS Visual Analog Scale, NRS Numeric Ratio Scale, VDS Verbal Descriptive Scale, FES Facial Expression Scale

Conflicts of interest

The authors declare no conflicts of interest.

References

- Albrecht, G. (1996). Using subjective health assessments in practice and policy-making. *Health Care Analysis*, 4(4), 284-292. doi:10.1007/BF02249315.
- Basch, E., Spertus, J., Dudley, R. A., Wu, A., Chuahan, C., Cohen, P., . . . Christensen, K. (2015). Methods for developing patient-reported outcome-based performance measures (PRO-PMs). *Value in health*, 18(4), 493-504. doi: 10.1016/j.jval.2015.02.018
- Beck, A. T., Steer, R. A., & Brown, G. K. (1987). *Beck depression inventory*: Harcourt Brace Jovanovich New York: doi: 10.1007/978-1-4419-9893-4_8.
- Black, N. (2013). Patient reported outcome measures could help transform healthcare. *Bmj*, 346. doi: 10.1136/bmj.f167
- Blanchard, E. B., Hickling, E. J., Taylor, A. E., Forneris, C. A., Loos, W., & Jaccard, J. (1995). Effects of varying scoring rules of the Clinician-Administered PTSD Scale (CAPS) for the diagnosis of post-traumatic stress disorder in motor vehicle accident victims. *Behaviour Research and Therapy*, 33(4), 471-475. doi: 10.1016/0005-7967(94)00064-Q
- Bradley, C. (1994). *The well-being questionnaire*: Chur, Switzerland, Harwood Academic Publishers.
- Curtis, K., Roach, K., Brooks Applegate, E., Amar, T., Benbow, C., Genecco, T., & Gualano, J. (1995). Development of the wheelchair user's shoulder pain index (WUSPI). *Spinal Cord*, 33(5), 290-293. doi: 10.1038/sc.1995.65
- Devaux, M., Sassi, F. (2016). Social disparities in hazardous alcohol use: Self-report bias may lead to incorrect estimates. *Eur J Public Health*, 26(1):129-134. doi:10.1093/eurpub/ckv190
- Dawson, J., Doll, H., Fitzpatrick, R., Jenkinson, C., & Carr, A. J. (2010). The routine use of patient reported outcome measures in healthcare settings. *Bmj*, 340. doi: 10.1136/bmj.c186
- Edwards, T. B., Bostick, R. D., Greene, C. C., Baratta, R. V., & Drez, D. (2002). Interobserver and intraobserver reliability of the measurement of shoulder internal rotation by vertebral level. *Journal of shoulder and elbow surgery*, 11(1), 40-42. doi: 10.1067/mse.2002.119853
- Feeny, D. H., Eckstrom, E., Whitlock, E. P., & Perdue, L. A. (2013). A primer for systematic reviewers on the measurement of functional status and health-related quality of life in older adults.
- Ferreira-Valente, M. A., Pais-Ribeiro, J. L., & Jensen, M. P. (2011). Validity of four pain intensity rating scales. *Pain®*, 152(10), 2399-2404. doi: 10.1016/j.pain.2011.07.005
- Fischbach, F. T., & Dunning, M. B. (2009). *A manual of laboratory and diagnostic tests*: Lippincott Williams & Wilkins.
- Francis, D. O., McPheeters, M. L., Noud, M., Penson, D. F., & Feurer, I. D. (2016). Checklist to operationalize measurement characteristics of patient-reported outcome measures. *Systematic reviews*, 5(1), 1-11. doi: 10.1186/s13643-016-0307-4
- Garra, G., Singer, A. J., Taira, B. R., Chohan, J., Cardoz, H., Chisena, E., & Thode Jr, H. C. (2010). Validation of the Wong-Baker FACES pain rating scale in pediatric emergency department patients. *Academic Emergency Medicine*, 17(1), 50-54. doi:10.1111/j.1553-2712.2009.00620.x
- Grove, B. E., Schougaard, L. M. V., Ivarsen, P. R., Kyte, D., Hjollund, N. H., & de Thurah, A. (2021). Patient-reported outcome measures for clinical decision-making in outpatient follow-up: validity and reliability of a renal disease questionnaire. *Journal of patient-reported outcomes*, 5(1), 1-11. doi: 10.1186/s41687-021-00384-0
- Guyatt, G., & Schunemann, H. (2007). How can quality of life researchers make their work more useful to health workers and their patients? *Quality of Life Research*, 16(7), 1097-1105. doi: 10.1007/s11136-007-9223-3
- Hunt, S. M., McKenna, S., McEwen, J., Williams, J., & Papp, E. (1981). The Nottingham Health Profile: subjective health status and medical consultations. *Social Science & Medicine. Part A: Medical Psychology & Medical Sociology*, 15(3), 221-229. doi: 10.1016/0271-7123(81)90005-5
- Hutchings, H. A., Cheung, W.-Y., Russell, I. T., Durai, D., Alrubaiy, L., & Williams, J. G. (2015). Psychometric development of the Gastrointestinal Symptom Rating Questionnaire (GSRQ) demonstrated good validity. *Journal of clinical epidemiology*, 68(10), 1176-1183. doi: 10.1016/j.jclinepi.2015.03.019
- Jaarsma, T., Årestedt, K. F., Mårtensson, J., Dracup, K., & Strömberg, A. (2009). The European Heart Failure Self-care Behaviour scale revised into a nine-item scale (EHFScB-9): a reliable and valid international instrument. *European journal of heart failure*, 11(1), 99-105. doi: 10.1093/eurjhf/hfn007
- Lilienfeld, S. O., & Fowler, K. A. (2006). The Self-Report Assessment of Psychopathy: Problems, Pitfalls, and Promises. *Handbook of psychopathy*, 107-132, ISBN 1-59385-212-6
- Loeser, J. D., & Melzack, R. (1999). Pain: an overview. *The lancet*, 353(9164), 1607-1609. doi: 10.1016/S0140-6736(99)01311-2
- McDonald, J. D. (2008). Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire*, 1(1), 1-19.
- Melzack, R., & Raja, S. N. (2005). The McGill pain questionnaire: from description to measurement. *The Journal of the American Society of Anesthesiologists*, 103(1), 199-202. doi: 10.1097/0000542-200507000-00028
- Merluzzi, T. V., & Martinez Sanchez, M. A. (1997). Assessment of self-efficacy and coping with cancer: development and validation of the cancer behavior inventory. *Health Psychology*, 16(2), 163. doi: 10.1037/0278-6133.16.2.163

- Michener, L. A. (2011). Patient-and clinician-rated outcome measures for clinical decision making in rehabilitation. *Journal of Sport Rehabilitation*, 20(1), 37-45. doi: 10.1123/jsr.20.1.37
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology*, 63(7), 737-745. doi: 10.1016/j.jclinepi.2010.02.006
- Mollaoglu, M., Durna, Z., & Bolayir, E. (2015). Validity and reliability of the quality of life in epilepsy inventory (QOLIE-31) for Turkey. *Nöro Psikiyatri Arşivi*, 52(3), 289. doi: 10.5152/npa.2015.8727
- Najafov, E., Özal, Ş., Kaptan, A. Y., Ulucaköy, C., Kanatlı, U., Ataoğlu, B., & Başar, S. (2020). Validity and reliability of the Turkish version of LHB score. *Journal of Sport Rehabilitation*, 30(1), 30-36. doi: 10.1123/jsr.2019-0364
- Nelson, E. C., Eftimovska, E., Lind, C., Hager, A., Wasson, J. H., & Lindblad, S. (2015). Patient reported outcome measures in practice. *Bmj*, 350. doi: 10.1136/bmj.g7818
- Özal, Ş., Güzel, N. A., Kaptan, A. Y., Eren, T. K., & Kafa, N. (2021). Validity and Reliability of the Turkish Version of Western Ontario Osteoarthritis of the Shoulder Index. *Joint diseases and related surgery*, 32(2), 497. doi: 10.52312/jdrs.2021.141
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. *Handbook of research methods in personality psychology*, 1(1), 224-239 ISBN 9781606236123.
- Powers III, J. H., Patrick, D. L., Walton, M. K., Marquis, P., Cano, S., Hobart, J., . . . Molsen, E. (2017). Clinician-reported outcome assessments of treatment benefit: report of the ISPOR clinical outcome assessment emerging good practices task force. *Value in health*, 20(1), 2-14. doi: 10.1016/j.jval.2016.11.005
- Roach, K. E., Budiman-Mak, E., Songsiridej, N., & Lertratanakul, Y. (1991). Development of a shoulder pain and disability index. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, 4(4), 143-149. doi: 10.1002/art.1790040403
- Snyder, A. R., Valovich McLeod, T. C., & Sauers, E. L. (2007). Defining, valuing, and teaching clinical outcomes assessment in professional and post-professional athletic training education programs. *Athletic Training Education Journal*, 2(2), 31-41. doi: 10.4085/1947-380X-2.2.31
- Vagias, W. M. (2006). Likert-type scale response anchors. *Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University*. doi: 10.13140/RG.2.2.21022.82243
- Van de Water, A., Davidson, M., Shields, N., Evans, M. C., & Taylor, N. F. (2016). The Shoulder Function Index (SFInX): evaluation of its measurement properties in people recovering from a proximal humeral fracture. *BMC musculoskeletal disorders*, 17(1), 1-12. doi: 10.1186/s12891-016-1138-0
- Washburn, R. A., Smith, K. W., Jette, A. M., & Janney, C. A. (1993). The Physical Activity Scale for the Elderly (PASE): development and evaluation. *Journal of clinical epidemiology*, 46(2), 153-162. doi: 10.1016/0895-4356(93)90053-4
- Wyrwich, K. W., & Wolinsky, F. D. (2000). Identifying meaningful intra-individual change standards for health-related quality of life measures. *Journal of evaluation in clinical practice*, 6(1), 39-49. doi: 10.1046/j.1365-2753.2000.00238.x
- Zheng, H., Li, W., Harrold, L., Ayers, D. C., & Franklin, P. D. (2014). Web-based comparative patient-reported outcome feedback to support quality improvement and comparative effectiveness research in total joint replacement. *EGEMS*, 2(1). doi: 10.13063/2327-9214.1130