12.1.1 Rationale for a systematic review of measurement properties

Systematic reviews synthesize the best available evidence and are the keystone of evidence-based practice (Aromataris & Pearson, 2014). A well-conducted systematic review provides a robust, transparent, rigorous method of answering a research question and there are several reasons why a research question relating to the measurement properties of instruments may be asked.

Clinicians and researchers frequently want an instrument to measure a particular attribute that best suits their context. This may require synthesis of published and unpublished results of psychometric testing of instruments claiming to measure the specific attribute e.g. caregiver burden (Whalen & Buchholz, 2009) to find the best instrument for their purpose. Alternatively, there may be a need to establish the relevance and applicability of a specific instrument prior to implementing research findings into practice or using the instrument in research. Conducting a systematic review of the measurement properties of instruments may provide clinicians and researchers with the gold standard instrument and identify settings or contexts in which instruments should or should not be used. Alternatively, a systematic review of measurement properties may identify a gap in knowledge demonstrating where a reliable, valid instrument needs to be developed.

There are many measurement properties to consider when seeking answers to the above questions. However, the predominant domains into which these properties fit are validity (measure what it is supposed to measure), reliability (consistency), and responsiveness (ability to detect change over time). Several publications are available to assist systematic reviewers to identify the domains and items that should be considered when assessing the quality of studies reporting measurement properties of instruments and how best to synthesize them (Francis et al., 2016; Mokkink et al., 2018a; Polit & Yang, 2016; Prinsen et al., 2018).

An international panel of experts in health measurement properties used a consensus approach to develop a taxonomy, terminology, and definitions of measurement properties. The panel also sought consensus on the relevancy of evaluating each property when appraising an instrument. The resultant list, COnsensus-based Standards for the selection of health status Measurement Instruments (COSMIN), formed standards which can be used for the selection of health measurement instruments, peer reviewing a manuscript, designing or reporting a study on measurement properties, or for educational purposes (Mokkink et al., 2010b). A subsequent checklist applying a four-point scale was developed to enable quantifying the overall methodological quality of a study on measurement properties (Terwee et al., 2012), which was updated in 2018 (Mokkink et al., 2018a). The COSMIN initiative has also developed guidelines for systematic reviews of measurement properties for patient-reported outcome measures (PROMs) (Prinsen et al., 2018).

A similar taxonomy was developed (Polit, 2015) and subsequent debate regarding how to organize the measurement properties in domains and what measurements of error should be included has followed (Mokkink et al., 2016; Polit, 2016). Authors acknowledge that the work in this area is evolving. An alternative checklist was created to operationalize measurement characteristics of instruments measuring patient reported outcomes (Francis et al., 2016). The authors viewed the four domain, 119 item COSMIN checklist as too complex and offered a checklist of six domains with a total of 18 items, that could be used by both those with measurement theory expertise and less experienced clinicians (Francis et al., 2016). However, members of the COSMIN initiative have identified shortcomings of this reduced checklist, which may introduce bias in the ratings. One argument is that the criteria presented in the shortened checklist are not detailed enough to provide a transparent and systematic rating of the quality of an instrument (Terwee et al., 2016b). While both checklists by the COSMIN initiative and Francis et al. (2016) were established using patient reported outcomes, they have utility for assessing measurement properties of other instruments. There are several other critical appraisal tools developed for appraisal of studies of measurement properties, but the COSMIN checklist remains the benchmark in this field (Rosenkoetter & Tate, 2017).

JBI has been a leader in synthesizing findings from multiple studies within the framework of evidence-based healthcare. JBI has published guidelines for systematic reviews of many research designs, therefore, it is timely to consider the guidelines available for conducting systematic reviews of measurement properties and provide guidance to systematic reviewers who work within the JBI framework. There are some similarities between systematic reviews of measurement properties and diagnostic test accuracy systematic reviews (see JBI Manual for Evidence Synthesis – Chapter 9), however diagnostic test accuracy reviews specifically focus on the comparison of two tests (index test and reference test) to establish accuracy in identifying the presence or absence of a condition. On the other hand, systematic reviews of measurement properties are used to establish validity, reliability, and responsiveness of one or more instruments that may be used to measure a wide variety of outcomes.

This chapter outlines and describes guidance for synthesizing evidence related to the measurement properties of instruments and contributes to the emerging field of systematic review methodologies. The systematic review of studies to answer questions of validity, reliability, and responsiveness adheres to the same basic principles of systematic reviews of other types of data. An a priori protocol must precede and inform the conduct of the systematic review, comprehensive searching must be performed, and critical appraisal of eligible studies must be carried out by two independent reviewers, followed by data extraction and synthesis. These steps will be further discussed in the following sections of this chapter. Additionally, reviewers should refer to two statements/checklists: one for transparent reporting of a systematic review of various research study designs (Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA)) (Moher et al., 2009) and one for the COSMIN guidance for systematic reviews of patient-reported outcome measures, which provides guidance on standards for methodological quality of studies reporting measurement properties (Mokkink et al., 2018a; Prinsen et al., 2018).