Patient-reported outcome measures for adverse events: A systematic review and COSMIN evaluation study

by

Myles Hancock

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Medical Sciences – Pediatrics

University of Alberta

ABSTRACT

Background: Health care requires constant improvement. Harms in health care are becoming a priority, as is incorporating the patient's voice in both clinical research and clinical care. Patients have been found to provide a more subjective, detailed perspective of their treatment experiences compared with health care providers; this is especially true of potential harms they have experienced. Measurement instruments must be valid and reliable; a new field of Patient Reported Outcome Measures (PROM) has emerged to capture the patient's perspective and experience.

Methods: A systematic review was conducted to identify all patient-reported outcome measures for adverse events (PROM-AE) currently published in the health literature databases. These measures were compared to establish similarities and differences, and to determine if any core characteristics existed.

Results: The most commonly used PROM AE in clinical research and clinical practice were evaluated further with regards to their measurement properties.

Conclusion: Important gaps, such as minimal harms reporting in clinical research and practice, were identified that could help advance the field of PROM AE and thereby enhance patient safety in both research and clinical settings.

Preface

This thesis is an original work by Myles Hancock. No part of this thesis has been previously published

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Harms

There is a drive to continually improve treatment in health care (1). Patient safety is becoming a priority as the seriousness of treatment-related harms is increasingly recognized (2). Compared to treatment efficacy, harms have been grossly underreported in both clinical care and research (2). A 2016 analysis estimated that if deaths due to harms from medical error were incorporated into the US Centre for Disease Control (CDC) yearly reports, it would be the third highest cause of death in the US (3). This estimate of over 250,000 yearly deaths is likely lower than the true number, for only available health records and inpatient death data were utilized (3). Additionally, these projections do not include the yearly occurrence of treatment-related harms not resulting in patient deaths. Under-reporting of harms also plagues clinical research (4).

Over a decade ago, researchers identified the need to improve the quantity and quality of harms reporting. In 2004, the *Con*solidated *S*tandards *o*f *R*eporting *T*rials (CONSORT) group added a harms extension to their original guidelines for collecting data and reporting on harms in randomized controlled trials (RCT) (2, 5). Even though the CONSORT guidelines are widely-endorsed by journals (5), there remains a concerning lack of improvement in harms reporting (6). RCTs have not followed CONSORT guidance and trials continue to provide less than acceptable information about harms (6). Recent systematic reviews found most RCT reporting to have half of the required harms components stated in the CONSORT guidelines (6, 7). For example, a 2014 review of all RCTs of psychological interventions for behavioral disorders published since 2010, discovered that 79% of trials failed to provide any harms reporting (8).

Terminology related to harms has been suggested by a variety of international organizations, including the Canadian Patient Safety Institute (CPSI) (9), the Institute for Health Improvement (IHI) in the US (10), and the International Council for Harmonization (ICH) in Europe (9). Within research, notable groups also involved in defining harms include CONSORT (2004) for RCTs(5), the Cochrane Collaboration (2011)(11) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (2016)(4) for systematic review methodology and reporting standards. The table below illustrates a selection of these definitions:

TABLE 1.1: HARMS TERMINOLOGY

TERM	CONTEXT	REFERENCES
Harm	The totality of possible adverse consequences of an intervention or therapy.	Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-reported outcomes (PROs): putting the patient perspective in patient-centered outcomes research. Medical care. 2013;51(8 0 3):S73.
Adverse Effect	Harms that occur "during or after the use of a drug or other intervention and the causal relation between the intervention and the event is at least a reasonable possibility."	Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. bmj. 2016;352:i157.
Adverse Event (AE)	An unfavorable outcome that occurs during or after the use of a drug or other intervention but is not necessarily caused by it.	Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. bmj. 2016;352:i157.
Adverse Drug Reaction (ADR)	An adverse effect specific to a drug.	Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. bmj. 2016;352:i157.

For the purposes of this thesis, we will use the term “adverse event” (AE) (4) interchangeably with “harms”, as we are interested in an inclusive approach for any/all adverse consequences occurring during or after an intervention or treatment that may or may not be related to it.

Given the current evidence, there is a pressing need to enhance the collecting and reporting of adverse events in clinical practice and clinical research. It is unclear from the literature whether harms associated with health interventions vary across different conditions/diseases. It is only possible to assess and confirm this with considerably more data than currently available. Patient involvement in harms reporting could help address this issue. It is known that patients report harms more often than health care providers (HCPs), and that the quality of their reporting is more detailed and subjective (8). Thus in clinical care, patients may be uniquely able to enhance both the quantity and quality of harms reporting with their experiences (9).

Patient-Reported Outcome Measurement Tools (PROMs)

The patient perspective is increasingly recognized as an essential component to understanding whether health care is beneficial and safe (1). In keeping with this, patient-reported outcomes (PROs) have recently been more utilized in research. PROs are defined as “any report coming directly from a patient about a health condition or treatment” (12). PROs are typically gathered by administering patient-reported outcome measures (PROMs). These are measurement tools that provide the patient’s self-reported viewpoint on their health condition (1, 13, 14). These PROM- related developments have helped the patient voice grow in importance for improving health outcomes (15). Noted regulatory bodies such as the World Health Organization, the US National Institute of Health (NIH) and the US Food and Drug

Administration (FDA) have recommended PROM use in research reports; these policy decisions further illustrate the importance being given to patient feedback in health research (12, 15).

Additionally, PROMs are being considered to augment routine clinical practice (17). Emerging research suggests that use of PROMs may improve care (1, 13). Studies have found that patients provide unique perspectives different from those obtained by HCPs (16). Within the last decade, limited implementation of PROMs has begun in daily clinical care and surveillance (13, 17). The United Kingdom (UK) has committed to working with researchers to test PROM use in their primary health care network (13, 17). Starting in 2008, pilot studies on PROM implementation were conducted in areas such as elective surgeries (12, 13). Patient involvement was tested to guide treatment selection for patients with chronic conditions, physicians' decision-making on scheduling surgeries, and treatment efficacy (13). Similar smaller-scale studies have been published from Sweden and the US (13). Despite this progress, evidence for the impact of PROMs in daily care is yet to be established (13).

PROM Types

Instruments may be classified as generic or disease-specific, and standardized or individualized. A generic PROM would provide patients with any disorder with an opportunity to express their treatment status and current health condition (18). However, this survey-type instrument may possess limited ability to describe experiences common to a specific disease (18). Researchers have therefore developed "disease/condition-specific" measures for PROMs designed to be responsive to patients with a specified disorder (13, 18).

Measurement instruments can also be standardized or individualized. In standardized PROMs, the question content is determined *a priori* (18). This approach facilitates comparisons between respondents. However, standardized measurement instruments have a limited ability to

capture specific patient experiences (18). Conversely, individualized measures allow the patient to identify the constructs most important to them, and then select their desired response (19). The primary limitation of individualized measures is the difficulty of direct comparisons between patients of with different conditions/disorders. A combination of standardized and individualized in a PROM is possible. The Measure Yourself Medical Outcome Profile (MYMOP) questionnaire provides patients with a standardized structured question structure, leaving the symptom open-ended for the patient to add (25). A sample of the tool is described further below.

These key differences (generic vs. disease-specific, standardized vs. individualized) may be best illustrated by looking at examples of PROMs.

Generic and standardized PROM:

The Short Form Health Survey (SF-36) is a well-used generic and standardized tool that is designed for patient reports on their health-related quality of life (HRQOL) (20). To clarify, HRQOL is defined by the US Center for Disease Control (CDC) as a person's "physical and mental health perceptions (e.g., energy level, mood) and their correlates, including health risks and conditions, functional status, social support, and socioeconomic status" (21). The SF-36 is generic because the question types are not geared towards a specific disorder (22); it is termed standardized because the questions are pre-determined; for example:

> "During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or group?"
>
> (i) Not at all, (ii) Slightly, (iii) Moderately, (iv) Quite a bit, (v) Extremely

The Expanded Prostate Cancer Index Composite (EPIC) is a disease specific measure that is also standardized. The question types are geared toward the prostate and urinary health (23): for example:

> "Over the past 4 weeks, how often have you had pain or burning with urination?"
>
> (i) More than once a day, (ii) About once a day, (iii) More than once a week, (iv) About once a week, (v) Rarely or never

Disease specific, individualized PROM:

The Audit of Diabetes Dependent QoL (ADDQoL) is an example of an individualized disease-specific PROM (24). Patients are provided with a list of 13 different items to choose from that help communicate perceptions of their own experience living with diabetes (24). As the PROM is individualized, patients only answer the questions relevant to their experiences. A sample of the first five items and answer choices are listed below:

> "If I did not have diabetes, my_____would (be)..."
>
> 1. ..my employment/career opportunities
> 2. ..my social life
> 3. ..my family relationships
> 4. ..my friendships
> 5. ..my sex life
>
> Answer choices:
>
> (-3) a great deal better, (-2) quite a lot better, (-1) a little better, (0) the same, (1) a little worse, (2) quite a lot worse, (3) a great deal worse

Generic, individualized PROM:

In the Measure Yourself Medical Outcome Profile (MYMOP) questionnaire (25), patients are given the opportunity to list symptoms of their choice, and then measure their severity using a 7 point scale (25):

> Choose one or two symptoms (physical or mental) which bother you the most. Write them on the lines.
>
> Now consider how bad each symptom is, over the last week, and score it by circling your chosen number.
>
> SYMPTOM 1: .....
>
> (As good as it could be) 0 1 2 3 4 5 6 (As bad as it could be)
>
> Now choose one activity (physical, social or mental) that is important to you, and that your problem makes difficult or prevents you doing. Score how bad it has been in the last week.
>
> ACTIVITY: .....
>
> (As good as it could be) 0 1 2 3 4 5 6 (As bad as it could be)

Measurement Properties of PROMs

Accurate knowledge of the quality of measurement properties in health instruments is essential for proper measurement of patient outcomes; this is also true of PROMs (26). The COSMIN group conducted a Delphi study using international expert consensus to determine the key components for outcome measurement instruments, known as domains (27). First is the domain validity; defined as whether the PROM measures the intended variable in the target population (26, 28). Under validity, COSMIN further defined several validity-related measurement properties. Content validity is confirmed when a PROM measures the construct with all essential components of the outcome present and working together (26, 29). Construct

validity is established if the measurement tool demonstrates a strong correlation to similar measures in a relevant patient sample (26, 29). Criterion validity can be measured. This is the degree to which measurement as true as an external evaluation considered to be a "gold standard" measurement tool (26). However, experts have stated that there is no true gold standard for PROMs, given the immense variation of patient outcomes (30).

The second fundamental domain is reliability. COSMIN defines reliability as the extent to which "the measurement is free of measurement error" (28). Therefore, measures with higher reliability possess minimal error (29). The COSMIN group classified reliability into three measurement properties. The first is reliability, defined as "the extent to which scores for patients who have not changed are the same for repeated measurements" (27). COSMIN further defines reliability into different contexts; test-retest ("over time"), inter-rater ("by different persons on the same occasion"), and intra-rater ("by the same persons on different occasions") reliability types (27). The other two measurement properties are internal consistency ("degree of inter-relatedness of the items" (27)); and measurement error ("systematic and random error not attributed to true changes in the construct to be measured" (31)).

The third main domain is responsiveness, defined by COSMIN as "the ability of a PRO instrument to detect change over time in the construct to be measured" (27). This is useful when constructs can vary, such as the severity of an adverse event (AE) through the duration of treatment (26). The fourth domain is interpretability. This is the extent a reviewer "can assign qualitative meaning—that is, clinical or commonly understood connotations—to an instrument's quantitative scores or change in scores" (27).

Thesis Objective

As harms measurement is essential to advancing patient safety, this thesis is focused on PROM-AEs in both clinical research and clinical practice. After a systematic review to identify PROM-AEs, further assessment of measurement properties is conducted on selected instruments.

Chapter Based Thesis Objectives

Chapter 2, "Identifying Patient Reported Outcome Measurement Instruments for Adverse Events (PROM-AE): A Systematic Review", aims to find PROMs that have been reported in the peer-reviewed published health literature to measure AEs. The main goal is to bring all existing PROM-AEs together, compare them, and determine any gaps in the knowledge that may warrant further investigation.

Chapter 3, the "COSMIN Checklist Evaluation of Patient Reported Outcome Measurement tools of Adverse Events (PROM-AE) Validation Studies: PRO-CTCAE & Yellow Card Scheme (YCS)", examines two key PROM-AEs identified in chapter 2. PRO-CTCAE and YCS represent the leading PROM AE in clinical research and clinical practice, respectively. The methodological quality of studies evaluating their measurement properties is examined using the COSMIN group's expert-consensus based checklist.

TABLE 2.2: THESIS OBJECTIVES

Thesis Objective	Study Design	Thesis Chapter
1. Identify all PROM-AEs in the literature and summarize properties	Systematic Review	Chapter 2
2. Evaluate the measurement properties reporting of select PROM-AE validation studies	Identify and describe studies reporting on the measurement properties of select PROM-AEs identified in objective 1.	Chapter 3

REFERENCES

1. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. Journal of evaluation in clinical practice. 2006;12(5):559-68.
2. Ioannidis JP, Evans SJ, Gøtzsche PC, O'neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Annals of internal medicine. 2004;141(10):781-8.
3. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ: British Medical Journal (Online). 2016;353.
4. Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. bmj. 2016;352:i157.
5. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. BMC medical research methodology. 2001;1(1):2.
6. Hodkinson A, Kirkham JJ, Tudur-Smith C, Gamble C. Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension. BMJ open. 2013;3(9):e003436.
7. Shukralla AA, Tudur-Smith C, Powell GA, Williamson PR, Marson AG. Reporting of adverse events in randomised controlled trials of antiepileptic drugs using the CONSORT criteria for reporting harms. Epilepsy research. 2011;97(1):20-9.
8. Jonsson U, Alaie I, Parling T, Arnberg FK. Reporting of harms in randomized controlled trials of psychological interventions for mental and behavioral disorders: A review of current practice. Contemporary Clinical Trials. 2014;38(1):1-8.
9. Pohlman KA, O'Beirne M, Thiel H, Cassidy JD, Mior S, Hurwitz EL, et al. Development and validation of providers' and patients' measurement instruments to evaluate adverse events after spinal manipulation therapy. European Journal of Integrative Medicine. 2014;6(4):451-66.
10. Griffin FA, Resar RK. IHI Global Trigger Tool for measuring adverse events. Institute for Healthcare Improvement Innovation Series White Paper. 2009.
11. Higgins JPT, Green S. Cochrane Handbook on Systematic Reviews of Interventions 2011 [Web Page]. Available from: http://handbook.cochrane.org/.
12. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-reported outcomes (PROs): putting the patient perspective in patient-centered outcomes research. Medical care. 2013;51(8 0 3):S73.
13. Black N. Patient reported outcome measures could help transform healthcare. BMJ (Clinical research ed). 2013;346:f167.
14. Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? Social science & medicine. 2005;60(4):833-43.
15. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Medical care. 2007;45(5 Suppl 1):S3.
16. Bateman E, Keefe D, editors. Patient-reported outcomes in supportive care. Seminars in oncology; 2011: Elsevier.
17. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. Bmj. 2010;340:c186.
18. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Annals of internal medicine. 1993;118(8):622-9.
19. Paterson C. Measuring outcomes in primary care: a patient generated measure, MYMOP, compared with the SF-36 health survey. Bmj. 1996;312(7037):1016-20.

20. Ware Jr JE. SF-36 health survey update. Spine. 2000;25(24):3130-9.
21. Prevention CfDCa. Health Related Quality of Life (HRQOL) concepts: U.S. Department of Health & Human Services; 2016 [cited 2017 March 27th]. Available from: https://www.cdc.gov/hrqol/concept.htm.
22. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Medical care. 1992:473-83.
23. Wei JT, Dunn RL, Litwin MS, Sandler HM, Sanda MG. Development and validation of the expanded prostate cancer index composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer. Urology. 2000;56(6):899-905.
24. Bradley C, Todd C, Gorton T, Symonds E, Martin A, Plowright R. The development of an individualized questionnaire measure of perceived impact of diabetes on quality of life: the ADDQoL. Quality of Life Research. 1999;8(1):79-91.
25. Paterson C, Britten N. In pursuit of patient-centred outcomes: a qualitative evaluation of the 'Measure Yourself Medical Outcome Profile'. Journal of health services research & policy. 2000;5(1):27-36.
26. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Group MFP-ROCM. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value in Health. 2007;10:S94-S105.
27. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. Journal of clinical epidemiology. 2010;63(7):737-45.
28. De Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide: Cambridge University Press; 2011.
29. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of clinical epidemiology. 2007;60(1):34-42.
30. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury. 2011;42(3):236-40.
31. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of Life Research. 2010;19(4):539-49.

CHAPTER 2: Patient Reported Outcome Measurement Instruments for Adverse Events (PROM-AE): A Systematic Review

Myles Hancock, *Masters Candidate*[1], Liliane Zorzela, *PhD*[1], Lisa Hartling *Professor*[1,2]„ Don Voaklander, *Professor*[3], Sunita Vohra, *Professor*[1,3]

[1] Department of Pediatrics, Faculty of Medicine & Dentistry, University of Alberta

Edmonton Clinic Health Academy
11405-87 Avenue
Edmonton, Alberta
Canada T6G 1C9
[2]Alberta Research Center for Health Evidence, University of Alberta
Edmonton Clinic Health Academy
11405-87 Avenue
Edmonton, AB T6T 1C9
[3]School of Public Health, University of Alberta
3-300 Edmonton Clinic Health Academy
11405 – 87 Ave
Edmonton, Alberta
Canada T6G 1C9

Correspondence to: S Vohra svohra@ualberta.ca

ABSTRACT

Objective: To search the published literature for patient-reported outcome measures for adverse events reporting (PROM-AE). Once compiled, compare and establish trends and gaps in knowledge.

Methods: Systematic review with comprehensive search in five major databases. Eligible studies were: (i) Published in English language; (i) any patient population; (iii) documented the development, evaluation, or use of a PROM-AE. Screening and data extraction were performed by one reviewer and verified by a second reviewer.

Results: Screening of 7776 citations yielded 533 articles for potential inclusion. 69 articles met inclusion criteria. Most of the 69 included were validation studies (n=24, 34%). The number of PROM-AEs identified was 35, with the majority designed for use in the field of cancer research (n=17, 60%). The most referenced PROM-AE was the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) (n=10, 15.6%).

Conclusion: This is the first systematic review to search the literature to find all PROM-AEs published in research. Most tolls were used in cancer research, with less focus in other areas. There is potential for further research and daily clinical application of PROM-AEs. We encourage the incorporation of patient feedback on their AE experiences in both settings, which may enhance communication and improving patient care.

INTRODUCTION

Patient Reported Outcome Measures

Patient-reported outcomes (PRO) have gained prominence in current medical research (1). PROs have been defined as "any report coming directly from the patient about a health condition or treatment" (2). Recent progress in PRO use indicates an increasing desire in the healthcare field to know the patient's perspective, as more subjective experiences of treatment are likely only known to them (3). If the patient's perspective is not sought, important outcomes may be lost. Recognizing the importance of this field, in 2010 the United States Congress launched the Patient Centered Outcomes Research Institute (PCORI) (4). The overarching goal of PCORI is to bring attention to outcomes that directly concern the patient, and to find evidence-based information on these outcomes to guide better health care decisions (4). Recently Canada has joined this movement, developing a Strategy for Patient Oriented Research (SPOR) (7).

Patient reported outcome measures (PROMs) are measurement instruments that are developed for patients to complete to express their views on their own condition (5). The US Food and Drug Administration (FDA), the World Health Organization (WHO) and the US National Institute of Health (NIH) encourage the use of PROMs in research, underscoring their importance (2, 3, 6). Obtaining a patient's direct report on their current treatment may play a vital role in the course of treatment decisions in clinical care as well as clinical research. The United Kingdom has implemented PROM testing in daily clinical care nationwide (8). Initiatives that utilize patient input directly in this way show promise to enhance treatment satisfaction and enable any required course corrections (8). Ultimately, a balanced evaluation of treatment effect

requires measuring and reporting both benefit and harms in both clinical practice and clinical research (9).

Researchers have further classified PROMs into different types; one is generic (applicable to any patient) or disease/condition specific (designed for a particular condition or patient type) (10). Another variation of PROMs is standardized (set questions that all patients answer) or individualized (patients identify what questions are most relevant to them) (8, 10).

Harms

Since the turn of the 21st century, widely endorsed research groups and regulatory bodies have put effort into defining harms in the context of health research and daily care (9, 11-13). The term "harms" has been stated as being any potential adverse outcomes in medical treatment (11). There are a number of related terms, from "adverse effect" and "adverse event", to "adverse drug reaction" (Please see Table 1 for full definitions). For the purposes of this study, we used the term "adverse events" (AE), defined as "an unfavorable outcome that occurs during or after the use of a drug or other intervention but is not necessarily caused by it" (9).

Harms have been underreported in both research and clinical settings (14). While CONSORT, the world's leading clinical trial reporting guideline, recommends that harms should be required outcome reporting in clinical trials (5), AEs are not yet routinely sought nor well reported by trial authors (11). Lack of harms reporting may create the misconception that an intervention is safe, or that the benefits outweigh potential harms; often it is more accurate to state that it is not known whether a health intervention is safe or not (14). Effective identification and reporting of harms can improve patient safety and reduces preventable AEs in both clinical research and care (15). Direct patient input has been found to promote timely course correction of treatment to limit patient suffering (14).

Identifying and reporting treatment-related harms in clinical practice is important. However, relatively little attention has been paid to whose perspective should consistently be sought: the provider, the patient, or both (14, 15). Historically, documentation of AEs mainly relied on physicians; however, evidence shows that they often overlook identifying and reporting treatment-related harms (16). As in clinical research, there is increasing awareness in the health care system that patient reports of harms are more comprehensive than information provided by health care providers alone (14, 15). Studies comparing patient versus physician outcome measure appraisal scores have shown statistically significant differences in reported symptom severity; this indicates that physicians may not fully understand a patient's subjective experience of their symptoms (16).

Given the importance of further investigating patient AE reporting, the main objective of this systematic review is to identify any PROM-AE measurement tools in the peer reviewed literature.

Review Question

What Patient-Reported Outcome Measures (PROM) for Adverse Events (AE) have been reported in the peer reviewed medical literature?

METHODS

The "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" (PRISMA) statement was used as a reporting guideline for this review (17).

Search and Information Sources

With guidance from a health research librarian, a systematic search of the following five databases was completed, spanning from database inception to April 2016: Medline, Embase,

CINAHL, Cochrane Database of Systematic Reviews (CDSR), and the Cochrane database of methodology register. Search strategies for the databases are provided in Appendix A. A filter for AEs developed by Su Golder of the Cochrane Collaboration Adverse Effects Methods Group was used with her consultation and approval in the search strategy (18). The terms in Table 1 were also used in the search strategy to identify relevant studies whose authors may use various terms to describe AEs.

Eligibility Criteria

Any articles found that documented the development, evaluation, or use of a specified PROM-AE was included. Retrieved studies that mentioned a PROM-AE measure but did not provide information on the relevant measurement properties were included; systematic, scoping and literature reviews that mentioned PROM-AEs were also accepted (Appendix B).

We anticipated that deciding between PROM-AE and health-related quality of life (HRQOL) measurement instruments may be difficult, as some HRQOL tools measure pain and/or harm-related aspects of the patient's condition. Therefore, the context of PROM-AE use within each study was examined to make the final decision. We used the terms "adverse event" (AE) (4) and "harms" interchangeably, as we were interested in any and all adverse consequences occurring during or after an intervention or treatment, that may or may not be related. Sometimes HRQOL is used in the context of treatment-related outcome assessment. Therefore, if a tool is designed for measuring patient QOL but is used specifically for a treatment-related AE, it will be included.

Studies included in this review also met the following basic criteria: (i) published in English; (ii) studied clinical populations of any age (pediatric, adult, or mixed); (iii) full text original research concerning PROM-AEs or studies that reported such an instrument in their

methods. Hand searching of reference lists of included articles was done to find any further information on the specific tool mentioned; those articles were included as well. The main goal of hand-searching was to find the PROM-AEs themselves if the original included articles did not contain or describe its components. Therefore, not every article of each PROM-AE identified was included. Two independent researchers applied the criteria to include relevant articles (MH, LZ). A third reviewer, the senior investigator (SV), helped resolve any discrepancies until consensus was reached.

Data Extraction and Analysis

Data were extracted from the included articles and entered into a Microsoft Excel document by two independent reviewers (MH, LZ). Information retrieved comprised of intervention, patient type (disorder, age, and sex), setting, and whether the tool was standardized (patient answers all questions) or individualized (patient chooses topic/questions to answer), condition-specific (geared to a certain patient diagnosis) or generic (applicable to all patients). We extracted all AEs mentioned in the PROM-AEs and listed them on Table 2A/2B. We also noted the key components, or constructs, of each PROM-AE, compiled them, and documented frequency and what proportion of included tools possessed these constructs.

The goal of this systematic review was not a quantitative assessment of treatment effect (or harm), but to identify and describe the reported PROMs for AEs. Therefore, a meta-analysis to combine data was not appropriate, and no quality appraisal/risk of bias assessment was done. Relevant data that were extracted from the included articles were displayed graphically or tabulated as appropriate, and described narratively.

RESULTS

The systematic search found 7706 entries; initial screening identified 533 potentially relevant studies. After full text examination, 52 articles were included. From hand searching reference lists of included articles, an additional 17 were added, such that 69 articles were included in total. The reason for excluding the 486 references was no mention or use of a PROM-AE tool. Please see the PRISMA flow diagram for further details (Figure 1).

Within the 69 included articles were: validation studies (n=24, 34.8%); RCTs (n=11, 15.9%); cohort studies (n=7, 10.1%), literature reviews (n=6, 8.7%), systematic reviews (n=4, 5.8%), editorials (n=4, 5.8%), and pilot studies (n=4, 5.8%). Please see Table 3 for a full list of study types. Publication year ranged from 1990 to 2016, with a median of 2010 (IQR: 2004, 2014). Of the 69 included studies, 51 were experimental and recruited patients for an intervention. Among these articles, the patient sample size ranged from 19 to 3759, with a median of 173 (IQR=80, 467.5). The majority of included studies had an adult sample (n=49), while two other studies had pediatric samples. Of the 49 adult patient studies, 31 reported an age range. Of these, the median of the minimum age was 24 years (IQR: 21, 32), and the maximum age was 79 years (IQR:73, 84). The age ranges for the two pediatric studies were 7-16 years and 5-13 years.

The included articles focused on a variety of interventions; chemotherapy was the most common (n=28, 40.6%), followed by radiotherapy (n=7, 10.1%), nonspecific treatment/drugs (n=5, 7.2%), surgery (n=4, 5.8%) and opioid drugs (n=4, 5.8%). Please see Table 4 for a full list of treatments.

This review identified 35 different PROM-AE measures among the 69 included articles. We compared the characteristics of the 35 PROM-AEs found. Please see Table 2 for a list of PROM-AEs and the included articles. Almost all were condition-specific with standardized questions (n=33, 94.3%). Generic with standardized questions, generic with individualized each had one PROM-AE tool (n=1, 2.9%). There were no PROM-AEs found that were disease-specific and individualized. Please see Table 2A for more details of the condition-specific PROM-AEs, and Table 2B for all generic tools found.

Among the PROM-AEs, the most common condition assessed was cancer (n=17, 60%) with various types included in the study samples. The second most prevalent diagnoses were schizophrenia and post-surgery, with two PROM-AEs each (5.7%). Single studies were found in a range of conditions (n=10); please see Table 4 for the full list.

PROM AEs were measured across three main settings: inpatient, outpatient (day treatment, clinic), and home. The majority of PROM-AEs (n=20, 57.1%) were used in both outpatient and home settings. Notably, the sole PROM-AE identified that had regular use outside of a research setting was the Yellow Card Scheme (YCS).

We examined what core constructs comprised the identified PROM-AEs. Among the 35 included measures, the most commonly used construct was patient-reported severity of the AE (n=29, 82.9%). The second most common construct identified was the frequency of the AE occurrences (n=14, 40%). Tied for third most (n=2, 5.7%) were six separate constructs: the name of medicine/treatment in question, the dosage (if applicable), reason for taking the treatment, age, sex, and current patient status. The PROM-AE that encompassed the most constructs identified was the YCS. The only generic and individualized PROM-AE, the YCS provides the

patient the opportunity to select what adverse symptoms matter most and describe the context in their preferred way. Frequency was the only construct the YCS did not include. Please see Table 2 for a list of constructs corresponding to each PROM-AE, and Table 6 for a list of the proportion of each construct found in the included measurement tools**.**

The most frequently used measure was the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) (n=10, 15.6%). The PRO-CTCAE was developed by working with the initial version of the CTCAE survey created by the National Cancer Institute. Initially intended for physician use, the CTCAE terminology was modified to substitute lay terms into the PRO-CTCAE to allow for use by patients (19). Once the preliminary version was created, a large scale validation study was conducted in the US. In total, 975 adult cancer patients in various select major cancer care centers were enrolled, and completed the questionnaire on tablet computers (20). The study demonstrated validity, reliability and responsiveness for the PRO-CTCAE (20); detailed findings are presented elsewhere (Chapter 3).

The structure of the PRO-CTCAE spans a total of 124 items across different areas of potential AE outcomes. Depending on the cancer type, the patient will be directed to answer questions about disorders within the realms of cardiac, ear, eye, gastrointestinal system, general pain/swelling, appetite, musculoskeletal, nervous system, psychiatric, renal/urinary, reproductive system, respiratory system, skin/subcutaneous tissue, and the vascular system (20). Researchers using the tool are able to select from the list of items what body systems are relevant to their investigation (20). As such, the PRO-CTCAE is a disease-specific standardized instrument. The timeframe used for asking patients about their selected AE(s) is within the past week; the instrument also asks about the frequency, severity, and level of interference with a patient's daily

life (19). The PRO-CTCAE does not ask about the following constructs: name, reason for, and dosage of treatment(s), as well as age, sex, weight, prescribing doctor, and current status of the patient.

DISCUSSION

To our knowledge, this is the first systematic review to compile all PROM-AEs reported in the peer-reviewed medical literature. There were four other systematic reviews retrieved during screening that had relevant studies for inclusion. The foci in these reviews were on drug-related harms and their effect on HRQOL (21-23), and dermatological AEs during cancer treatment (24). This review does not duplicate these works; we sought to establish a complete representation of the resources available in the peer-reviewed medical literature for patient AE reporting.

The most cited PROM-AE was a disease-specific standardized tool developed in 2014 for patients with cancer, the PRO-CTCAE (19). The only PROM-AE identified for regular use outside a research setting was the YCS, a generic individualized tool that included the most constructs found among the PROM-AEs except frequency (24). The vast majority of PROM types were, like the PRO-CTCAE, disease-specific and standardized. This indicates that research has focused on synthesizing data between patients under a given condition. Adding all cancers together amounts to 60% of all PROM-AEs. Most PROM-AEs were utilized in both outpatient and home settings. This is likely due to trial methods having initial PROM-AE administration in an outpatient clinic, with subsequent follow-up in the clinic or at home.

There is untapped potential for research and clinical application of PROM-AE. While AE reporting has become a part of the standard reporting methodology in health research fields such

as cancer and psychiatry, this remains the exception, not the norm, for other areas of clinical research (15, 25). Treatment-related AEs must be sought and clearly reported in all areas of clinical research. PROM-AEs provide the opportunity to collect patient information on AEs on a larger and more systematic scale than obtaining AE information without them. To our knowledge, there is no PROM-AE required by regulatory agencies for AE reporting in clinical trials. However, in Canada and the UK, national regulatory agencies collect information from patients about AEs experienced in routine clinical care through the use of PROM-AE (26,27). Patient experience is useful and important, and should be accounted for in treatment decisions. PROM-AEs may help overcome existing knowledge gaps while promoting awareness of patient experience. They also may play an important role in clinical practice as well as in clinical research, as seen with PROM use in the UK (28). Routine use of PROM-AEs in clinical care may encourage patient feedback, enhancing communication, and improving patient care. Accurate knowledge of treatment-related harms is essential to patient safety.

STRENGTHS AND LIMITATIONS

This systematic review brought together studies that used PROM-AE tools in various areas of research and clinical care. Strengths of this study include the use of systematic review search methods using PRISMA guidelines (17) with inclusive search terms to identify all relevant peer-reviewed literature. The marked variation in terminology for harms was accepted, as authors tend to use a range of terms to describe harms. Limitations included the search being restricted to English language only for feasibility reasons. In addition, we collected only PROM-AEs in the published literature. We made this restriction to identify current tools in the published literature. Therefore, PROM-AEs that have either undergone validation, used in the field, or have published protocols of development, would be retrieved.

CONCLUSION

A range of PROM-AEs have been developed, most commonly for use in cancer research. Given the number of PROM-AEs, we recommend potential development of a core outcome set to reduce variability and ensure all key constructs are measured consistently between studies. Given the relative lack of harms reporting at present in clinical research and clinical care, and the considerable risks posed by health interventions, further investigation of PROM-AE to enhance harms identification and reporting is both urgent and important.

REFERENCES

1. Sloan JA, Berk L, Roscoe J, Fisch MJ, Shaw EG, Wyatt G, et al. Integrating patient-reported outcomes into cancer symptom management clinical trials supported by the National Cancer Institute–sponsored clinical trials networks. Journal of Clinical Oncology. 2007;25(32):5070-7.
2. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-reported outcomes (PROs): putting the patient perspective in patient-centered outcomes research. Medical care. 2013;51(8 0 3):S73.
3. Administration UFaD. Guidance for industry: Patient-reported outcome measures—Use in medical product development to support labeling claims. Federal Register. 2009;235(74):65132-3.
4. Frank L, Basch E, Selby JV. The PCORI perspective on patient-centered outcomes research. Jama. 2014;312(15):1513-4.
5. Trotti A, Colevas AD, Setser A, Basch E. Patient-reported outcomes and the evolution of adverse event reporting in oncology. Journal of Clinical Oncology. 2007;25(32):5121-7.
6. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Medical care. 2007;45(5 Suppl 1):S3.
7. Research CIoH. Strategy for Patient-Oriented Research 2016 [cited 2017 5/19/2017]. Available from: http://www.cihr-irsc.gc.ca/e/41204.html.
8. Black N. Patient reported outcome measures could help transform healthcare. BMJ (Clinical research ed). 2013;346:f167.
9. Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. bmj. 2016;352:i157.
10. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Annals of internal medicine. 1993;118(8):622-9.
11. Ioannidis JP, Evans SJ, Gøtzsche PC, O'neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Annals of internal medicine. 2004;141(10):781-8.
12. Higgins JPT, Green S. Cochrane Handbook on Systematic Reviews of Interventions 2011 [Web Page]. Available from: http://handbook.cochrane.org/.
13. Griffin FA, Resar RK. IHI Global Trigger Tool for measuring adverse events. Institute for Healthcare Improvement Innovation Series White Paper. 2009.
14. Zorzela L, Golder S, Liu Y, Pilkington K, Hartling L, Joffe A, et al. Quality of reporting in systematic reviews of adverse events: systematic review. 2014.
15. Basch E, Iasonos A, McDonough T, Barz A, Culkin A, Kris MG, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. The lancet oncology. 2006;7(11):903-9.
16. Bateman E, Keefe D, editors. Patient-reported outcomes in supportive care. Seminars in oncology; 2011: Elsevier.
17. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. International journal of surgery (London, England). 2010;8(5):336-41.
18. Golder S, Loke YK. The performance of adverse effects search filters in MEDLINE and EMBASE. Health Information & Libraries Journal. 2012;29(2):141-51.

19. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). Journal of the National Cancer Institute. 2014;106(9):dju244.
20. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, et al. Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). JAMA Oncology. 2015;1(8):1051.
21. Foster JM, van dM, Caeser M, Hannaford P. The use of questionnaires for measuring patient-reported side effects of drugs: its importance and methodological challenges. Pharmacoepidemiology and drug safety. 2008;17(3):278-96.
22. Gaertner J, Siemens W, Camilleri M, Davies A, Drossman DA, Webster LR, et al. Definitions and outcome measures of clinical trials regarding opioid-induced constipation: a systematic review. Journal of clinical gastroenterology. 2015;49(1):9.
23. Longden E, Read J. Assessing and reporting the adverse effects of antipsychotic medication: A systematic review of clinical studies, and prospective, retrospective, and cross-sectional research United States: Lippincott Williams and Wilkins; 2016 [cited 39 1]. 29]. Available from:

24. Chan A, Cameron MC, Garden B, Boers-Doets C, Schindler K, Epstein JB, et al. A systematic review of patient-reported outcome instruments of dermatologic adverse events associated with targeted cancer therapies. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer. 2015;23(8):2231.

25. Williamson PR, Altman DG, Blazeby JM, Clarke M, Gargon E. The COMET (core outcome measures in effectiveness trials) initiative. Trials. 2011;12(Suppl 1):A70.
26. Health Canada. Adverse Reaction and Medical Device Problem Reporting [Internet], Canada: Health Canada; 2016 [updated 2016 Feb 4]. Available from: https://www.canada.ca/en/health-canada/services/drugs-health-products/medeffect-canada/adverse-reaction-reporting.html#a1
27. The Medicines and Healthcare products Regulatory Agency (MHRA). About Yellow Card [Internet], United Kingdom: MHRA: 2017. Available from: https://yellowcard.mhra.gov.uk/the-yellow-card-scheme/
28. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. Bmj. 2010;340:c186.

APPENDICES

Figure 2.1: Flow Diagram of Selection of Studies in PROM-AE Systematic Review

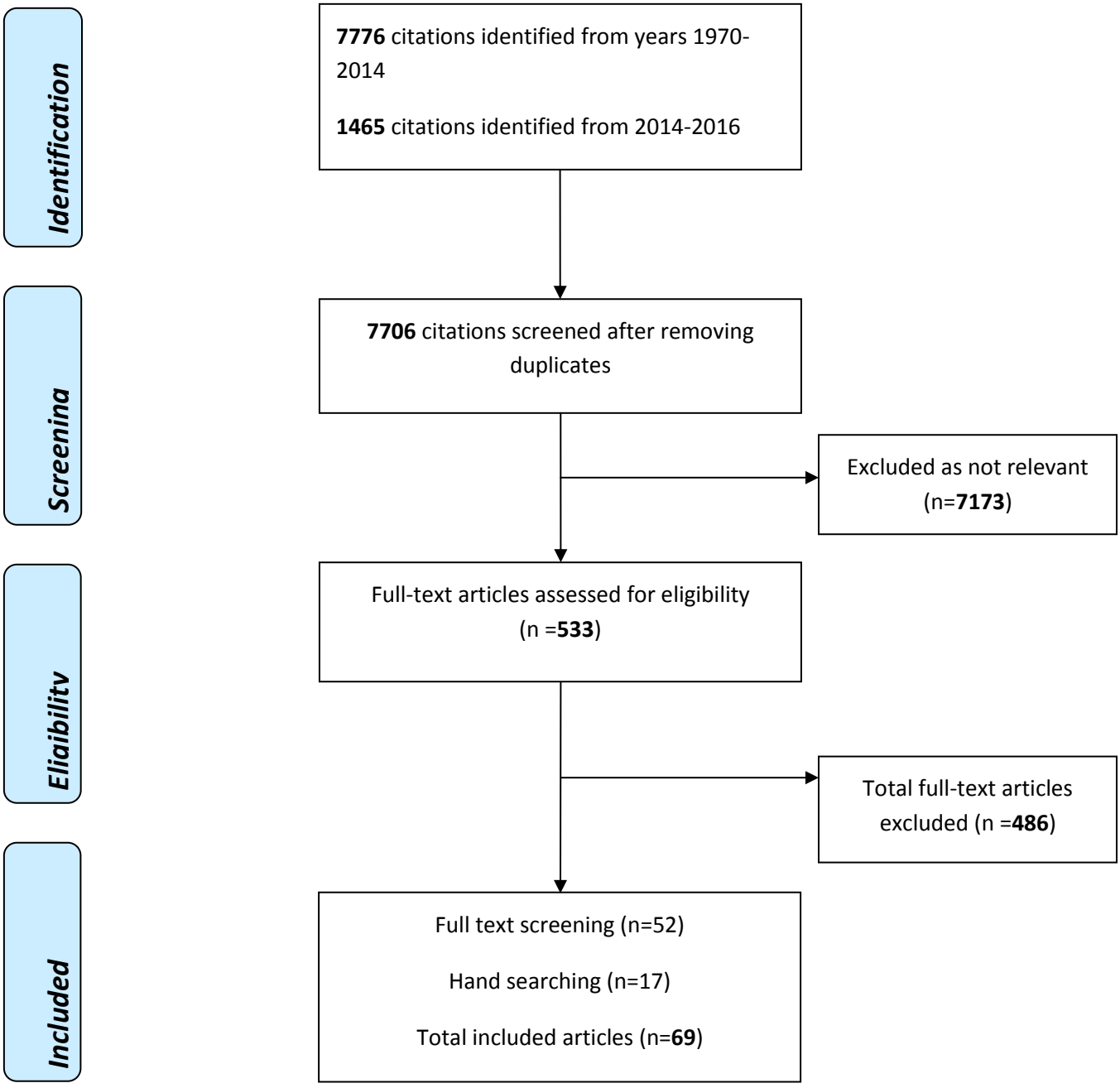


TABLE 2.1: HARMS TERMINOLOGY

Term	Definition	Reference
Harm	The totality of possible adverse consequences of an intervention or therapy	Ioannidis JP, Evans SJ, Gøtzsche PC, O'neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Annals of internal medicine. 2004;141(10):781-8.
Adverse Effect	Harms that occur "during or after the use of a drug or other intervention and the casual relation between the intervention and the event is at least a reasonable possibility.	Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. bmj. 2016;352:i157.
Adverse Event	An unfavorable outcome that occurs during or after the use of a drug or other intervention but is not necessarily caused by it.	Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. bmj. 2016;352:i157.
Adverse Drug Reaction	Causality link to the tested intervention is well established and strong enough (sensitive and specific) to warrant attribution of the event to the intervention.	Ioannidis JP, Evans SJ, Gøtzsche PC, O'neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Annals of internal medicine. 2004;141(10):781-8.

TABLE 2.2: PROM-AE RETRIEVED REFERENCES

PROM-AE Name	References	Constructs
A-B Neuropsychological Assessment Schedule (ABNAS)	• Foster JM et al, 2008(1) • Aldenkamp AP et al, 2002(2) • Brooks J et al, 2001(3)	Severity
Approaches to Schizophrenia Communication-Self-Report Checklist (ASC-SR)	• Foster JM et al, 2008(1) • Dott SG et al, 2001(4)	Presence, Name of medicine/treatment and reason for taking, Dosage of medicine/treatment
Intra-hospital post-operative AE Survey	• Hart RA et al, 2013(5)	Severity
Barkley Stimulant Side Effect Rating Scale (BSSERS)	• Adamo N et al, 2015(6) • Barkley RA et al, 1990(7)	Severity
Cancer Symptom Experience Inventory (CSEI)	• Wagland R et al, 2015(8) • Hoffman AJ et al, 2007(9)	Frequency, Severity
C-SAS: Chemotherapy Symptom Assessment Scale	• Foster JM et al, 2008(1) • Brown V et al, 2001(10)	Severity
Checklist for Patients with Endocrine Therapy (C-PET)	• Foster JM et al, 2008(1) • Malionvsky KM et al, 2004(11)	Presence
CTC-AE Based Questionnaire	• Farnell DJJ et al, 2010(12) • Ho KF et al, 2009 (13)	Frequency, Severity
Disabilities of Arm, Shoulder and Hand Scale (DASH)	• Land SR et al, 2010(14) • Hudak PL et al, 1996(15)	Severity
The Expanded Prostate Cancer Index Composite (EPIC)	• Wei JT et al, 2000(16) • Talcott et al, 2014(17)	Frequency, Severity
Functional Assessment of Cancer Therapy: Anemia (FACT-An)	• Tefferi A et al, 2014(18)	Severity
Functional Assessment of Cancer Therapy Prostate (FACT-P)	• Henderson A et al, 2006(19) • Esper P et al, 1997(20)	Severity
Hip Osteoarthritis Outcome Score (HOOS)	• Boyce MB et al, 2015(21) • Davis AM et al, 2009(22)	Frequency, Severity
HIV Symptom Index Questionnaire (SIQ)	• Brunetta J et al, 2015(23) • Justice AC et al, 2001(24)	Frequency, Severity
Inhaled Corticosteroid Questionnaire (ICQ)	• Van der Molen T et al, 2010(25) • Foster JM et al, 2006(26)	Severity
International Prostate Symptom Score (IPSS)	• Venderbos LD et al, 2015(27) • Glass AS et al, 2014(28)	Frequency
Late Effects of Normal Tissue (LENT) prostate patient questionnaire	• Farnell DJJ et al, 2010(29)	Frequency, Severity
LUNSERS: The Liverpool University Neuroleptic Side Effect Rating Scale	• Foster JM et al, 2008(1) • Day J et al, 1995(30) • Longden E at al, 2016(31)	Frequency
M.D. Anderson Dysphagia Inventory (MDADI)	• Roe JWG et al, 2014(32) • Chen AY et al, 2001(33)	Severity
Memorial Symptom Assessment Scale (MSAS)	• Wagland R et al, 2015(8) • Portenoy RK et al, 1994(34) • Traeger L et al, 2015 (35)	Frequency, Severity
MPQ: Medication Problems Questionnaire	• Foster JM et al, 2008(1) • Mohr DC et al, 1998(36)	Severity

Opioid- Related Symptom Distress Scale (OR-SDS)	• Langford RW et al, 2009(37) • Apfelbaum JL et al, 2004(38)	Frequency, Severity, Presence
Oral Mucositis Daily Questionnaire (OMDQ)	• Lucchese A et al, 2016(39) • Elting LS et al, 2008(40) • Stiff PJ et al, 2006(41)	Severity, Frequency
Oral Mucositis Weekly Questionnaire -Head and Neck Cancer (OMWQ-HN)	• Bateman E & Keefe D, 2011(42) • Epstein JB et al, 2007(42)	Severity
Patient Assessment of Constipation Symptoms (PAC-SYM)	• Gaertner J et al, 2015(43) • Slappendel R et al, 2006(44)	Severity
Patient Neurotoxicity Questionnaire (PNQ)	• Bennett BK et al, 2012(45) • Shimozuma K et al, 2009(46) • Extra?	Severity, Frequency,
Patient Reported Chemotherapy Indicators of Symptoms and Experience (PR-CISE)	• Wagland R et al, 2015(8) • Armes J et al, 2014(47)	Severity
PCM 2.0 Review of Systems Survey	• Abernathy AP et al, 2010(48) • Fortner R et al, 2003(49)	Severity
PRO-CTCAE	• Hay AL et al, 2014(50) • Banerjee AK et al, 2013(51) • Basch E et al, 2011(52) • Basch E et al, 2014(53) • Basch E et al, 2014(54) • Bennett AV et al, 2016(55) • Bruner DW et al, 2011(56) • Dueck AC et al, 2011(57) • Siddiqui F et al, 2014(58) • Smith TG et al, 2016(59)	Severity, Frequency, Interference, Presence,
Rotterdam Symptom Checklist	• Wagland R et al, 2015(8) • De Jong N et al, 2004(60)	Severity
SES-HP: Side Effect Scoring System for Helicobacter Pylori treatment	• Foster JM et al, 2008(1) • Armuzzi A et al, 2001(61)	Severity
Severity of Dyspepsia Assessment (SODA)	• Vardi M et al, 2015(62) • Rabeneck L et al, 2001(63)	Severity,
Skindex-16 toxicity measurement instrument	• Chan A et al, 2015(64) • Neben-Wittich MA et al, 2011(65) • Chren MM et al, 2001(66)	Frequency
Symptom Tracking and Reporting (STAR) online reporting system	• Judson TJ et al, 2013(67) • Basch E et al, 2007(68)	Age, Sex, Education, cancer stage,
Yellow Card Scheme Questionnaire	• Anderson C et al, 2011(69)	Age, Sex, Weight, Severity, Frequency, Current patient status, Name of medicine/treatment and reason for taking, Dosage of medicine/treatment, Doctor's name,

REFERENCE LIST OF INCLUDED ARTICLES

1. Foster J, van der Molen T, Caeser M, Hannaford P. The use of questionnaires for measuring patient-reported side effects of drugs: its importance and methodological challenges. Pharmacoepidemiology and drug safety. 2008;17(3):278.
2. Aldenkamp A, Van Meel H, Baker G, Brooks J, Hendriks M. The A–B neuropsychological assessment schedule (ABNAS): the relationship between patient-perceived drug related cognitive impairment and results of neuropsychological tests. Seizure. 2002;11(4):231-7.
3. Brooks J, Baker GA, Aldenkamp AP. The A–B neuropsychological assessment schedule (ABNAS): the further refinement of a patient-based scale of patient-perceived cognitive functioning. Epilepsy research. 2001;43(3):227-37.
4. Dott SG, Weiden P, Hopwood P, Awad AG, Hellewell JS, Knesevich J, et al. An innovative approach to clinical communication in schizophrenia: the approaches to schizophrenia communication checklists. CNS spectrums. 2001;6(04):333-8.
5. Hart RA, Cabalo A, Bess S, Akbarnia BA, Boachie-Adjei O, Burton D, et al. Comparison of patient and surgeon perceptions of adverse events after adult spinal deformity surgery. Spine. 2013;38(9):732-6.
6. Adamo N, Seth S, Coghill D. Pharmacological treatment of attention-deficit/hyperactivity disorder: Assessing outcomes United Kingdom: Expert Reviews Ltd; 2015 [cited 8 4]. QOL tool].

7. Barkley RA, McMurray MB, Edelbrock CS, Robbins K. Side effects of metlyiphenidate in children with attention deficit hyperactivity disorder: a systemic, placebo-controlled evaluation. Pediatrics. 1990;86(2):184-92.
8. Wagland R, Richardson A, Armes J, Hankins M, Lennan E, Griffiths P. Treatment-related problems experienced by cancer patients undergoing chemotherapy: a scoping review. European Journal of Cancer Care. 2015;24(5):605.
9. Hoffman AJ, Given BA, Von Eye A, Gift AG, Given CW, editors. Relationships among pain, fatigue, insomnia, and gender in persons with lung cancer. Oncology nursing forum; 2007.
10. Brown V, Sitzia J, Richardson A, Hughes J, Hannon H, Oakley C. The development of the Chemotherapy Symptom Assessment Scale (C-SAS): a scale for the routine clinical assessment of the symptom experiences of patients receiving cytotoxic chemotherapy. International journal of nursing studies. 2001;38(5):497-510.
11. Malinovszky K, Cameron D, Douglas S, Love C, Leonard T, Dixon J, et al. Breast cancer patients' experiences on endocrine therapy: monitoring with a checklist for patients on endocrine therapy (C-PET). The Breast. 2004;13(5):363-8.
12. Farnell DJJ, Routledge J, Hannon R, Logue JP, Cowan RA, Wylie JP, et al. Efficacy of data capture for patient-reported toxicity following radiotherapy for prostate or cervical cancer. European journal of cancer. 2010;46(3):534.
13. Ho KF, Farnell DJ, Routledge JA, Burns MP, Sykes AJ, Slevin NJ, et al. Developing a CTCAEs patient questionnaire for late toxicity after head and neck radiotherapy. European Journal of Cancer. 2009;45(11):1992-8.
14. Land SR, Kopec JA, Julian TB, Brown AM, Anderson SJ, Krag DN, et al. Patient-reported outcomes in sentinel node–negative adjuvant breast cancer patients receiving sentinel-node biopsy or axillary dissection: National Surgical Adjuvant Breast and Bowel Project Phase III Protocol B-32. Journal of Clinical Oncology. 2010;28(25):3929-36.
15. Hudak PL, Amadio PC, Bombardier C, Beaton D, Cole D, Davis A, et al. Development of an upper extremity outcome measure: the DASH (Disabilities of the Arm, Shoulder, and Hand). American journal of industrial medicine. 1996;29(6):602-8.

16. Wei JT, Dunn RL, Litwin MS, Sandler HM, Sanda MG. Development and validation of the expanded prostate cancer index composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer. Urology. 2000;56(6):899-905.
17. Talcott JA, Manola J, Chen RC, Clark JA, Kaplan I, D'amico AV, et al. Using patient-reported outcomes to assess and improve prostate cancer brachytherapy. BJU international. 2014;114(4):511-6.
18. Tefferi A, Hudgens S, Mesa R, Gale RP, Verstovsek S, Passamonti F, et al. Use of the Functional Assessment of Cancer Therapy--anemia in persons with myeloproliferative neoplasm-associated myelofibrosis and anemia. Clinical therapeutics. 2014;36(4):560.
19. Henderson A, Andreyev H, Stephens R, Dearnaley D. Patient and physician reporting of symptoms and health-related quality of life in trials of treatment for early prostate cancer: considerations for future studies. Clinical Oncology. 2006;18(10):735-43.
20. Esper P, Mo F, Chodak G, Sinner M, Cella D, Pienta KJ. Measuring quality of life in men with prostate cancer using the functional assessment of cancer therapy-prostate instrument. Urology. 1997;50(6):920-8.
21. Boyce MB, Browne JP. The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient-reported outcome measures—results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study. BMJ open. 2015;5(7):e008325.
22. Davis AM, Perruccio AV, Canizares M, Hawker GA, Roos EM, Maillefert J-F, et al. Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. Osteoarthritis and Cartilage. 2009;17(7):843-7.
23. Brunetta J, Moreno GS, Antinori A, Yeni P, Wade B, Johnson M, et al. Patient-Reported Outcomes After a Switch to a Single-Tablet Regimen of Rilpivirine, Emtricitabine, and Tenofovir DF in HIV-1-Positive, Virologically Suppressed Individuals: Additional Findings From a Randomized, Open-Label, 48-Week Trial Switzerland: Springer International Publishing; 2015 [cited 8 3]. 257].

24. Justice A, Holmes W, Gifford A, Rabeneck L, Zackin R, Sinclair G, et al. Development and validation of a self-completed HIV symptom index. Journal of Clinical Epidemiology. 2001;54(12):S77-S90.
25. van der Molen T, Foster JM, Caeser M, Müller T, Postma DS. Difference between patient-reported side effects of ciclesonide versus fluticasone propionate. Respiratory medicine. 2010;104(12):1825-33.
26. Foster JM, Van Sonderen E, Lee AJ, Sanderman R, Dijkstra A, Postma DS, et al. A self-rating scale for patient-perceived side effects of inhaled corticosteroids. Respiratory research. 2006;7(1):1.
27. Venderbos LDF, Bangma CH, Korfage IJ. Disease-specific and patient-reported outcomes under active surveillance. Current opinion in urology. 2015;25(3):267.
28. Glass AS, Hilton JF, Cowan JE, Washington SL, Carroll PR. Serial prostate biopsy and risk of lower urinary tract symptoms: results from a large, single-institution active surveillance cohort. Urology. 2014;83(1):33-9.
29. Farnell DJ, Mandall P, Anandadas C, Routledge J, Burns MP, Logue JP, et al. Development of a patient-reported questionnaire for collecting toxicity data following prostate brachytherapy. Radiotherapy and Oncology. 2010;97(1):136-42.
30. Day JC, Wood G, Dewey M, Bentall RP. A self-rating scale for measuring neuroleptic side-effects. Validation in a group of schizophrenic patients. The British Journal of Psychiatry. 1995;166(5):650-3.
31. Longden E, Read J. Assessing and Reporting the Adverse Effects of Antipsychotic Medication: A Systematic Review of Clinical Studies, and Prospective, Retrospective, and Cross-Sectional Research. Clinical neuropharmacology. 2016;39(1):29-39.

32. Roe JWG, Drinnan MJ, Carding PN, Harrington KJ, Nutting CM. Patient-reported outcomes following parotid-sparing intensity-modulated radiotherapy for head and neck cancer. How important is dysphagia? United Kingdom: Elsevier Ltd; 2014 [cited 50 12]. QOL tool].

33. Chen AY, Frankowski R, Bishop-Leone J, Hebert T, Leyk S, Lewin J, et al. The development and validation of a dysphagia-specific quality-of-life questionnaire for patients with head and neck cancer: the MD Anderson dysphagia inventory. Archives of Otolaryngology–Head & Neck Surgery. 2001;127(7):870-6.
34. Portenoy RK, Thaler HT, Kornblith AB, Lepore JM, Friedlander-Klar H, Kiyasu E, et al. The Memorial Symptom Assessment Scale: an instrument for the evaluation of symptom prevalence, characteristics and distress. European Journal of Cancer. 1994;30(9):1326-36.
35. Traeger L, McDonnell TM, McCarty CE, Greer JA, El-Jawahri A, Temel JS. Nursing intervention to enhance outpatient chemotherapy symptom management: Patient-reported outcomes of a randomized controlled trial. Cancer. 2015;121(21):3905-13.
36. Mohr DC, Likosky W, Boudewyn AC, Marietta P, Dwyer P, Van Der Wende J, et al. Side effect profile and adherence to in the treatment of Multiple Sclerosis with interferon beta-la. Multiple Sclerosis. 1998;4(6):487-9.
37. Langford RM, Joshi GP, Gan TJ, Mattera MS, Chen W-H, Revicki DA, et al. Reduction in opioid-related adverse events and improvement in function with parecoxib followed by valdecoxib treatment after Non-cardiac surgery. Clinical drug investigation. 2009;29(9):577-90.
38. Apfelbaum JL, Gan TJ, Zhao S, Hanna DB, Chen C. Reliability and validity of the perioperative opioid-related symptom distress scale. Anesthesia & Analgesia. 2004;99(3):699-709.
39. Lucchese A, Matarese G, Ghislanzoni LH, Gastaldi G, Manuelli M, Gherlone E. Efficacy and effects of palifermin for the treatment of oral mucositis in patients affected by acute lymphoblastic leukemia United Kingdom: Taylor and Francis Ltd; 2016 [cited 57 4]. 820].
40. Elting LS, Keefe DM, Sonis ST, Garden AS, Spijkervet F, Barasch A, et al. Patient-reported measurements of oral mucositis in head and neck cancer patients treated with radiotherapy with or without chemotherapy. Cancer. 2008;113(10):2704-13.
41. Stiff P, Erder H, Bensinger W, Emmanouilides C, Gentile T, Isitt J, et al. Reliability and validity of a patient self-administered daily questionnaire to assess impact of oral mucositis (OM) on pain and daily functioning in patients undergoing autologous hematopoietic stem cell transplantation (HSCT). Bone marrow transplantation. 2006;37(4):393-401.
42. Bateman E, Keefe D, editors. Patient-reported outcomes in supportive care. Seminars in oncology; 2011: Elsevier.
43. Gaertner J, Siemens W, Camilleri M, Davies A, Drossman DA, Webster LR, et al. Definitions and outcome measures of clinical trials regarding opioid-induced constipation: a systematic review. Journal of clinical gastroenterology. 2015;49(1):9.
44. Slappendel R, Simpson K, Dubois D, Keininger DL. Validation of the PAC-SYM questionnaire for opioid-induced constipation in patients with chronic low back pain. European Journal of Pain. 2006;10(3):209-17.
45. Bennett BK, Park SB, Lin CS-Y, Friedlander ML, Kiernan MC, Goldstein D. Impact of oxaliplatin-induced neuropathy: a patient perspective. Supportive Care in Cancer. 2012;20(11):2959-67.
46. Shimozuma K, Ohashi Y, Takeuchi A, Aranishi T, Morita S, Kuroi K, et al. Feasibility and validity of the Patient Neurotoxicity Questionnaire during taxane chemotherapy in a phase III randomized trial in patients with breast cancer: N-SAS BC 02. Supportive Care in Cancer. 2009;17(12):1483-91.
47. Armes J, Wagland R, Finnegan-John J, Richardson A, Corner J, Griffiths P. Development and testing of the patient-reported chemotherapy indicators of symptoms and experience: patient-reported

outcome and process indicators sensitive to the quality of nursing care in ambulatory chemotherapy settings. Cancer nursing. 2014;37(3):E52-E60.
48. Abernethy AP, Zafar SY, Uronis H, Wheeler JL, Coan A, Rowe K, et al. Validation of the Patient Care Monitor (Version 2.0): a review of system assessment instrument for cancer patients. Journal of pain and symptom management. 2010;40(4):545-58.
49. Fortner B, Okon T, Schwartzberg L, Tauer K, Houts AC. The Cancer Care Monitor: psychometric content evaluation and pilot testing of a computer administered system for symptom screening and quality of life in adult cancer patients. Journal of pain and symptom management. 2003;26(6):1077-92.
50. Hay JL, Atkinson TM, Reeve BB, Mitchell SA, Mendoza TR, Willis G, et al. Cognitive interviewing of the US National Cancer Institute's patient-reported outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Quality of Life Research. 2014;23(1):257-69.
51. Banerjee AK, Okun S, Edwards IR, Wicks P, Smith MY, Mayall SJ, et al. Patient-reported outcome measures in safety event reporting: PROSPER consortium guidance. Drug safety. 2013;36(12):1129-49.
52. Basch E, Bennett A, Pietanza MC. Use of patient-reported outcomes to improve the predictive accuracy of clinician-reported adverse events. Journal of the National Cancer Institute. 2011;103(24):1808-10.
53. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). Journal of the National Cancer Institute. 2014;106(9):dju244.
54. Basch E. The rationale for collecting patient-reported symptoms during routine chemotherapy. American Society of Clinical Oncology educational book / ASCOAmerican Society of Clinical OncologyMeeting. 2014:161.
55. Bennett AV, Dueck AC, Mitchell SA, Mendoza TR, Reeve BB, Atkinson TM, et al. Mode equivalence and acceptability of tablet computer-, interactive voice response system-, and paper-based administration of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Health & Quality of Life Outcomes. 2016;14(1):24.
56. Bruner DW, Hanisch LJ, Reeve BB, Trotti AM, Schrag D, Sit L, et al. Stakeholder perspectives on implementing the National Cancer Institute's patient-reported outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Translational Behavioral Medicine. 2011;1(1):110.
57. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, et al. Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). JAMA Oncology. 2015;1(8):1051.
58. Siddiqui F, Liu AK, Watkins-Bruner D, Movsas B. Patient-reported outcomes and survivorship in radiation oncology: overcoming the cons. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2014;32(26):2920.
59. Smith TG, Castro KM, Troeschel AN, Arora NK, Lipscomb J, Jones SM, et al. The rationale for patient-reported outcomes surveillance in cancer and a reproducible method for achieving it United States: John Wiley and Sons Inc. (P.O.Box 18667 Newark NJ 07191-8667 United States); 2016 [cited 122 3]. 344].

60. De Jong N, Candel M, Schouten H, Abu-Saad HH, Courtens A. Prevalence and course of fatigue in breast cancer patients receiving adjuvant chemotherapy. Annals of oncology. 2004;15(6):896-905.
61. Armuzzi A, Cremonini F, Bartolozzi F, Canducci F, Candelli M, Ojetti V, et al. The effect of oral administration of Lactobacillus GG on antibiotic-associated gastrointestinal side-effects during Helicobacter pylori eradication therapy. Alimentary pharmacology & therapeutics. 2001;15(2):163-9.
62. Vardi M, Cryer BL, Cohen M, Lanas A, Schnitzer TJ, Lapuerta P, et al. The effects of proton pump inhibition on patient-reported severity of dyspepsia when receiving dual anti-platelet therapy with

clopidogrel and low-dose aspirin: analysis from the Clopidogrel and the Optimization of Gastrointestinal Events Trial. Alimentary Pharmacology & Therapeutics. 2015;42(3):365.
63. Rabeneck L, Cook KF, Wristers K, Souchek J, Menke T, Wray NP. SODA (severity of dyspepsia assessment): a new effective outcome measure for dyspepsia-related health. Journal of clinical epidemiology. 2001;54(8):755-65.
64. Chan A, Cameron MC, Garden B, Boers-Doets C, Schindler K, Epstein JB, et al. A systematic review of patient-reported outcome instruments of dermatologic adverse events associated with targeted cancer therapies. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer. 2015;23(8):2231.
65. Neben-Wittich MA, Atherton PJ, Schwartz DJ, Sloan JA, Griffin PC, Deming RL, et al. Comparison of provider-assessed and patient-reported outcome measures of acute skin toxicity during a Phase III trial of mometasone cream versus placebo during breast radiotherapy: the North Central Cancer Treatment Group (N06C4). International Journal of Radiation Oncology* Biology* Physics. 2011;81(2):397-402.
66. Chren M-M, Lasek RJ, Sahay AP, Sands LP. Measurement properties of Skindex-16: a brief quality-of-life measure for patients with skin diseases. Journal of cutaneous medicine and surgery. 2001;5(2):105-10.
67. Judson TJ, Bennett AV, Rogak LJ, Sit L, Barz A, Kris MG, et al. Feasibility of long-term patient self-reporting of toxicities from home via the Internet during routine chemotherapy. Journal of Clinical Oncology. 2013;31(20):2580-5.
68. Basch E, Artz D, Iasonos A, Speakman J, Shannon K, Lin K, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. Journal of the American Medical Informatics Association. 2007;14(3):264-8.
69. Anderson C, Krska J, Murphy E, Avery A. The importance of direct patient reporting of suspected adverse drug reactions: a patient perspective. British journal of clinical pharmacology. 2011;72(5):806-22.

TABLE 2.3A: CHARACTERISTICS OF CONDITION-SPECIFIC PROM-AEs REPORTED IN MEDICAL LITERATURE

Elements	Intervention Used	Patient age group/Disorder	Sex	Setting of use	Question Type	Local/Constitutional	Frequency Used (y/n)	Severity Used (y/n)	AE symptoms listed
A-B Neuropsychological Assessment Schedule (ABNAS)	Antiepileptic drugs	Adult Epilepsy	M/F	Outpatient; Home	Standardized	Constitutional	n	y	Decreased mood, problems with memory/ concentration, fatigue, decreased motor ability, speech trouble, slower reaction time, lack of stamina
Approaches to Schizophrenia Communication-Self-Report Checklist (ASC-SR)	Antipsychotic therapy	Adult Schizophrenia	M/F	Inpatient; Outpatient	Standardized	Constitutional	n	n	Fatigue, slowness/difficulties with movement, restlessness, muscle stiffness, excessive drooling, muscles trembling/shaking, weight gain/loss, excessive/difficulty sleeping, insomnia, dry mouth, blurry vision, constipation, trouble urinating, problems with memory problems with memory/ concentration
Barkley Stimulant Side Effect Rating Scale (BSSERS)	ADHD Medication	Adult ADHD	M/F	Outpatient	Standardized	Constitutional	y	y	Decreased appetite, Insomnia, Stomachaches, Headaches, Prone to crying, Tics/nervous movements, Dizziness, Drowsiness, Nail biting, Talks less, Anxious, Disinterested in others, Euphoria, Irritable Nightmares, Sadness, Staring
Cancer Symptom Experience Inventory	Chemotherapy	Adult Cancer (Lung)	M/F	Outpatient	Standardized	Constitutional	y	y	Fatigue, Pain, Nausea, Constipation, Insomnia, Poor appetite, Cough, Dry mouth, Difficulty breathing, Diarrhea, Difficulty concentrating, Vomiting, Fever, Mouth sores
Checklist for Patients with Endocrine Therapy (C-PET)	Endocrine therapy	Adult Cancer (Breast)	F	Outpatient; Home	Standardized	Genitourinary	n	n	Hot flashes, increased sweating, weight gain, nausea, fatigue, fluid retention, irritability, decreased sex drive, skin rash, shortness of breath, bleeding/discharge/dryness of vagina
Chemotherapy Symptom Assessment Scale (CSAS)	Chemotherapy	Adult Cancer (General)	M/F	Outpatient; Home	Standardized	Constitutional	y	y	Nausea, vomiting, sore/sensitive throat, hair loss, fatigue, depression, diarrhea, pain, constipation, anxiety, numbness in hands/feet, shortness of breath, feeling fearful, nosebleed, headache, sore/scratchy/dry eyes, feeling angry/aggressive, weight gain/loss, vision changes, indigestion, irritability, (bleeding/spotting, hot flashes, irregular periods (F)), appetite changes, mood swings, dry/itchy/inflamed skin, problems with memory/ concentration, sore/achy joints, change in taste, restlessness, change in skin sensitivity, frequent urination, decreased sex drive, change in sense of smell.

CTC-AE Based Questionnaire	Chemotherapy, radiotherapy	Adult Cancer (Prostate & Gynecological)	M /F	Inpatient	Standardized	Constitutional	y	Y	Males: presence/duration of discomfort in rectum/bowel, bladder/urethra, ureter/kidney, deficiencies in sexual function; females: presence/duration of discomfort in ovary/reproductive system, rectum/bowel, bladder/urethra, vagina, deficiencies in sexual function.
Disabilities of Arm, Shoulder and Hand Scale (DASH)	Surgery	Adult Cancer (Head & Neck)	M /F	Inpatient: Home	Standardized	Musculoskeletal	n	y	Difficulty with sexual activity, pain, pain when moving, tingling, weakness, stiffness, difficulty sleeping
Expanded Prostate Cancer Index Composite (EPIC)	Brachytherapy, Androgen Deprivation Therapy	Adult Cancer (Prostate)	M	Outpatient; Home	Standardized	Genitourinary, Gastrointestinal	y	Y	Leaked urine, Blood in urine, Pain with urination, Urinary control, Stool/fecal leakage, Bloody stools, Pain with bowel movement, Abdominal/pelvic/rectal pain
Functional Assessment of Cancer Therapy: Prostate (FACT-P)	Cancer treatment (Various)	Adult Cancer (Prostate)	M	Outpatient; Home	Standardized	Genitourinary, Gastrointestinal	y	Y	Weight loss, appetite, aches/pains, difficulty with bowel movement, difficulty with/frequent urination, difficulty with erections
Functional Assessment of Cancer Therapy: Anemia (FACT-An)	Cancer Treatment (Various)	Adult Cancer (General)		Outpatient; Home	Standardized	Constitutional	y	y	Fatigue, tiredness, energy, walking trouble, dizziness, headaches, shortness of breath, pain in chest
Hip Osteoarthritis Outcome Score (HOOS)	Surgery (Hip Replacement)	Adult Arthritis	M /F	Outpatient; Home	Standardized	Musculoskeletal	y	y	Presence of grinding, difficulty moving legs, severity of hip joint stiffness, frequency of hip pain, severity of hip pain, difficulty with everyday tasks /exercise due to hip
HIV Symptom Index Questionnaire (SIQ)	Retroviral treatment	Adult HIV	M /F	Outpatient	Standardized	Constitutional	n	N	Fatigue, fever, chills, sweats, dizziness, hand/foot pain, memory loss, nausea, diarrhea, sadness, anxiety, sleep trouble, skin problems, cough/shortness of breath, headache, appetite loss, bloating/gas, muscle/joint pain, sex problems, body image, weight loss, hair loss
Inhaled Corticosteroid Questionnaire (ICQ)	Drug treatment	Adult Asthma	M /F	Outpatient; Home	Standardized	Respiratory	y	y	Voice changes (hoarse/rough/breaking), changes in voice sensation, lack singing ability, loss of speech volume, fatigue/pain when talking, coughing (phlegm, thick mucus), sore/dry/itchy throat, bad taste, change/loss in taste ability, loss of appetite, swollen face, dry/thin skin, bruising, brittle/breaking nails, hair loss, decreased vision, increased sweating, tooth decay/staining, difficulty sleeping, fatigue, dry eyes
International Prostate Symptom Score (IPSS)	Active surveillance	Adult Cancer (Prostate)	M	Outpatient; Home	Standardized	Genitourinary	y	y	Incomplete emptying, frequency, intermittency, urgency, weak stream, straining, nocturia

Intra-hospital post-operative AE Survey	Surgery	Adult Post-Surgery	M /F	Inpatient	Standardized	Constitutional	n	y	Inflammation, infection (urinary tract, sepsis, wound), infusion/drug error (time, dose, omission), error in treating wrong side of body, wrong treatment, fall in hospital
Late Effects of Normal Tissue prostate patient questionnaire (LENT)	Chemotherapy	Adult Cancer (Prostate)	M	Outpatient; Home	Standardized	Genitourinary, Gastrointestinal	y	y	Concerning bowel movements: Pain, increased/decreased activity, diarrhea, difficulty with control, bleeding, constipation, black/sticky/slimy motions; concerning urination: pain, increased/decreased desire/frequency,
Liverpool University Neuroleptic Side Effect Rating Scale (LUNSERS)	Drug treatment	Adult Schizophrenia	M /F	Outpatient; Home	Standardized	Constitutional	y	y	Skin rash, wakefulness difficulty, runny nose, increased dreaming, headache, dry mouth, swollen/tender chest, chilblains, problems with concentration, constipation, hair loss, darker urine, period problems (decreased) (F), tension, dizziness, nausea, increased/decreased sex drive, fatigue, muscle stiffness, heart palpitations, problems with memory, weight loss, weak fingernails, depression, increased sweating, mouth ulcers, slowing of movements, greasy skin, excessive sleeping, difficulty urinating, flushed face, muscle spasms, increased sun sensitivity, diarrhea, drooling mouth, blurry vision, weight gain, restlessness, difficulty sleeping, shakiness, pins and needles sensation, painful joints, decreased sex drive, new/unusual skin marks, body parts moving involuntarily (e.g. foot moving up and down), itchy skin, frequent urination.
M.D. Anderson Dysphagia Inventory (MDADI)	Radiotherapy	Adult Cancer (Head & Neck)	M /F	Inpatient	Standardized	Constitutional	n	n	Severity of swallowing difficulty, coughing when drinking liquids, weight issues
Medication Problems Questionnaire (MPQ)	Drug Treatment	Adult (General)	M /F	Inpatient	Standardized	Constitutional	y	y	Flu-like symptoms, muscle aches, chills, loss of feeling/numbness, tingling/loss of control in limb(s), depression, fever, headache, dizziness, nausea, fatigue, redness/ pain at injection site, sleep difficulties, upper respiratory tract infection, sinus inflammation, indigestion, pain, loss of strength, diarrhea
Opioid- Related Symptom Distress Scale (OR-SDS)	Post-surgery Opioid Treatment	Adult Post-Surgery	M /F	Inpatient: Home	Standardized	Constitutional	y	y	Fatigue, drowsiness, problems with concentration, confusion, nausea, dizziness, constipation, itching, difficulty with urination, retching/vomiting

Oral Mucositis Daily Questionnaire (OMDQ)	Drug Treatment	Adult Cancer (Head & Neck)	M /F	Inpatient; Home	Standardized	Gastrointestinal	y	y	Mouth/throat soreness, diarrhea, nausea, vomiting
Oral Mucositis Weekly Questionnaire - Head and Neck Cancer (OMWQ-HN)	Radiotherapy	Adult Cancer (Head & Neck)	M /F	Outpatient; Home	Standardized	Gastrointestinal	y	y	Mouth/throat soreness, diarrhea, nausea, vomiting
Patient Assessment of Constipation Symptoms (PAC-SYM)	Opioid treatment	Adult Chronic Low Back Pain (CBLP)	M /F	Outpatient	Standardized	Gastrointestinal	y	y	Discomfort in stomach, pain in your stomach, bloating in your stomach, stomach cramps, painful bowel movements, rectal burning during or after a bowel movement, rectal bleeding or tearing during or after a bowel movement, incomplete bowel movement, bowel movements that were too hard, bowel movements that were too small, straining or squeezing to try to pass bowel movements, feeling like you had to pass a bowel movement but you could not (‘‘false alarm’’)
Patient Neurotoxicity Questionnaire (PNQ)	Chemotherapy	Adult Cancer (General)	M /F	Outpatient; Home	Standardized	Neurological	n	y	Numbness, pain, tingling, weakness in hands/arms/legs/feet
Patient Reported Chemotherapy Indicators of Symptoms and Experience (PR-CISE)	Chemotherapy	Adult Cancer (General)	M /F	Outpatient	Standardized	Constitutional	n	y	Nausea, vomiting, IV line pain and irritation, mouth problems, weakness, signs of infection, tiredness, feeling low/depressed
PCM 2.0 Review of Systems Survey	Chemotherapy	Adult Cancer (General)	M /F	Outpatient; Home	Standardized	Constitutional	y	y	Fatigue/tiredness/weakness, trouble breathing, coughing, rash/dry/itchy skin, problems with focus/concentration, difficulty sleeping, problems with urination, daytime sleepiness, numbness/tingling, pain, decreased sex drive, sweating, weight gain/loss, change in taste ability, hair loss, appetite changes, nausea, vomiting, sore throat/trouble swallowing, dry mouth, diarrhea, constipation, acute distress, despair, impaired ambulation
PRO-CTCAE	Chemotherapy	Adult Cancer (General)	M /F	Inpatient; Outpatient; Home	Standardized	Constitutional	y	y	Heart palpitations, ringing ears, blurry/flashing lights/spots (floaters) vision, abdominal pain/bloating, constipation, diarrhea, loss of bowel control, increased flatulence, dry/ skin cracking at corners/sores in mouth, heartburn, nausea, vomiting, shivering/shaking chills, swelling/tingling/numbness in arm(s)/hand(s)/leg(s)/feet, fatigue/tiredness,

									difficulty concentrating, pain/swelling/redness at IV/injection site, bruise easily, burned skin, decreased appetite, aching joints/muscles, problems tasting food/drink, headache, problems with memory, unable/long time to orgasm, decreased sex drive, pain during vaginal sex, problems with ejaculation, difficulty with erections, enlarged/tender breasts, irregular periods, vaginal dryness/discharge, anxiety, depressed mood, change in urine color, frequent/ sudden urges/loss of control of urination, pain/burning with urination, difficulty sleeping, insomnia, coughing, difficulty breathing, wheezing sound when breathing, nosebleed, hiccups, changes in/ hoarse voice, hair loss, body odor, dry skin, unexpected excessive/decreased sweating, change in color/ridges/bumps/loss of nails, hand-foot syndrome, skin sensitive to light, itchy skin/acne/pimples, chest rash, stretch marks, darkening of skin, bed sores, hives, hot flashes
Rotterdam Symptom Checklist	Chemotherapy	Adult Cancer (General)	M /F	Outpatient; Home	Standardized	Constitutional	y	n	Acid indigestion, shivering, tingling hands or feet difficulty concentrating, sore mouth/pain when swallowing, loss of hair, burning/sore eyes shortness of breath, dry mouth
Severity of Dyspepsia Assessment (SODA)	Anti-platelet therapy	Adult dyspepsia	M /F	Outpatient; Home	Standardized	Gastrointestinal	y	y	Pain, abdominal discomfort, burping/belching, heartburn, bloating, passing gas, sour taste, nausea, bad breath
Side Effect Scoring System for Helicobacter Pylori treatment (SES-HP)	Treatment	Adult Helicobacter Pylori	M /F	Outpatient; Home	Standardized	Gastrointestinal	n	y	Severe discomfort, requiring discontinuation of treatment, taste disturbance, loss of appetite, nausea, vomiting, stomach pain, bloating, diarrhea, constipation, skin rash
Skindex-16 toxicity measurement instrument	Drug treatment	Adult Cancer (Breast)	M /F	Outpatient; Home	Standardized	Dermatological, Mental Health	n	y	Itching, burning, stinging, hurting, irritated skin, persistence/reoccurrence of skin condition
Symptom Tracking and Reporting online reporting system (STAR)	Chemotherapy	Adult Cancer (General)	M /F	Outpatient; Home	Standardized	Constitutional	n	y	Pain, fatigue, appetite loss, vomiting, increased bowel movement, diarrhea, constipation, weight gain/loss

TABLE 2.3B: CHARACTERISTICS OF GENERIC PROM-AEs REPORTED IN MEDICAL LITERATURE

Elements	Intervention Used	Patient age group/Disorder	Sex	Setting of use	Question Type	Local/Constitutional	Frequency Used (y/n)	Severity Used (y/n)	AE symptoms listed
Memorial Symptom Assessment Scale (MSAS)	Chemotherapy	Adult Cancer (General)	M / F	Inpatient	Standardized	Constitutional	y	y	Difficulty concentrating, pain, lack of energy, cough, feeling nervous, dry mouth, nausea, feeling drowsy, numbness/tingling in hands/feet, difficulty sleeping, feeling bloated, problems with urination, vomiting, shortness of breath, diarrhea, feeling sad, sweats, worrying, problems with sexual, interest or activity, itching, lack of appetite, dizziness, difficulty swallowing, feeling irritable, mouth sores, change in the way food tastes, weight loss, hair loss, constipation, swelling of arms or legs
Yellow Card Scheme Questionnaire (YCS)	Treatment (General)	Adult & Pediatric (General)	M / F	Home	Individualized	Constitutional	y	y	None(open ended)

LEGEND:

Question Type: Whether the questions in the survey have predetermined answers (standardized) or open ended questions for patients to report their experiences (individualized)

Local/Constitutional: The particular body system for which the PROM-AE is designed, or is applicable to; constitutional means a non-specific and generalized location.

TABLE 2.4: TYPES OF INCLUDED ARTICLES

Study type	Frequency	%
Validation Study	24	34.8
RCT	11	15.9
Cohort Study	7	10.1
Literature Review	6	8.7
Systematic Review	4	5.8
Editorial	4	5.8
Pilot study	5	7.2
Survey	3	4.3
Case Series	3	4.3
Scoping Review	1	1.4
Descriptive Study	1	1.4
Total	69	100

TABLE 2.5: PATIENT DIAGNOSES IN INCLUDED STUDIES

Patient Sample Diagnosis	Frequency	%
Cancer (General)	8	22.9
Cancer (Prostate)	4	11.4
Cancer (Head and Neck)	4	11.4
Cancer (Breast)	3	8.6
Cancer (Lung)	2	5.7
Schizophrenia	2	5.7
Surgery	2	5.7
Multiple Sclerosis	1	2.9
HIV	1	2.9
General	1	2.9
Heliobacter Pylori	1	2.9
Dyspepsia	1	2.9
Epilepsy	1	2.9
ADHD	1	2.9
Arthritis	1	2.9
Asthma	1	2.9
Chronic Low Back Pain (CBLP)	1	2.9
Total	35	100

TABLE 2.6: STUDY TREATMENTS OF INCLUDED STUDIES

Treatment Type	Frequency	%
Chemotherapy	28	40.6
Radiotherapy	7	10.1
Treatment/Drugs (General)	5	7.2
Surgery	4	5.8
Opioid drugs	4	5.8
Antipsychotic drugs	3	4.3
Chemotherapy & Radiotherapy	2	2.9
Corticosteroid drugs	2	2.9
Anti-ADHD drugs	2	2.9
Anti-HIV drugs	2	2.9
Anti-epileptic drugs	2	2.9
Anti-platelet therapy	1	1.4
Endocrine therapy	1	1.4
Interferon drugs	1	1.4
Neuroleptic drugs	1	1.4
Probiotics	1	1.4
Stem cell transplant	1	1.4
Anti-dyspepsia drugs	1	1.4
Anti-anemia drugs	1	1.4
Total	69	100

TABLE 2.7: MAIN CONSTRUCTS

PROM-AE Constructs	#	%
Severity of AE	29	82.9
Frequency of AE	14	40
Name of medicine/treatment	2	5.7
Reason for taking medicine/treatment	2	5.7
Dosage of medicine/treatment	2	5.7
Age	2	5.7
Sex	2	5.7
Current patient status	2	5.7
Interference of AE in daily life	1	2.9
Weight	1	2.9
Prescribing doctor	1	2.9
Total # PROM-AEs	35	100

APPENDICES

Appendix A: Search Strategies

i) MEDLINE

1. exp Drug Interactions/

2. Drug Hypersensitivity/

3. Adverse Drug Reaction Reporting Systems/

4. Drug Toxicity/

5. (ae or co or de).fs.

6. (safe or safety).mp. or side effect*.ti,ab. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]

7. undesirable effect*.ti,ab.

8. (treatment emergent or tolerability).mp. or toxicity.ti,ab. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]

9. (adverse adj2 (effect or effects or reaction or reactions or event or events or outcome or outcomes)).ti,ab.

10. or/1-9

11. patient reported outcomes.mp.

12. patient centered.mp. or patient centred.ti,ab. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]

13. patient generated.ti,ab.

14. consumer generated.mp. or patient oriented.ti,ab. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]

15. patient perspective$.ti,ab.

16. patient generated index.ti,ab.

17. (repertory grid or subjective quality of life profile or disease repercussion profile).mp. or patient enablement instrument.ti,ab. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]

18. well being questionnaire.ti,ab.

19. 11 or 12 or 13 or 14 or 15 or 16 or 17

20. 10 and 19

21. randomized controlled trial.pt.

22. clinical trial.pt.

23. guideline.pt.

24. review.pt.

25. randomi?ed.ti,ab.

26. placebo.ti,ab.

27. dt.fs.

= don't use RCT filter

28. 21 or 22 or 23 or 24 or 25 or 26 or 27

29. 20 and 28

30. limit 29 to humans

Total references: 1636 (without RCT filter: 3344)

ii) EMBASE

1. Drug Interactions/

2. drug hypersensitivity/

3. drug surveillance program/

4. drug toxicity/

5. (ae or co or de).fs.

6. drug/ae [Adverse Drug Reaction]

7. (safe or safety or side effect*).mp. or undesirable effect*.ti,ab. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]

8. (treatment emergent or tolerability).mp. or toxicity.ti,ab. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]

9. (adverse adj2 (effect or effects or reaction or reactions or event or events or outcome or outcomes)).ti,ab.

10. or/1-9

11. patient reported outcome$.mp.

12. patient centered.ti,ab.

13. patient centred.ti,ab.

14. patient generated.ti,ab.

15. consumer generated.ti,ab.

16. patient oriented.ti,ab.

17. patient perspective$.ti,ab.

18. (repertory grid or subjective quality of life profile or disease repercussion profile or patient enablement instrument).ti,ab.

19. well being.mp. or wellbeing questionnair$.ti,ab. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]

20. 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19

21. 10 and 20

**22. randomized controlled trial/

23. controlled clinical trial/

24. placebo.ti,ab.

25. randomi?ed.ti,ab.

26. "systematic review (topic)"/

27. practice guideline/

28. 22 or 23 or 24 or 25 or 26 or 27

29. 21 and 28

30. limit 29 to human

(iii) CINAHL

S1 MH Drug Hypersensitivity

S2 (MH "Adverse Drug Event+")

S3 (MH "Drug Interactions+")

S4 (MH "Drug Toxicity")

S5 AB safe or safety

S6 AB side effect*

S7 AB undesirable effect*

S8 AB toxicity

S9 AB "adverse effect*"

S10 AB "adverse reaction*"

S11 AB "adverse event*"

S12 (S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11)

S13 "patient reported outcome*"

S14 "patient generated"

S15 "patient perspective$"

S16 "wellbeing questionnaire$"

S17 "well being questionnaire$"

S18 "patient oriented"

S19 (S13 OR S14 OR S15 OR S16 OR S17 OR S18)

S20 ((S13 OR S14 OR S15 OR S16 OR S17 OR S18)) AND (S12 AND S19)

S21 (MH "Randomized Controlled Trials") OR (MH "Clinical Trials+")

S23 S20 AND S21

Appendix B: Inclusion/Exclusion Sheet

Ref ID	ILL	Library	Retreived (Y/N)	Study Type (1=RCT, 2=SR, 3=other)	Not Human	Not PRO-AE	Duplicate Publication	Language	I/E	Notes	Citation

Appendix C: Data Extraction Sheet

			Study Design					Measurement Tool					
Ref ID	Publication Name	Year	Study Type	Patient characteristics	Sample size	Sex	Intervention Type	Name	Generic (1) Disease-specific (2)	Standardized (1) Individualized (2)	Constructs	AEs reported	References

CHAPTER 3: COSMIN Checklist Evaluation of Patient Reported Outcome Measurement tools of Adverse Events (PROM-AE) Validation studies: PRO-CTCAE & Yellow Card Scheme (YCS)

Myles Hancock, *Masters Candidate*[1], Lisa Hartling *Associate Professor*[1,2]„ Don Voaklander, *Professor*[3], Sunita Vohra, *Professor*[1,3]

[1] Department of Pediatrics, Faculty of Medicine & Dentistry, University of Alberta
Edmonton Clinic Health Academy
11405-87 Avenue
Edmonton, Alberta
Canada T6G 1C9
[2]Alberta Research Center for Health Evidence, University of Alberta
Edmonton Clinic Health Academy
11405-87 Avenue
Edmonton, AB T6T 1C9
[3]School of Public Health, University of Alberta
3-300 Edmonton Clinic Health Academy
11405 – 87 Ave
Edmonton, Alberta
Canada T6G 1C9

Correspondence to: S Vohra svohra@ualberta.ca

ABSTRACT

Objective: To retrieve validation studies of the Yellow Card Scheme (YCS) and the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE), and evaluate the measurement properties reporting of articles using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Checklist.

Methods: Evaluation study using the Terwee Search Filter, searched for relevant articles on measurement properties in the MEDLINE database. Inclusion criteria: i) English Language; ii) focuses on either the YCS or PRO-CTCAE; iii) validated one or more measurement properties of the YCS or PRO-CTCAE. Accepted articles were evaluated for their measurement properties reporting using the COSMIN checklist four point scoring system.

Results: The search filter retrieved one article for the YCS, reporting on face validity; three PRO-CTCAE validation articles were included, looking at content validity, construct validity, test-retest reliability, measurement error, and responsiveness. All studies scored excellent on reporting with our interpretation of the COSMIN checklist.

Conclusion: The main goal of this study was to evaluate the quality of measurement properties reporting in studies validating the YCS and PRO-CTCAE. We found evidence of validation being done for PROM-AEs used most used in research. We recommend that more effort be put into ensuring PROM-AEs are comprehensive with measurement properties assessments.

BACKGROUND

PROM-AEs

Patient reported outcome measurement tools of adverse events (PROM-AEs) are a method of reporting harms that is gaining traction (1). Patients are uniquely able to provide a detailed perspective on AEs that they experience, often with specific information that may be overlooked by their health care providers (2). Some federal regulatory agencies now incorporate patient self-reports of AEs, rather than relying on reports from physicians or other health care providers (3, 4), PROM-AEs are also gaining traction in health research (3). In order to enhance patient safety in both research and clinical settings, it is important to seek patient reports using instruments that are valid and reliable.

PROM Validation

Key elements of outcome measurement are validity, reliability, and responsiveness (4). These domains assess if the instrument is measuring what it is supposed to, that it measures consistently, and that it can successfully detect change over time, respectively. If research relies on instruments lacking these constructs, it is difficult to trust the study's conclusions.

A systematic review (Chapter 2) was undertaken to find all PROM-AE tools in the peer-reviewed medical literature published in English. From this systematic review, two PROM-AEs were selected for further evaluation; one generic and one disease-specific. We selected the Yellow Card Scheme (YCS), as it was the most widely used generic PROM-AE found in the review. The YCS is an example of a PROM-AE designed for patient harms reporting from any treatment in any setting. The Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) tool was the most referenced disease-specific

PROM-AE in the review. The PRO-CTCAE provides an example of a patient harms reporting tool designed and used in a research setting.

Yellow Card Scheme (YCS)

The Yellow Card Scheme (YCS) was a pioneer AE reporting measure, first developed in the United Kingdom (UK) in 1964 for health care professionals (HCP) to record any patient harms potentially associated with treatment (5). Since 2005, patients have also been able use the PROM-AE version to report any AEs to the Medicines & Healthcare products Regulatory Agency (MHRA) of the UK government (6). The YCS asks patients to provide relevant demographic information as well as details surrounding their AE occurrence, including age, sex, weight, and family physician of the individual affected; the specifics of the AE asked about are: (i) name of the medicine/treatment; (ii) reason for taking it; (iii) dosage of the medicine (if applicable); and (iv) the current status of the affected patient (7).

A study of patient AE reporting was completed in 2011 using the YCS in the UK. It was found that in a 2 year span over 26,000 patients sent in the form citing a potential drug related harm to the government (1). Another study in 2011 found that patient reports using the patient-geared YCS were more detailed in symptom description and reported more adverse drug reactions per submission (median 3, IQR (2-5)) than HCPs (median 2, IQR (1-3)) (5). These studies indicate the importance of patient reporting to enhance the quality and quantity of AE reports, as not all AEs are identified or reported by health care professionals (8).

Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE)

The PRO-CTCAE was developed for use with cancer patients to obtain AE information from toxicity or other treatment-related difficulties (9). The initiative for this process gained traction when the United States National Cancer Institute (NCI) determined that clinicians using the original physician-reporting version of the CTCAE were underreporting AEs (10). Given the research evidence establishing benefits of obtaining the patient perspective of their treatment, development began on a patient-reporting version of this instrument (10).

The PRO-CTCAE contains 124 items across various body systems and symptoms experienced by patients with cancer. Systems include eye, ear, respiratory, cardiovascular, gastrointestinal, renal/urinary, reproductive, musculoskeletal, nervous, psychiatric, and skin/subcutaneous tissue while common symptoms include pain, swelling, and appetite (10). The tool asks the patient to specify the frequency, severity, and level of interference in daily life caused by the AE (10). Researchers administer the tool by first selecting the relevant body system(s) and symptoms to their study (10). The PRO-CTCAE has been used in clinical trials and continues to be tested, translated into different languages, and formatted for different platforms (paper, internet and telephone administration) (9).

MAIN OBJECTIVES

In this evaluation study we aimed to: (i) retrieve studies on the measurement properties of the YCS and PRO-CTCAE tools; (ii) evaluate the methodological quality of the studies reporting validity and reliability of both PROM-AEs.

METHODS

Search Strategy

An experienced health research librarian entered the validated Terwee PubMed search filter into the MEDLINE database to identify studies on the measurement properties of the YCS and PRO-CTCAE (13). One database was chosen for my thesis for feasibility reasons. Medline was selected to be appropriate, given its coverage of life sciences articles spanning from 1964 to present, which totals more than 24 million references. It is considered to be the largest database for medical literature, and comprises the majority of literature available on the Pubmed search engine (12). Prior to peer-reviewed publication, I will ensure the search is run in other major databases. This specialized filter contains a combination of search terms that have been demonstrated to be highly sensitive and specific in finding articles that assess outcomes using measurement properties (11). Please see Appendix A for the full list of search terms. The reference list of included articles was also examined for other relevant literature.

Study Selection

In the first stage of screening, the titles and abstracts of study entries were screened by one person (MH). During the second stage full texts of potentially relevant studies were retrieved and selected. Any issues were resolved by consulting the supervisory committee for guidance. Studies were included if they: i) were published in English ii) mentioned use of the PRO-CTCAE or YCS measures: and iii) evaluated one or more measurement properties of the PRO-CTCAE or YCS (Please see Figure 1 for the flow diagram).

Quality Assessment of Studies

To evaluate the methodological quality of included studies, the COSMIN checklist was used (12). This checklist was constructed via an international Delphi study with four rounds of international expert feedback to decide the key components of measurement-related methodological quality. The COSMIN checklist describes items pertinent to nine measurement properties and presents specific reporting standards for each. These are grouped within the 3 main domains: (i) reliability (internal consistency, measurement error, reliability); (ii) validity (content, construct, criterion); and (iii) responsiveness (12). The COSMIN scoring system was developed to enable reporting and comparison of methodological quality between studies (11).

COSMIN defines validity as whether the PROM measures the intended variable in the target population (11,12). Under validity, COSMIN further classified several validity-related measurement properties. Content validity is confirmed when a PROM measures the construct with all essential components of the outcome present and working together (11,12). Construct validity is established if the measurement tool demonstrates a strong correlation to similar measures in a relevant patient sample (12). Criterion validity can be measured. This is the degree to which measurement as true as an external evaluation considered to be a "gold standard" measurement tool (12).

Reliability is the extent to which "the measurement is free of measurement error" (12). Therefore, measures with higher reliability possess minimal error (12). The COSMIN group classified reliability into three measurement properties. The first is reliability, defined as "the extent to which scores for patients who have not changed are the same for repeated

measurements" (11). COSMIN further defines reliability into different contexts; test-retest ("over time"), inter-rater ("by different persons on the same occasion"), and intra-rater ("by the same persons on different occasions") reliability types (11). The other two measurement properties are internal consistency ("degree of inter-relatedness of the items" (11)); and measurement error ("systematic and random error not attributed to true changes in the construct to be measured" (12)).

COSMIN defines responsiveness as "the ability of a PRO instrument to detect change over time in the construct to be measured" (12). This is useful when constructs can vary, such as the severity of an adverse event (AE) through the duration of treatment (12). The fourth domain is interpretability. This is the extent a reviewer "can assign qualitative meaning—that is, clinical or commonly understood connotations—to an instrument's quantitative scores or change in scores" (11).

The checklist was used by one reviewer (MH) in the following three steps. First, it was used to identify measurement properties that were reported in each study. Second, each measurement property identified in the first step was rated on a 4-point scale (excellent, good, fair, or poor) in accordance with COSMIN evaluation standards (12). Finally, as the COSMIN guidelines dictate, the lowest rating represented the overall methodological quality score. The rating "excellent" applied if adequate evidence was given regarding the measurement property in question. COSMIN uses the term "adequate" to define when the reporting fulfils the requirement for proper measurement properties reporting (13). The "good" rating was given "when relevant information is missing, but it can still be assumed there is adequate quality in the measurement property" (13). A "fair" rating was assigned when it was unclear if there was adequate

information on a specific measurement property. A “poor” rating was assigned when there was inadequate information (12). Please see the COSMIN checklist (Appendix B) for further details.

Data Extraction

Relevant data were extracted from eligible studies by a single reviewer (MH). The primary goal was to gather the essential study information required for results reporting and completion of the COSMIN checklist. The main data extracted included author, publication year, study design, patient population, PROM-AE tool utilized, measurement properties described, and relevant references. At the time of thesis submission, data were not independently verified; however, this is planned prior to submission for publication.

RESULTS

Yellow Card Scheme (YCS)

The search retrieved one published article for the Yellow Card Scheme that focused on evaluation of the tool. There were two validation studies completed within the report. The first was a survey questionnaire done in the UK, where researchers provided the option of paper, telephone, or online module modes of administration to patient participants. Any patients who submitted a YCS form between March and December 2008 received a notice asking for participation. The goal was to obtain patient opinion on their experiences using the YCS form, and any suggestions they may have to improve it (6). Any patients who submitted a YCS form between March and December 2008 received a notice asking for participation; of 2008 potential participants, 1362 replied (68%). Other patient characteristic of note included gender (66.8% female), and ethnicity (93.5% white). The second study involved a series of adult volunteer focus groups with a usability testing component; researchers gave patient participants a similar

opportunity to voice their suggested improvement to the YCS form, but in a face-to-face environment (6). Patient feedback included replacing medical language with lay terminology on the online form, more space on the paper version form to write personal experiences, and larger print on both modes for those with vision problems (6); main characteristics described included 40 participants completing the sessions, 67.5% of them were female, and 74% were reported to be over 50 years old.

Methodological Quality

The first study reported a confirmation of face validity. This was evaluated by providing a questionnaire for patient feedback on the use of the YCS for future amendments for improvement. Please see Table 1 for further characteristics of each study. COSMIN classifies face validity as the extent that PROM appears to be "an adequate reflection of the construct to be measured" (14). COSMIN also labels face validity as an aspect of the measurement property content validity; defined as the degree to which the components of the measurement tool are each an "adequate reflection of the construct to be measured" (14). The second study verbally assessed content validity with focus group feedback (6). According to the COSMIN checklist, we rated the study as "excellent" in reporting of content validity. The study fulfilled all requirements, assessing patient feedback on whether all items were relevant for the purpose of application, and whether all items were comprehensive enough to reflect the construct being measured; in addition, the sample sizes of the first and second studies was greater than the minimum required 10 patients (n=1362, n=40).

PRO-CTCAE

The search identified one commentary that summarized the development and validation, and six studies that evaluated measurement properties of the PRO-CTCAE. Three studies about the validation of the tool translated into Spanish, Danish and German were excluded, as they were outside the scope of this assessment (15-17). The commentary was published in 2014, and outlined the current and future validation plans (10). Although it mentioned the three included validation studies while each was in progress (8, 18, 19), it did not evaluate any measurement properties and was therefore excluded.

Methodological Quality

Within the three included studies, the reporting of content validity, construct validity, test-retest reliability, measurement error, and responsiveness were evaluated. As there is no "gold standard" for PROMs (18), it was not possible to evaluate criterion validity. Inter-rater reliability was not deemed relevant, as in PROMs there is only one patient reporting on the specified outcome. Internal consistency was also not applicable, as not all question items on a particular symptom construct needed to be present to have the specified symptom in the PRO-CTCAE (13).

The first study was a validation of the PRO-CTCAE across the main areas of measurement properties. It assessed the construct validity, test-retest reliability, measurement error, and responsiveness of the PROM-AE in a large patient sample undergoing chemotherapy treatment for various cancers (9). Out of the 975 patients enrolled, 96.4% (n=940) completed the first visit, and 90.6% (n=852) of 940 completed the second visit. Patient characteristics included age (median 59 years), sex (57.3% female), ethnicity (88.5% no Hispanic/Latino), and cancer

type (majority: 35% lung, head or neck) (9). The study found the PRO-CTCAE performed well in all of these measurement properties (9). Construct validity is defined by the COSMIN checklist as the extent a PROM is consistent with "internal relationships, relationships to scores of other instruments, or differences between relevant groups" and is measuring what it is supposed to (14). The study compared the PRO-CTCAE with a well-used health-related quality of life (HRQOL) PROM, the European Organization for Research and Treatment of Cancer (QLQ-C30). The authors reversed the direction of scores for the EORTC QLQ-C30 to mirror the PRO-CTCAE scores. In the results section, they reported that 122 (98.4%) of the 124 items on the PRO-CTCAE had a Pearson correlation in the expected direction with the EORTC QLQ-C30. Overall COSMIN assessment for construct validity reporting was scored as excellent; authors provided the percentage of missing items and how they were handled, had an adequate sample size ($n \geq 100$), gave an a priori hypothesis with expected direction and magnitude of correlation, and gave an adequate description and measurement properties of the instrument that was used for comparison (14).

Test-retest reliability is defined as whether the PROM correctly measures the same condition as unchanged over time (14). Measurement error, a measurement property of the larger domain reliability, is the "systematic and random error of a patient's score not attributed to true changes in the construct to be measured" (14). Using COSMIN, we scored the reporting of these properties as excellent; the adequate requirements for test-retest reliability and measurement error included having at least two measurements of survey administration in similar conditions were available for comparison, and stating an appropriate time interval, which in this case was one day (9). The interclass correlation (ICC) was calculated, and the patients were assumed to be stable throughout the survey. To establish this stability, the COSMIN tool required an

assessment of a global rating of change (14); the article reported that it did so, citing the "Global Impressions of Change (GIC)" was used on patients' follow up visits (9). Responsiveness is the ability of a PROM to measure change over time (14). The reporting of responsiveness was rated as excellent quality. A longitudinal design was utilized with a stated time interval of one to six weeks apart (9). Any changes in patient condition or occurrences were reported to an adequate degree, as the researchers stated the patient's chemotherapy sessions (14).

The second study, also done in the US, assessed content validity by conducting cognitive interviews of cancer patients and getting feedback on their experiences using the PRO-CTCAE. 127 patients participated in three interview rounds, and were evaluated on" comprehension, memory retrieval, judgement, and response mapping"(20); characteristics included education level (35% high school education or lower), sex (75% female), ethnicity (91% white/non-Hispanic), and disease site (majority: 21% genitourinary). Content validity is defined by COSMIN as "the degree to which the content of a HR-PRO (health related—patient reported outcome) instrument is an adequate reflection of the construct to be measured" (14). The checklist found the study to have excellent quality of reporting; the requirements for an adequate "excellent" score included whether the study made assessments to determine whether all items were relevant to the measured construct; in this case, the semi-structured interviews (20)**.** In addition, the researchers involved the target population, cancer patients, in evaluating the PRO-CTCAE items (20). The article also possessed adequate patient sample size (>100), and comprehensiveness in reflecting the construct, and was therefore relevant to the purpose of the application (4).

The third study contributed to test-retest reliability and measurement error (19). It scored overall excellent in reporting both measurement properties. The third study, carried out in the US, compared different types of test administration (tablet/computer, interactive voice response, and paper based) for the PRO-CTCAE amongst a subset of 112 cancer patients undergoing chemotherapy or radiotherapy (12). Noteworthy characteristics included sex (59.8% female), age (median 56.5 years), ethnicity (76.8% white), and cancer type (majority: 34% breast) (19). The COSMIN checklist states the essential elements for both test-retest reliability and measurement error are to be having more than one independent measurement, a description and explanation given of the percentage of missing items (12). The article provided detailed information about missing participant data, and explained that there was very low missing data because the survey administration occurred in the clinic (19). They also reported an adequate sample size (>100). The time interval was also reported (one hour), and the test conditions were stated to be similar (19). Lastly, regarding test-retest reliability, the ICC and kappa were calculated for continuous and dichotomous scores respectively (12, 19).

DISCUSSION

Our main objective was to evaluate the quality of the studies' measurement properties reporting in a PROM-AE and a condition-specific PROM-AE using the COSMIN checklist.

For the PRO-CTCAE, we were able to examine the reporting of test-retest reliability, measurement error, content validity, construct validity, and responsiveness. We found the methodological quality scores in the included studies to be excellent, according to the COSMIN checklist. The researchers described their steps to ensure validation was carried out comprehensively. Any missing information in the main paper was provided in the supplemental material, and the study text provided relevant links.

The YCS multiple-study paper was deemed excellent in assessing content validity. The results affirmed that the patients' ease of using the YCS was improved by obtaining direct feedback from them. However, the report does not discuss measurement properties in detail when compared with the PRO-CTCAE validation studies, and is more qualitative in nature having focus group methodology (6). We contacted the UK Medicines & Healthcare products Regulatory Agency (MHRA) to inquire if any more YCS validation studies were completed or in progress. Their reply referenced the included report as the most current published evidence. In the email correspondence, the MHRA also confirmed that all patient feedback had been applied to a newer version of the form, and that the patient version of the YCS is now treated equally to physician-based form submissions in their regulatory assessment of reported AEs.

While we found the COSMIN scores to be of excellent quality, more work is required to assess and validate measurement properties not yet addressed by the PROM-AEs. The included PRO-CTCAE articles provided evidence of excellent test-retest reliability, content validity, construct validity, and responsiveness. The measurement properties relevant to the tool have been evaluated, but the PRO-CTCAE does not necessarily include all relevant constructs (discussed further below). The YCS possesses content validity from the published study. However, construct validity, reliability, measurement error, and responsiveness have not been investigated. The field would benefit from additional work to ensure instruments are valid and reliable in what outcomes they measure.

Both PROM-AEs possess some of the core constructs found in the prior systematic review (Chapter 2), but not all of them. The PRO-CTCAE has the constructs severity, frequency, and level of interference in the patient's life. The constructs the PRO-CTCAE does not capture include: reason for taking medicine/treatment, age, sex, weight, current patient status, and

identifying information (patient's name as well as prescribing physician). Conversely, the YCS possesses all of these constructs save for frequency of AE. We recommend that researchers establish what constructs are most relevant to PROM-AEs, and then ensure that these are measured in a valid and reliable fashion.

The methodological quality of this assessment of validation depended on several factors. The first main component was that the Terwee search filter only retrieved studies with properties of measurement tools; the filter was shown prior to be efficacious (12). The PRO-CTCAE, a tool currently being used in cancer clinical trials, had a number of retrievable validation articles with mostly strong methodological quality (9). In contrast, there was one YCS validation article found. This may have been because the YCS survey is used mainly for patient self-reporting of harms in nationwide surveillance, and less in clinical research (1). The second main contributor to methodological quality in this study was using the COSMIN checklist to examine the articles with regards to methodological quality of the validation studies. Delphi consensus was reached with international experts, strengthening the accountability of results when using these methods (4, 14).

Regarding limitations, this validation study was done by one reviewer, and ratings are unconfirmed by a second independent reviewer. Prior to publication, the findings presented in this chapter will be independently verified by a second reviewer, with discrepancies discussed with a senior reviewer and resolved through consensus. We limited the inclusion criteria to English language only; while this means relevant papers in foreign languages were not included, it is unclear if these assessments would contribute to the assessment of the English instrument. Patient comprehension and validation of measurement instruments can vary depending on language, culture, and setting (21).

CONCLUSION

The study examined the reporting of measurement properties in validation studies using internationally recommended guidelines. We found evidence showing validation of measurement properties for the PROM AEs used most often in clinical research, the PRO-CTCAE, and substantially less information about the validation of a widely used PROM-AE in routine health care settings, the YCS. We recommend further work to ensure PROM AE instruments are comprehensive and that their measurement properties are formally assessed.

REFERENCES

1. McLernon DJ, Bond CM, Hannaford PC, Watson MC, Lee AJ, Hazell L, et al. Adverse drug reaction reporting in the UK: a retrospective observational comparison of yellow card reports submitted by patients and healthcare professionals. Drug Safety. 2010;33(9):775-88.
2. Fromme EK, Eilers KM, Mori M, Hsieh Y-C, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. Journal of Clinical Oncology. 2004;22(17):3485-90.
3. Aagaard L, Nielsen LH, Hansen EH. Consumer reporting of adverse drug reactions. Drug safety. 2009;11(32):1067-74.
4. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of Life Research. 2010;19(4):539-49.
5. McLernon DJ, Bond CM, Lee AJ, Watson MC, Hannaford PC, Fortnum H, et al. Patient views and experiences of making adverse drug reaction reports to the Yellow Card Scheme in the UK. Pharmacoepidemiology and drug safety. 2011;20(5):523-31.
6. Avery AJ, Anderson C, Bond C, Fortnum H, Gifford A, Hannaford PC, et al. Evaluation of patient reporting of adverse drug reactions to the UK'Yellow Card Scheme': literature review, descriptive and qualitative analyses, and questionnaire surveys. Health Technology Assessment. 2011;15(20):1-234.
7. (MHRA) MHpRA. Reporting Forms United Kingdom2017 [cited 2017 April 25th]. Available from: https://yellowcard.mhra.gov.uk/downloadable-information/.
8. Zorzela L, Golder S, Liu Y, Pilkington K, Hartling L, Joffe A, et al. Quality of reporting in systematic reviews of adverse events: systematic review. 2014.
9. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, et al. Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). JAMA Oncology. 2015;1(8):1051-9.
10. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). Journal of the National Cancer Institute. 2014;106(9):dju244.
11. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Quality of Life Research. 2009;18(8):1115-23.
12. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of Life Research. 2012;21(4):651-7.
13. Mokkink LB, Terwee, Caroline B, Patrick, Donald L, Alonso, Jordi et al. . COSMIN checklist manual Amsterdam, the Netherlands2012 [cited 2017 6/1]. Available from: www.cosmin.nl/images/upload/files/COSMIN checklist manual v9.pdf.
14. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. Journal of clinical epidemiology. 2010;63(7):737-45.
15. Arnold B, Mitchell SA, Lent L, Mendoza TR, Rogak LJ, Barragan NM, et al. Linguistic validation of the Spanish version of the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Supportive Care in Cancer. 2016;24(7):2843-51.
16. Baeksted C, Nissen A, Pappot H, Bidstrup PE, Mitchell SA, Basch E, et al. Danish Translation and Linguistic Validation of the U.S. National Cancer Institute's Patient-Reported Outcomes Version of the

Common Terminology Criteria for Adverse Events (PRO-CTCAE). Journal of Pain & Symptom Management. 2016.
17. Kirsch M, Mitchell SA, Dobbels F, Stussi G, Basch E, Halter JP, et al. Linguistic and content validation of a German-language PRO-CTCAE-based patient-reported outcomes instrument to evaluate the late effect symptom experience after allogeneic hematopoietic stem cell transplantation. European Journal of Oncology Nursing. 2015;19(1):66-74.
18. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury. 2011;42(3):236-40.
19. Bennett AV, Dueck AC, Mitchell SA, Mendoza TR, Reeve BB, Atkinson TM, et al. Mode equivalence and acceptability of tablet computer-, interactive voice response system-, and paper-based administration of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Health & Quality of Life Outcomes. 2016;14:24.
20. Hay JL, Atkinson TM, Reeve BB, Mitchell SA, Mendoza TR, Willis G, et al. Cognitive interviewing of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Quality of Life Research. 2014;23(1):257-69.
21. Sperber AD. Translation and validation of study instruments for cross-cultural research. Gastroenterology. 2004;126:S124-S8.

APPENDICES

Figure 3.1: Flow diagram of selection of studies COSMIN Measurement roperties Study

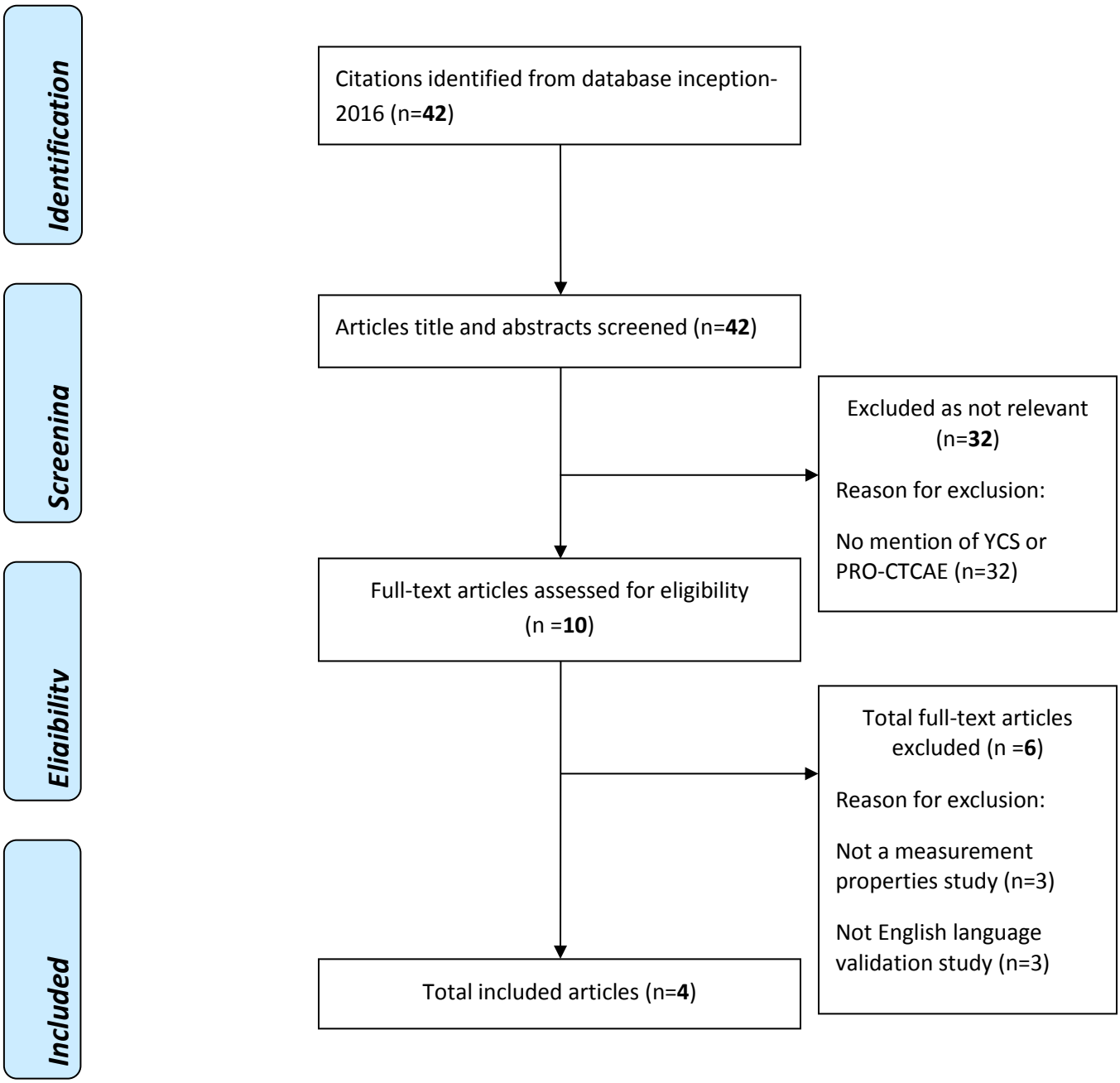


TABLE 3.1: STUDY CHARACTERISTICS

PROM-AE	Setting	Patient type	Sample size	Reference/retrieved study
YCS	UK	Prior experience filling out the Yellow Card Scheme	1362	Avery AJ, Anderson C, Bond CM, Fortnum H, Gifford A, Hannaford PC, Hazell L, Krska J, Lee AJ, Mclernon DJ, Murphy E. Evaluation of patient reporting of adverse drug reactions to the UK'Yellow Card Scheme': literature review, descriptive and qualitative analyses, and questionnaire surveys. Health Technology Assessment. 2011;15(20):1-234.
PRO-CTCAE	US	Any type of cancer undergoing chemotherapy or radiotherapy	112	Bennett AV, Dueck AC, Mitchell SA, Mendoza TR, Reeve BB, Atkinson TM, et al. Mode equivalence and acceptability of tablet computer-, interactive voice response system-, and paper-based administration of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Health & Quality of Life Outcomes. 2016;14:24.
PRO-CTCAE	US	Any type of cancer undergoing chemotherapy or radiotherapy	975	Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, et al. Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). JAMA Oncology. 2015;1(8):1051-9.
PRO-CTCAE	US	Any type of cancer undergoing chemotherapy or radiotherapy	127	Hay JL, Atkinson TM, Reeve BB, Mitchell SA, Mendoza TR, Willis G, et al. Cognitive interviewing of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Quality of Life Research. 2014;23(1):257-69.

TABLE 3.2: METHODOLOGICAL QUALITY OF STUDIES USING THE COSMIN SCORING SYSTEM

Outcome Measure		**Reliability**			**Validity**		**Responsiveness**
	Internal Consistency	Measurement Error	Reliability	Content	Construct	Criterion	Responsiveness
PRO-CTCAE	NR	Excellent (56, 64)	Excellent (56, 64)	Excellent (63)	Excellent (56)	NR	Excellent(56)
Yellow Card Scheme (YCS)	NR	NR	NR	Excellent (55)	NR	NR	NR

NR: Not reported

Appendix A: MH PROAE Validation Search Strategy (Medline May 2016)

1. (instrumentation or methods).sh.

2. (Validation Studies or Comparative Study).pt.

3. exp Psychometrics/

4. psychometr*.ti,ab.

5. (clinimetr* or clinometr*).tw.

6. exp "Outcome Assessment (Health Care)"/

7. outcome assessment.ti,ab.

8. outcome measure*.tw.

9. exp Observer Variation/

10. observer variation.ti,ab.

11. exp Health Status Indicators/

12. exp "Reproducibility of Results"/

13. reproducib*.ti,ab.

14. exp Discriminant Analysis/

15. (reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or "internal consistency").ti,ab.

16. (cronbach* and (alpha or alphas)).ti,ab.

17. (item and (correlation* or selection* or reduction*)).ti,ab.

18. (agreement or precision or imprecision or "precise values" or test-retest).ti,ab.

19. (test and retest).ti,ab.

20. (reliab* and (test or retest)).ti,ab.

21. (stability or interrater or inter-rater or intrarater or intra-rater or intertester or intertester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intra-observer or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or inter-assay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappa's or kappas or repeatab*).ti,ab.

22. ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab.

23. (generaliza* or generalisa* or concordance).ti,ab.

24. (intraclass and correlation*).ti,ab.

25. (discriminative or "known group" or factor analysis or factor analyses or dimension* or subscale*).ti,ab.

26. (multitrait and scaling and (analysis or analyses)).ti,ab.

27. (item discriminant or interscale correlation* or error or errors or "individual variability").ti,ab.

28. (variability and (analysis or values)).ti,ab.

29. (uncertainty and (measurement or measuring)).ti,ab.

30. ("standard error of measurement" or sensitiv* or responsive*).ti,ab.

31. ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ti,ab.

32. (small* and (real or detectable) and (change or difference)).ti,ab.

33. (meaningful change or "ceiling effect" or "floor effect" or "Item response model" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross-cultural equivalence").ti,ab.

34. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33

35. (Patient-Reported Outcomes CTCAE or PRO-CTCAE or PROCTCAE or Patient-Reported Outcomes Common Terminology Criteria for Adverse Events ORPRO Common Terminology Criteria for Adverse Events).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]

36. (Yellow Card Scheme or Yellow Card Form or Yellow Card or YCS).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]

37. 35 or 36

38. 34 and 37

Appendix B: COSMIN Checklist

COSMIN checklist with 4-point scale

Contact
CB Terwee, PhD
VU University Medical Center
Department of Epidemiology and Biostatistics
EMGO Institute for Health and Care Research
1081 BT Amsterdam
The Netherlands
Website: www.cosmin.nl, www.emgo.nl
E-mail: cb.terwee@vumc.nl

Instructions

This version of the COSMIN checklist is recommended for use in systematic reviews of measurement properties. With this version it is possible to calculate overall methodological quality scores per study on a measurement property. A methodological quality score per box is obtained by taking the lowest rating of any item in a box ('worse score counts'). For example, if for a reliability study one item in the box 'Reliability' is scored poor, the methodological quality of that reliability study is rated as poor. The Interpretability box and the Generalizability box are mainly used as data extraction forms. We recommend to use the Interpretability box to extract all information on the interpretability issues described in this box (e.g. norm scores, floor-ceiling effects, minimal important change) of the instruments under study from the included articles. Similar, we recommend to use the Generalizability box to extract data on the characteristics of the study population and sampling procedure. Therefore no scoring system was developed for these boxes.

This scoring system is described in this paper:

Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of Life Research 2011, July 6 [epub ahead of print].

Step 1. Evaluated measurement properties in the article

	Internal consistency	Box A
	Reliability	Box B
	Measurement error	Box C
	Content validity	Box D
	Structural validity	Box E
	Hypotheses testing	Box F
	Cross-cultural validity	Box G
	Criterion validity	Box H
	Responsiveness	Box I

Step 2. Determining if the statistical method used in the article are based on CTT or IRT

Box General requirements for studies that applied Item Response Theory (IRT) models		excellent	good	fair	poor
1	Was the IRT model used adequately described? e.g. One Parameter Logistic Model (OPLM), Partial Credit Model (PCM), Graded Response Model (GRM)	IRT model adequately described	IRT model not adequately described		
2	Was the computer software package used adequately described? e.g. RUMM2020, WINSTEPS, OPLM, MULTILOG, PARSCALE, BILOG, NLMIXED	Software package adequately described	Software package not adequately described		
3	Was the method of estimation used adequately described? e.g. conditional maximum likelihood (CML), marginal maximum likelihood (MML)	Method of estimation adequately described	Method of estimation not adequately described		
4	Were the assumptions for estimating parameters of the IRT model checked? e.g. unidimensionality, local independence, and item fit (e.g. differential item functioning (DIF))	assumptions of the IRT model checked	assumptions of the IRT model partly checked	assumptions of the IRT model not checked or unknown	

To obtain a total score for the methodological quality of studies that use IRT methods, the 'worse score counts' algorithm should be applied to the IRT box in combination with the box of the measurement property that was evaluated in the IRT study. For example, if IRT methods are used to study internal consistency and item 4 in the IRT box is scored fair, while the items in the internal consistency box (box A) are all scored as good or excellent, the methodological quality score for internal consistency will be fair. However, if any of the items in box A is scored poor, the methodological quality score for internal consistency will be poor.

Step 3. Determining if a study meets the standards for good methodological quality

Box A. Internal consistency	**excellent**	**good**	**fair**	**poor**
1 Does the scale consist of effect indicators, i.e. is it based on a reflective model?				
Design requirements				
2 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
3 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
4 Was the sample size included in the internal consistency analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
5 Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	Factor analysis performed in the study population	Authors refer to another study in which factor analysis was performed in a similar study population	Authors refer to another study in which factor analysis was performed, but not in a similar study population	Factor analysis NOT performed and no reference to another study
6 Was the sample size included in the unidimensionality analysis adequate?	7* #items and ≥100	5* #items and ≥100 OR 6-7* #items but <100	5* #items but <100	<5* #items

7	Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?	Internal consistency statistic calculated for each subscale separately		Internal consistency statistic NOT calculated for each subscale separately
8	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods				
9	for Classical Test Theory (CTT), continuous scores: Was Cronbach's alpha calculated?	Cronbach's alpha calculated	Only item-total correlations calculated	No Cronbach's alpha and no item-total correlations calculated
10	for CTT, dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated	Only item-total correlations calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated
11	for IRT: Was a goodness of fit statistic at a global level calculated? E.g. χ^2, reliability coefficient of estimated latent trait value (index of (subject or item) separation)	Goodness of fit statistic at a global level calculated		Goodness of fit statistic at a global level NOT calculated

NB. Item 1 is used to determine whether internal consistency is relevant for the instrument under study. It is not used to rate the quality of the study.

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)				
	excellent	**good**	**fair**	**poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Were at least two measurements available?	At least two measurements			Only one measurement
5 Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	measurements NOT independent
6 Was the time interval stated?	Time interval stated		Time interval NOT stated	
7 Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate

9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods					
11	for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated
12	for dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
13	for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
14	for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		

Box C. Measurement error: absolute measures				
	excellent	**good**	**fair**	**poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Were at least two measurements available?	At least two measurements			Only one measurement
5 Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	measurements NOT independent
6 Was the time interval stated?	Time interval stated		Time interval NOT stated	
7 Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate

9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar
10	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods					
11	for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	SEM, SDC, or LoA calculated	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population

Box D. Content validity (including face validity)					
		excellent	**good**	**fair**	**poor**
General requirements					
1	Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	Assessed if all items refer to relevant aspects of the construct to be measured		Aspects of the construct to be measured poorly described AND this was not taken into consideration	NOT assessed if all items refer to relevant aspects of the construct to be measured

2	Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	Assessed if all items are relevant for the study population in adequate sample size (≥10)	Assessed if all items are relevant for the study population in moderate sample size (5-9)	Assessed if all items are relevant for the study population in small sample size (<5)	NOT assessed if all items are relevant for the study population OR target population not involved
3	Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)	Assessed if all items are relevant for the purpose of the application	Purpose of the instrument was not described but assumed	NOT assessed if all items are relevant for the purpose of the application	
4	Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	Assessed if all items together comprehensively reflect the construct to be measured		No theoretical foundation of the construct and this was not taken into consideration	NOT assessed if all items together comprehen-sively reflect the construct to be measured
5	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study

Box E. Structural validity				
	excellent	**good**	**fair**	**poor**
1 Does the scale consist of effect indicators, i.e. is it based on a reflective model?				
Design requirements				
2 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
3 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
4 Was the sample size included in the analysis adequate?	7* #items and ≥100	5* #items and ≥100 OR 5-7* #items but <100	5* #items but <100	<5* #items
5 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. rotation method not described)	Other important methodological flaws in the design or execution of the study (e.g. inappropriate rotation method)

Statistical methods				
6 for CTT: Was exploratory or confirmatory factor analysis performed?	Exploratory or confirmatory factor analysis performed and type of factor analysis appropriate in view of existing information	Exploratory factor analysis performed while confirmatory would have been more appropriate		No exploratory or confirmatory factor analysis performed
7 for IRT: Were IRT tests for determining the (uni-) dimensionality of the items performed?	IRT test for determining (uni)dimension-ality performed			IRT test for determining (uni)dimension-ality NOT performed

Box F. Hypotheses testing				
	excellent	**good**	**fair**	**Poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (<30 per analysis)

4	Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	Multiple hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
5	Was the expected *direction* of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated		
6	Was the expected absolute or relative *magnitude* of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
7	for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
8	for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)

9 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study
Statistical methods				
10 Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Box G. Cross-cultural validity

	excellent	**good**	**fair**	**poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	

3	Was the sample size included in the analysis adequate?	CTT: 7* #items and ≥100 IRT: ≥200 per group	CTT: 5* #items and ≥100 OR 5-7* #items but <100 IRT: ≥200 in 1 group and 100-199 in 1 group	CTT: 5* #items but <100 IRT: 100-199 per group	CTT: <5* #items IRT: (<100 in 1 or both groups
4	Were both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	Both source language and target language described			Source language NOT known
5	Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the disease(s) involved, expertise in the construct to be measured, expertise in both languages	Expertise of the translators described with respect to disease, construct, and language	Expertise of the translators with respect to disease or construct poor or not described	Expertise of the translators with respect to language not described	
6	Did the translators work independently from each other?	Translators worked independent	Assumable that the translators worked independent	Unclear whether translators worked independent	Translators worked NOT independent
7	Were items translated forward and backward?	Multiple forward and multiple backward translations	Multiple forward translations but one backward translation	One forward and one backward translation	Only a forward translation
8	Was there an adequate description of how differences between the original and translated versions were resolved?	Adequate description of how differences between translators were resolved	Poorly or NOT described how differences between translators were resolved		

9	Was the translation reviewed by a committee (e.g. original developers)?	Translation reviewed by a committee (involving other people than the translators, e.g. the original developers)	Translation NOT reviewed by (such) a committee		
10	Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	Translated instrument pre-tested in the target population	Translated instrument pre-tested, but unclear if this was done in the target population	Translated instrument pre-tested, but NOT in the target population	Translated instrument NOT pre-tested
11	Was the sample used in the pre-test adequately described?	Sample used in the pre-test adequately described		Sample used in the pre-test NOT (adequately) described	
12	Were the samples similar for all characteristics except language and/or cultural background?	Shown that samples were similar for all characteristics except language /culture	Stated (but not shown) that samples were similar for all characteristics except language /culture	Unclear whether samples were similar for all characteristics except language /culture	Samples were NOT similar for all characteristics except language /culture
13	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study

Statistical methods					
14	for CTT: Was confirmatory factor analysis performed?	Multiple-group confirmatory factor analysis performed			Multiple-group confirmatory factor analysis NOT performed
15	for IRT: Was differential item function (DIF) between language groups assessed?	DIF between language groups assessed			DIF between language groups NOT assessed

Box H. Criterion validity

		excellent	**good**	**fair**	**poor**
Design requirements					
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4	Can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'

5	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods				
6	for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated		Correlations or AUC NOT calculated
7	for dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated		Sensitivity and specificity NOT calculated

Box I. Responsiveness

		excellent	**good**	**fair**	**poor**
Design requirements					
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4	Was a longitudinal design with at least two measurement used?	Longitudinal design used			No longitudinal design used
5	Was the time interval stated?	Time interval adequately described			Time interval NOT described

6	If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	Anything that occurred during the interim period (e.g. treatment) adequately described	Assumable what occurred during the interim period	Unclear or NOT described what occurred during the interim period	
7	Was a proportion of the patients changed (i.e. improvement or deterioration)?	Part of the patients were changed (evidence provided)	NO evidence provided, but assumable that part of the patients were changed	Unclear if part of the patients were changed	Patients were NOT changed
Design requirements for hypotheses testing					
	For constructs for which a gold standard was not available:				
8	Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	Hypotheses formulated a priori		Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
9	Was the expected *direction* of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated		
10	Were the expected absolute or relative *magnitude* of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		

11	Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)		Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
12	Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	NO information on the measurement properties of the comparator instrument(s)
13	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study
Statistical methods					
14	Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate		Statistical methods applied NOT optimal	Statistical methods applied NOT

Design requirement for comparison to a gold standard					
	For constructs for which a gold standard was available:				
15	Can the criterion for change be considered as a reasonable gold standard?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
16	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods					
17	for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	Correlations or Area under the ROC Curve (AUC) calculated			Correlations or AUC NOT calculated
18	for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated

We recommend to use the Interpretability box to extract all information on the interpretability issues described in this box of the instruments under study from the included articles.

Box Interpretability

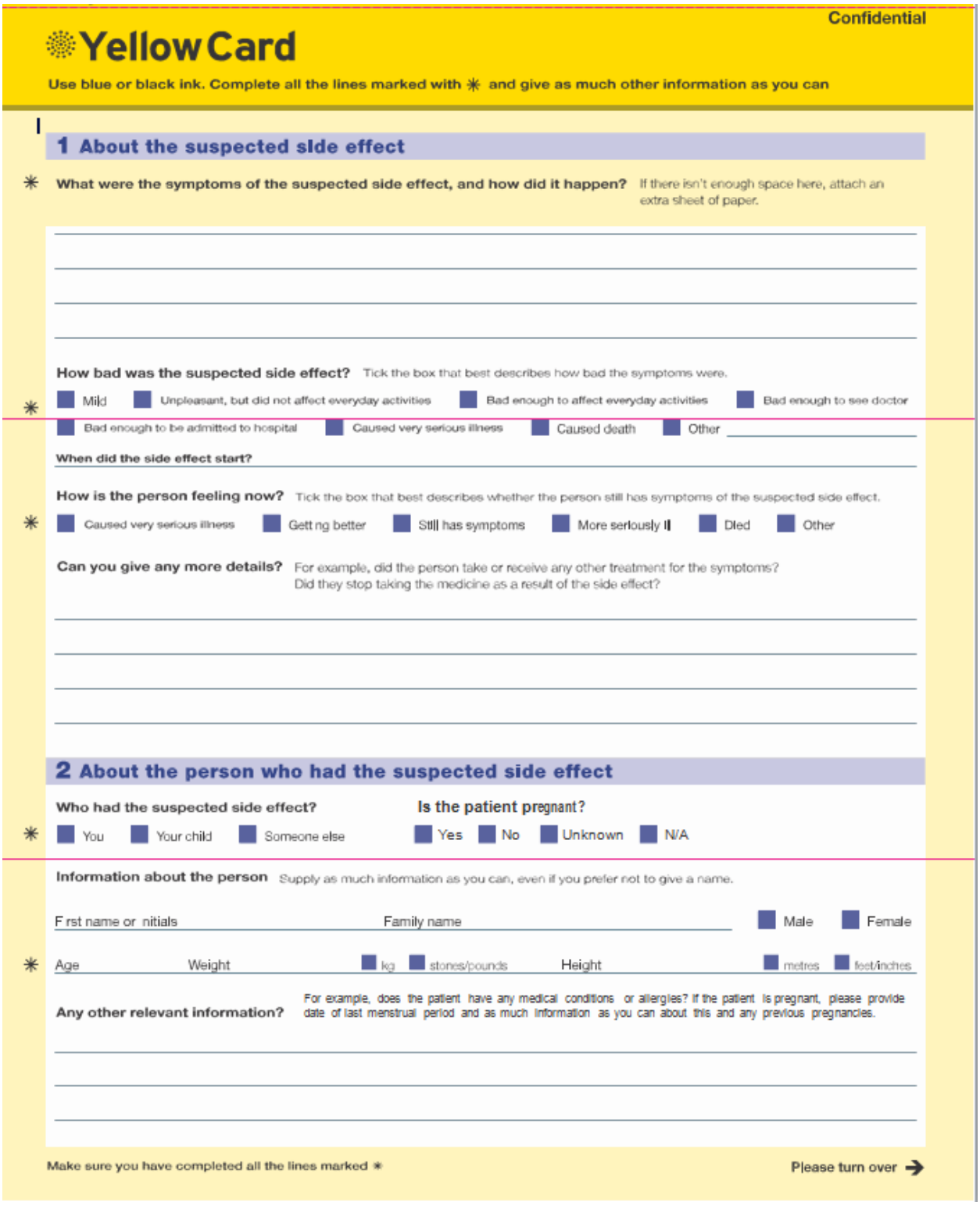
Percentage of missing items	
Description of how missing items were handled	
Distribution of the (total) scores	
Percentage of the respondents who had the lowest possible (total) score	
Percentage of the respondents who had the highest possible (total) score	
Scores and change scores (i.e. means and SD) for relevant (sub) groups, e.g. for normative groups, subgroups of patients, or the general population	
Minimal Important Change (MIC) or Minimal Important Difference (MID)	

We recommend to use the Generalizability box to extract data on the characteristics of the study populations and sampling procedures of the included studies.

Box Generalisability

Median or mean age (with standard deviation or range)	
Distribution of sex	
Important disease characteristics (e.g. severity, status, duration) and description of treatment	
Setting(s) in which the study was conducted (e.g. general population, primary care or hospital/rehabilitation care)	
Countries in which the study was conducted	
Language in which the HR-PRO instrument was evaluated	
Method used to select patients (e.g. convenience, consecutive, or random)	
Percentage of missing responses (response rate)	

Appendix C: Yellow Card Scheme (YCS)



Confidential

Yellow Card

Use blue or black ink. Complete all the lines marked with ✳ and give as much other information as you can

1 About the suspected side effect

✳ **What were the symptoms of the suspected side effect, and how did it happen?** If there isn't enough space here, attach an extra sheet of paper.

How bad was the suspected side effect? Tick the box that best describes how bad the symptoms were.

✳ Mild | Unpleasant, but did not affect everyday activities | Bad enough to affect everyday activities | Bad enough to see doctor | Bad enough to be admitted to hospital | Caused very serious illness | Caused death | Other

When did the side effect start?

How is the person feeling now? Tick the box that best describes whether the person still has symptoms of the suspected side effect.

✳ Caused very serious illness | Getting better | Still has symptoms | More seriously ill | Died | Other

Can you give any more details? For example, did the person take or receive any other treatment for the symptoms? Did they stop taking the medicine as a result of the side effect?

2 About the person who had the suspected side effect

Who had the suspected side effect? ✳ You | Your child | Someone else

Is the patient pregnant? Yes | No | Unknown | N/A

Information about the person Supply as much information as you can, even if you prefer not to give a name.

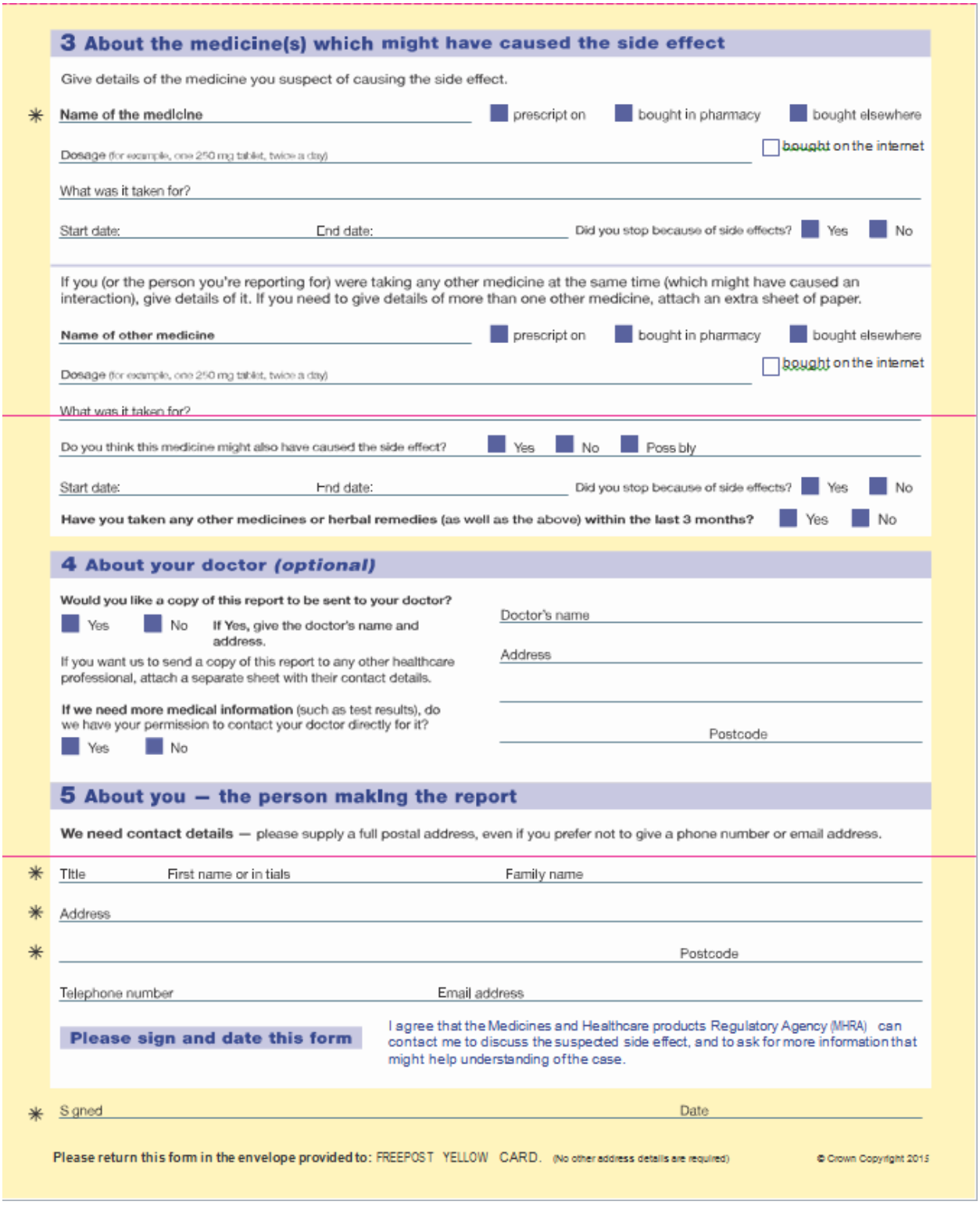
First name or initials | Family name | Male | Female

✳ Age | Weight | kg | stones/pounds | Height | metres | feet/inches

Any other relevant information? For example, does the patient have any medical conditions or allergies? If the patient is pregnant, please provide date of last menstrual period and as much information as you can about this and any previous pregnancies.

Make sure you have completed all the lines marked ✳

Please turn over ➔



3 About the medicine(s) which might have caused the side effect

Give details of the medicine you suspect of causing the side effect.

* **Name of the medicine** ☐ prescript on ☐ bought in pharmacy ☐ bought elsewhere ☐ bought on the internet

Dosage (for example, one 250 mg tablet, twice a day)

What was it taken for?

Start date: End date: Did you stop because of side effects? ☐ Yes ☐ No

If you (or the person you're reporting for) were taking any other medicine at the same time (which might have caused an interaction), give details of it. If you need to give details of more than one other medicine, attach an extra sheet of paper.

Name of other medicine ☐ prescript on ☐ bought in pharmacy ☐ bought elsewhere ☐ bought on the internet

Dosage (for example, one 250 mg tablet, twice a day)

What was it taken for?

Do you think this medicine might also have caused the side effect? ☐ Yes ☐ No ☐ Possibly

Start date: End date: Did you stop because of side effects? ☐ Yes ☐ No

Have you taken any other medicines or herbal remedies (as well as the above) within the last 3 months? ☐ Yes ☐ No

4 About your doctor *(optional)*

Would you like a copy of this report to be sent to your doctor?

☐ Yes ☐ No **If Yes, give the doctor's name and address.**

If you want us to send a copy of this report to any other healthcare professional, attach a separate sheet with their contact details.

If we need more medical information (such as test results), do we have your permission to contact your doctor directly for it?

☐ Yes ☐ No

Doctor's name

Address

Postcode

5 About you — the person making the report

We need contact details — please supply a full postal address, even if you prefer not to give a phone number or email address.

* Title First name or initials Family name

* Address

* Postcode

Telephone number Email address

Please sign and date this form

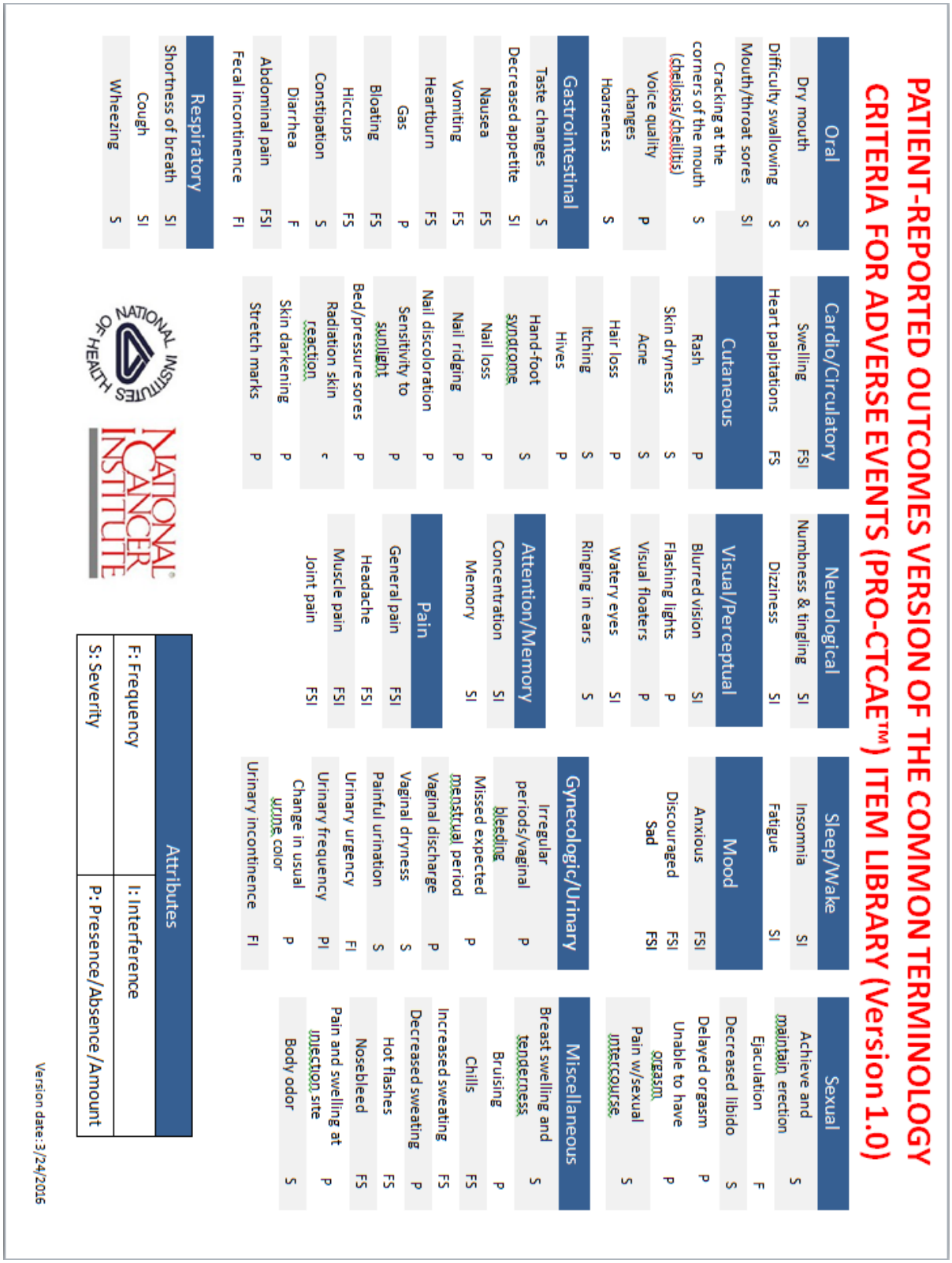
I agree that the Medicines and Healthcare products Regulatory Agency (MHRA) can contact me to discuss the suspected side effect, and to ask for more information that might help understanding of the case.

* Signed Date

Please return this form in the envelope provided to: FREEPOST YELLOW CARD. (No other address details are required)

https://yellowcard.mhra.gov.uk/_assets/files/2016-06-17-MHRA-Yellow-Card-patient-form-to-report-side-effects.pdf

Appendix D: PRO-CTCAE Item Library



https://healthcaredelivery.cancer.gov/pro-ctcae/item-library.pdf

CHAPTER 4: Conclusion.

Thesis Objectives and Results

We conducted two studies, the first to find all PROM-AEs used in the published literature (objective 1). The systematic review retrieved PROM-AEs targeting a variety of disorders and treatments. Most PROM-AEs identified in the review were condition-specific, as opposed to generic instruments. The highest proportion of measures was from cancer-related research, such as the PRO-CTCAE, the most highly cited measure identified. This indicates that AE reporting is a priority in cancer research (1). As evidenced by the many PROM-AEs identified, cancer is a field populated by varied efforts to get the patient perspective. We speculate that this may be due to the amount of research occurring in cancer relative to other fields; according to the NIH in 2016, cancer received the most funding out of all areas of health research (2).” In addition, cancer treatment (chemotherapy, radiation therapy, surgery) is well recognized to have important/serious associated adverse effects. We conclude that other areas of medical research with limited to no reported PROM-AE use would benefit from more vigilance, as harms occur across all conditions and treatments.

The COSMIN evaluation study (objective 2) evaluated the methodological quality of two PROM-AEs identified in the systematic review: one is the most commonly used disease-specific measure; the other is the longest-standing generic measure. Despite the YCS having a lengthy history of use (2) and the majority of main constructs established from the systematic review, we found little work published for its validation using the Terwee search filter. This paucity of studies may demonstrate that regulatory body-generated measures receive little formal scrutiny, as no other agency’s PROM-AE was found in the systematic review, even though they are now implemented across many federal regulatory agencies (e.g. Health Canada, FDA). In contrast,

robust validity and reliability assessment of PRO-CTCAE was available in the peer reviewed literature (3); however, it only possessed some of the constructs from the systematic review. Therefore, more work is needed to test the validity and reliability of the PROM-AE constructs.

Implications in Clinical Research

Researchers are placing increasing importance on patient responses to treatment-related AEs (4). Harms are well known to be under-reported in clinical trials and systematic reviews; PROM-AEs could improve both the quantity and quality of harms reporting. Thus far, only cancer research appears to have moved forward with the development of a valid and reliable PROM-AE instrument. There is variability in PROM-AE measurement in cancer, and a considerable gap in all other conditions. The field of patient AE reporting appears incomplete and may benefit from a core set of outcomes that are consistently measured between studies. Systematic review authors have confirmed the difficulty of synthesizing knowledge when outcome measurement instruments are heterogeneous (5).

Reducing heterogeneity in outcome measurement is the main focus of the Core Outcome Measures in Effectiveness Trials (COMET) Initiative. The international group was formed with the common goal to establish core outcome sets (COS) to improve clinical trial reporting across all diseases/conditions (6). COMET defines COS as a consensus-based minimum set of outcomes for a particular disorder or patient population; these agreed-upon outcomes should be measured and reported in all trials in that field (7). The COMET rationale is that through consensus, stakeholders are able to select the most relevant and important outcomes (7). Subsequent research could then yield results that can be efficiently "compared, contrasted and combined" (8).

We believe there would be value in developing a COS to enhance AE reporting. Disease-specific COS are relevant, given the differences in symptoms and outcomes. Given the lack of reporting of AEs, and the serious risk to patient safety posed by health interventions, a generic COS with a core set of constructs (e.g., severity, frequency) could facilitate AE reporting across all conditions. A COS would promote comparisons between studies for knowledge synthesis. Our systematic review identified a range of constructs measured in PROM-AEs (Table 1), which could be a starting point for the development of COS for AE reporting. Development of a COS would bring together the relevant stakeholders, such as PROM-AE researchers, patients, health care providers, and policy makers. Patients in particular should be a focus, as evidence shows, their perspective is essential for comprehensive AE reporting.

In addition, we encourage giving consideration to integrating PROM-AE reporting into all COS, as harms are relevant across all health conditions and interventions. Established research groups such as CONSORT and PRISMA have stressed in their guidelines the importance for harms as a main component of trial and systematic review reporting (9, 10). However, recent research confirms AE reporting remains very limited (9). This may be of concern, as researchers have estimated medical errors to be the third leading cause of death in the US (11). Therefore, there is an urgent need to include AE measurement and reporting in all clinical research.

If all COS are to include a PROM-AE component, the most feasible measure would be one that is generic and individualized. Such a PROM-AE would allow maximum utility across different disease/conditions, while capturing individual experience. The specificity of what symptoms the patient experienced can be reported with individualizing response options, as seen in the MYMOP tool. Patients provide the outcomes that are most relevant to them, but the core

constructs can remain consistent. The YCS is the most prominent example of a generic and individualized PROM-AE. The YCS possesses most of the core PROM-AE constructs identified in our systematic review, save for frequency. Nonetheless, in the absence of a formal COS process, we cannot be sure whether all the relevant constructs for AE reporting have been identified. The setting may influence what is needed from a measurement tool in that specific instance of AE reporting during research or daily care. Once we can establish the core components of AEs for COS, these can be applied across different settings and diseases. Given the limited evidence of published work, consensus on what further validation and evaluation of measurement properties is required.

Implications in Clinical Practice

In routine medical practice, harms are also under-reported (11). Patient reports tend to describe AE in more detail and higher severity than HCP reports (4). While there have been efforts testing PROMs in clinical care environments, we have yet to see such use of PROM-AEs (4). Future challenges for implementation of PROM-AEs are likely to mirror those encountered with PROM testing trials in the UK, especially the time and cost for collecting data, analyzing results, drawing conclusions from and presenting results to relevant stakeholders for feedback (4).

We believe there is much to gain from PROM-AE implementation in clinical practice. PROM-AEs can help patients report their AEs, and promote efficient refinement of treatment options (12). PROM-AE use may also help patients feel that HCPs are more invested in their health, and want to receive patient feedback (3). Improving self-reporting methods of patient harms will support the inclusion of patient priorities. Patient-reported satisfaction and support for PROM-AE implementation may subsequently increase as a result.

The collaboration and consensus of researchers, patients, HCPs and policy makers in the dissemination of clinical PROM-AE use results would raise the relevance of feedback reporting to regulatory agencies. Health Canada, the US Food and Drug Administration (FDA), and the UK Medicines and Healthcare products Regulatory Agency (MHRA) are all regulatory agencies that have PROM-AEs in use for nationwide surveillance of patient harms (13-15). We propose that if a PROM-AE is developed through COS consensus and validated, it could conceivably be harmonized from research to clinical care and across regulatory agencies. This way essential patient safety information on AEs could be shared efficiently across stakeholders.

Limitations

Both studies were limited to English-language only. Our rationale was that each language would require its own validation. As these two instruments were developed in English, we chose to look for measurement properties studies in English. Another limitation was the introduction of potential bias by only having a single assessor for the COSMIN checklist. The use of a validated checklist tool helps strengthen the objectivity of the evaluation; prior to publication, seconding will be sought and consensus achieved.

Conclusion

The main goal of this thesis was to identify patient-reported measurement tools for adverse events in clinical research and clinical care. While cancer has more than 15 PROM-AEs, most conditions have few to none. Opportunities for improvement include reducing excess variation in measurement through a consensus-based process to develop a core outcome set involving all relevant stakeholders. Such efforts could yield an instrument that has utility in both

clinical research and clinical care. Improved patient safety is a priority for all who have a stake in health care. PROM-AEs represent an area where this priority can be strengthened further.

References

1. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). Journal of the National Cancer Institute. 2014;106(9):dju244.
2. Anderson C, Krska J, Murphy E, Avery A, Yellow Card SC. The importance of direct patient reporting of suspected adverse drug reactions: a patient perspective. British journal of clinical pharmacology. 2011;72(5):806-22.
3. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. Bmj. 2010;340:c186.
4. Black N. Patient reported outcome measures could help transform healthcare. BMJ (Clinical research ed). 2013;346:f167.
5. Gaertner J, Siemens W, Camilleri M, Davies A, Drossman DA, Webster LR, et al. Definitions and outcome measures of clinical trials regarding opioid-induced constipation: a systematic review. Journal of clinical gastroenterology. 2015;49(1):9.
6. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012;13(1):132.
7. Williamson PR, Altman DG, Blazeby JM, Clarke M, Gargon E. The COMET (core outcome measures in effectiveness trials) initiative. Trials. 2011;12(Suppl 1):A70.
8. Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. PloS one. 2014;9(6):e99111.
9. Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. bmj. 2016;352:i157.
10. Ioannidis JP, Evans SJ, Gøtzsche PC, O'neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Annals of internal medicine. 2004;141(10):781-8.
11. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ: British Medical Journal (Online). 2016;353.
12. Foster JM, van dM, Caeser M, Hannaford P. The use of questionnaires for measuring patient-reported side effects of drugs: its importance and methodological challenges. Pharmacoepidemiology and drug safety. 2008;17(3):278-96.
13. Administration USFaD. Forms for Reporting to FDA 2015 [cited 2015 07/22/2015]. Available from: https://www.fda.gov/safety/medwatch/howtoreport/downloadforms/default.htm.
14. Agency MHPR. Yellow Card United Kingdom: Goverment of the United Kingdom; 2016 [cited 2016 November 10]. Available from: https://yellowcard.mhra.gov.uk/.
15. Canada H. Adverse Reaction and Medical Device Problem Reporting 2016 [Available from: http://www.hc-sc.gc.ca/dhp-mps/medeff/report-declaration/index-eng.php#a1.

TABLE 1: MAIN CONSTRUCTS FROM SYSTEMATIC REVIEW

PROM-AE Constructs	#	%
Severity of AE	28	80
Frequency of AE	13	37.1
Name of medicine/treatment	2	5.7
Reason for taking medicine/treatment	2	5.7
Dosage of medicine/treatment	2	5.7
Age	2	5.7
Sex	2	5.7
Current patient status	2	5.7
Interference of AE in daily life	1	2.9
Weight	1	2.9
Prescribing doctor	1	2.9
Total # PROM-AEs	35	100

ALL REFERENCES

1. (MHRA) MHpRA. Reporting Forms United Kingdom2017 [cited 2017 April 25th]. Available from: https://yellowcard.mhra.gov.uk/downloadable-information/.
2. Aagaard L, Nielsen LH, Hansen EH. Consumer reporting of adverse drug reactions. Drug safety. 2009;11(32):1067-74.
3. Administration UFaD. Guidance for industry: Patient-reported outcome measures—Use in medical product development to support labeling claims. Federal Register. 2009;235(74):65132-3.
4. Administration UFaD. Forms for Reporting to FDA 2015 [cited 2015 07/22/2015]. Available from: https://www.fda.gov/safety/medwatch/howtoreport/downloadforms/default.htm.
5. Agency MHPR. Yellow Card United Kingdom: Goverment of the United Kingdom; 2016 [cited 2016 November 10]. Available from: https://yellowcard.mhra.gov.uk/.
6. Anderson C, Krska J, Murphy E, Avery A, Yellow Card SC. The importance of direct patient reporting of suspected adverse drug reactions: a patient perspective. British journal of clinical pharmacology. 2011;72(5):806-22.
7. Arnold B, Mitchell SA, Lent L, Mendoza TR, Rogak LJ, Barragan NM, et al. Linguistic validation of the Spanish version of the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Supportive Care in Cancer. 2016;24(7):2843-51.
8. Avery AJ, Anderson C, Bond C, Fortnum H, Gifford A, Hannaford PC, et al. Evaluation of patient reporting of adverse drug reactions to the UK'Yellow Card Scheme': literature review, descriptive and qualitative analyses, and questionnaire surveys. Health Technology Assessment. 2011;15(20):1-234.
9. Baeksted C, Nissen A, Pappot H, Bidstrup PE, Mitchell SA, Basch E, et al. Danish Translation and Linguistic Validation of the U.S. National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Journal of Pain & Symptom Management. 2016.
10. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. Canadian medical association journal. 2004;170(11):1678-86.
11. Basch E. The missing voice of patients in drug-safety reporting. New England Journal of Medicine. 2010;362(10):865-9.
12. Basch E, Iasonos A, McDonough T, Barz A, Culkin A, Kris MG, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. The lancet oncology. 2006;7(11):903-9.
13. Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). Journal of the National Cancer Institute. 2014;106(9):dju244.
14. Bateman E, Keefe D, editors. Patient-reported outcomes in supportive care. Seminars in oncology; 2011: Elsevier.
15. Bennett AV, Dueck AC, Mitchell SA, Mendoza TR, Reeve BB, Atkinson TM, et al. Mode equivalence and acceptability of tablet computer-, interactive voice response system-, and paper-based administration of the U.S. National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Health & Quality of Life Outcomes. 2016;14:24.
16. Bhaloo Z, Adams D, Liu Y, et al. Primary outcomes reporting in trials (PORTal): A systematic review of pediatric randomized controlled trials. . Journal of Clinical Epidemiology (in press). Forthcoming 2016

17. Bhaloo ZK. Measurement Properties of Primary Outcome Measures in Pediatric RCTs: Inadequate Reporting, Insufficient Validation, Or Both? : University of Alberta; 2013.
18. Black N. Patient reported outcome measures could help transform healthcare. BMJ (Clinical research ed). 2013;346:f167.
19. Brazier J, Tsuchiya A. Preference-based condition-specific measures of health: what happens to cross programme comparability? Health economics. 2010;19(2):125-9.
20. Canada H. Adverse Reaction Information 2012. Available from: http://www.hc-sc.gc.ca/dhp-mps/medeff/advers-react-neg/index-eng.php.
21. Canada H. MedEffect Canada Canada: Government of Canada; 2016 [cited 2016 November 10]. Available from: http://www.hc-sc.gc.ca/dhp-mps/medeff/index-eng.php.
22. Canada H. Adverse Reaction and Medical Device Problem Reporting 2016 [updated February 24th, 2017; cited 2017 February 13]. Available from: http://www.hc-sc.gc.ca/dhp-mps/medeff/report-declaration/index-eng.php.
23. Carlson LE, Bultz BD, Morris DG. Individualized quality of life, standardized quality of life, and distress in patients undergoing a phase I trial of the novel therapeutic Reolysin (reovirus). Health and Quality of Life Outcomes. 2005;3(1):1.
24. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Medical care. 2007;45(5 Suppl 1):S3.
25. Chan A, Cameron MC, Garden B, Boers-Doets C, Schindler K, Epstein JB, et al. A systematic review of patient-reported outcome instruments of dermatologic adverse events associated with targeted cancer therapies. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer. 2015;23(8):2231.
26. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. Bmj. 2010;340:c186.
27. de Langen J, van Hunsel F, Passier A, de Jong-van den Berg L, van Grootheest K. Adverse drug reaction reporting by patients in The Netherlands three years of experience. Drug Safety. 2008;31(6):515-24.
28. De Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide: Cambridge University Press; 2011.
29. de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Quality and safety in health care. 2008;17(3):216-23.
30. de Vries ST, Mol PG, de Zeeuw D, Haaijer-Ruskamp FM, Denig P. Development and initial validation of a patient-reported adverse drug event questionnaire. Drug safety. 2013;36(9):765-77.
31. Dissemination CfRa. Systematic reviews: CRD's guidance for undertaking reviews in health care. United Kingdom: University of York; 2009.
32. Dueck AC, Mendoza TR, Mitchell SA, Reeve BB, Castro KM, Rogak LJ, et al. Validity and Reliability of the US National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). JAMA Oncology. 2015;1(8):1051-9.
33. Esper P, Mo F, Chodak G, Sinner M, Cella D, Pienta KJ. Measuring quality of life in men with prostate cancer using the functional assessment of cancer therapy-prostate instrument. Urology. 1997;50(6):920-8.
34. Foster JM, van dM, Caeser M, Hannaford P. The use of questionnaires for measuring patient-reported side effects of drugs: its importance and methodological challenges. Pharmacoepidemiology and drug safety. 2008;17(3):278-96.
35. Frank L, Basch E, Selby JV. The PCORI perspective on patient-centered outcomes research. Jama. 2014;312(15):1513-4.

36. Fromme EK, Eilers KM, Mori M, Hsieh Y-C, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. Journal of Clinical Oncology. 2004;22(17):3485-90.
37. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Group MFP-ROCM. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value in Health. 2007;10:S94-S105.
38. Gaertner J, Siemens W, Camilleri M, Davies A, Drossman DA, Webster LR, et al. Definitions and outcome measures of clinical trials regarding opioid-induced constipation: a systematic review. Journal of clinical gastroenterology. 2015;49(1):9.
39. Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. PloS one. 2014;9(6):e99111.
40. Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M. The COMET Initiative database: progress and activities from 2011 to 2013. Trials. 2014;15:279-6215-15-279.
41. Golder S, Loke YK. The performance of adverse effects search filters in MEDLINE and EMBASE. Health Information & Libraries Journal. 2012;29(2):141-51.
42. Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? Social science & medicine. 2005;60(4):833-43.
43. Griffin FA, Resar RK. IHI Global Trigger Tool for measuring adverse events. Institute for Healthcare Improvement Innovation Series White Paper. 2009.
44. Group C. CONSORT Statement Endorsers: Journals and Organizations [Web Page]. Available from: http://www.consort-statement.org/about-consort/endorsers#n.
45. Guilfoyle MR, Seeley H, Laing RJ. The Short Form 36 health survey in spine disease—validation against condition-specific measures. British journal of neurosurgery. 2009;23(4):401-5.
46. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Annals of internal medicine. 1993;118(8):622-9.
47. Hay JL, Atkinson TM, Reeve BB, Mitchell SA, Mendoza TR, Willis G, et al. Cognitive interviewing of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Quality of Life Research. 2014;23(1):257-69.
48. Higgins JPT, Green S. Cochrane Handbook on Systematic Reviews of Interventions 2011 [Web Page]. Available from: http://handbook.cochrane.org/.
49. Hodkinson A, Kirkham JJ, Tudur-Smith C, Gamble C. Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension. BMJ open. 2013;3(9):e003436.
50. Inch J, Watson MC, Anakwe-Umeh S, Collaboration YCS. Patient versus Healthcare Professional Spontaneous Adverse Drug Reaction Reporting. Drug safety. 2012;35(10):807-18.
51. Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. Archives of internal medicine. 2009;169(19):1737-9.
52. Ioannidis JP, Evans SJ, Gøtzsche PC, O'neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Annals of internal medicine. 2004;141(10):781-8.
53. Johnston BC, Supina AL, Vohra S. Probiotics for pediatric antibiotic-associated diarrhea: a meta-analysis of randomized placebo-controlled trials. Canadian Medical Association Journal. 2006;175(4):377-83.
54. Jonsson U, Alaie I, Parling T, Arnberg FK. Reporting of harms in randomized controlled trials of psychological interventions for mental and behavioral disorders: A review of current practice. Contemporary Clinical Trials. 2014;38(1):1-8.

55. Kirsch M, Mitchell SA, Dobbels F, Stussi G, Basch E, Halter JP, et al. Linguistic and content validation of a German-language PRO-CTCAE-based patient-reported outcomes instrument to evaluate the late effect symptom experience after allogeneic hematopoietic stem cell transplantation. European Journal of Oncology Nursing. 2015;19(1):66-74.
56. Lathyris D, Panagiotou OA, Baltogianni M, Ioannidis JP, Contopoulos-Ioannidis DG. Safety of medical interventions in children versus adults. Pediatrics. 2014;133(3):e666-e73.
57. Lipscomb J, Gotay CC, Snyder CF. Patient-reported Outcomes in Cancer: A Review of Recent Research and Policy Initiatives. CA: A Cancer Journal for Clinicians. 2007;57(5):278-300.
58. Longden E, Read J. Assessing and reporting the adverse effects of antipsychotic medication: A systematic review of clinical studies, and prospective, retrospective, and cross-sectional research United States: Lippincott Williams and Wilkins; 2016 [cited 39 1]. 29].
59. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. BMJ: British Medical Journal (Online). 2016;353.
60. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. Journal of evaluation in clinical practice. 2006;12(5):559-68.
61. McLernon DJ, Bond CM, Hannaford PC, Watson MC, Lee AJ, Hazell L, et al. Adverse drug reaction reporting in the UK: a retrospective observational comparison of yellow card reports submitted by patients and healthcare professionals. Drug Safety. 2010;33(9):775-88.
62. McLernon DJ, Bond CM, Lee AJ, Watson MC, Hannaford PC, Fortnum H, et al. Patient views and experiences of making adverse drug reaction reports to the Yellow Card Scheme in the UK. Pharmacoepidemiology and drug safety. 2011;20(5):523-31.
63. McTaggart-Cowan HM, Marra CA, Yang Y, Brazier JE, Kopec JA, FitzGerald JM, et al. The validity of generic and condition-specific preference-based instruments: the ability to discriminate asthma control status. Quality of Life Research. 2008;17(3):453-62.
64. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. International journal of surgery (London, England). 2012;10(1):28-55.
65. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. International journal of surgery (London, England). 2010;8(5):336-41.
66. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. BMC medical research methodology. 2001;1(1):2.
67. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. BMC Medical Research Methodology. 2010;10(1):82.
68. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC medical research methodology. 2010;10(1):22.
69. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of Life Research. 2010;19(4):539-49.
70. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. Journal of clinical epidemiology. 2010;63(7):737-45.
71. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Quality of Life Research. 2009;18(3):313-33.

72. Organization WH. The importance of pharmacovigilance. 2002.
73. Pakhomov S, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. The American journal of managed care. 2008;14(8):530.
74. Paterson C. Measuring outcomes in primary care: a patient generated measure, MYMOP, compared with the SF-36 health survey. Bmj. 1996;312(7037):1016-20.
75. Paterson C, Britten N. In pursuit of patient-centred outcomes: a qualitative evaluation of the 'Measure Yourself Medical Outcome Profile'. Journal of health services research & policy. 2000;5(1):27-36.
76. Pohlman KA, O'Beirne M, Thiel H, Cassidy JD, Mior S, Hurwitz EL, et al. Development and validation of providers' and patients' measurement instruments to evaluate adverse events after spinal manipulation therapy. European Journal of Integrative Medicine. 2014;6(4):451-66.
77. Research CIoH. Guide to knowledge translation planning at CIHR: Integrated and end-of-grant approaches. Canada: Canadian Institutes of Health Research; 2012.
78. Research CIoH. Strategy for Patient-Oriented Research 2016 [cited 2017 5/19/2017]. Available from: http://www.cihr-irsc.gc.ca/e/41204.html.
79. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? Injury. 2011;42(3):236-40.
80. Schunemann HJ, Griffith L, Jaeschke R, Goldstein R, Stubbing D, Austin P, et al. A comparison of the original chronic respiratory questionnaire with a standardized version. CHEST Journal. 2003;124(4):1421-9.
81. Shukralla AA, Tudur-Smith C, Powell GA, Williamson PR, Marson AG. Reporting of adverse events in randomised controlled trials of antiepileptic drugs using the CONSORT criteria for reporting harms. Epilepsy research. 2011;97(1):20-9.
82. Sloan JA, Berk L, Roscoe J, Fisch MJ, Shaw EG, Wyatt G, et al. Integrating patient-reported outcomes into cancer symptom management clinical trials supported by the National Cancer Institute–sponsored clinical trials networks. Journal of Clinical Oncology. 2007;25(32):5070-7.
83. Snyder C, Wu A. Putting the Patient Perspective in Patient-Centered Outcomes Research. 2012.
84. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-reported outcomes (PROs): putting the patient perspective in patient-centered outcomes research. Medical care. 2013;51(8 0 3):S73.
85. Tefferi A, Hudgens S, Mesa R, Gale RP, Verstovsek S, Passamonti F, et al. Use of the Functional Assessment of Cancer Therapy--anemia in persons with myeloproliferative neoplasm-associated myelofibrosis and anemia. Clinical therapeutics. 2014;36(4):560.
86. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of clinical epidemiology. 2007;60(1):34-42.
87. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Quality of Life Research. 2009;18(8):1115-23.
88. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of Life Research. 2012;21(4):651-7.
89. Terwee CBea. COSMIN definitions of domains, measurement properties, and aspects of measurement properties 2014 [cited 2016 February 24th, 2016]. Available from: http://www.cosmin.nl/images/upload/files/Tabel%20met%20definities-new.pdfhttp://www.cosmin.nl/images/upload/files/Tabel met definities-new.pdf.
90. Tönshoff B. Ethical conduct of clinical research involving children. Marilyn J. Field and Richard E. Behrman for the Committee on Clinical Research Involving Children of the Institute of Medicine, The

National Academies Press, Washington, DC, USA, 2004, 425 pp., USD 47.95, ISBN 0-309-09181-0. The Quality Assurance Journal. 2005;9(2):146-.
91. Trotti A, Colevas AD, Setser A, Basch E. Patient-reported outcomes and the evolution of adverse event reporting in oncology. Journal of Clinical Oncology. 2007;25(32):5121-7.
92. Vohra S, Cvijovic K, Boon H, Foster BC, Jaeger W, LeGatt D, et al. Study of natural health product adverse reactions (SONAR): active surveillance of adverse events following concurrent natural health product and prescription drug use in community pharmacies. PLoS One. 2012;7(9):e45196.
93. Wagland R, Richardson A, Armes J, Hankins M, Lennan E, Griffiths P. Treatment-related problems experienced by cancer patients undergoing chemotherapy: a scoping review. European Journal of Cancer Care. 2015;24(5):605.
94. Ware Jr JE. SF-36 health survey update. Spine. 2000;25(24):3130-9.
95. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Medical care. 1992:473-83.
96. Wei JT, Dunn RL, Litwin MS, Sandler HM, Sanda MG. Development and validation of the expanded prostate cancer index composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer. Urology. 2000;56(6):899-905.
97. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. Journal of clinical epidemiology. 2003;56(1):52-60.
98. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012;13(1):132.
99. Williamson PR, Altman DG, Blazeby JM, Clarke M, Gargon E. The COMET (core outcome measures in effectiveness trials) initiative. Trials. 2011;12(Suppl 1):A70.
100. Wood JM. Understanding and Computing Cohen's Kappa: A Tutorial. WebPsychEmpiricist Web Journal at http://wpe info/. 2007.
101. Zorzela L, Golder S, Liu Y, Pilkington K, Hartling L, Joffe A, et al. Quality of reporting in systematic reviews of adverse events: systematic review. 2014.
102. Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. bmj. 2016;352:i157.