Journal of Clinical Epidemiology 136 (2021) 26–36

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Patient-reported outcome measures in core outcome sets targeted overlapping domains but through different instruments

Oriana Ciani[a,b,*], Maximilian Salcher-Konrad[c,d], Michela Meregaglia[a], Kathrine Smith[e], Sarah L. Gorst[f], Susanna Dodd[f], Paula R. Williamson[f], Giovanni Fattore[a,g]

[a] *Centre for Research on Health and Social Care Management, SDA Bocconi, via Sarfatti 10, 20136, Milan, Italy*
[b] *Evidence Synthesis and Modeling for Health Improvement, College of Medicine and Health, University of Exeter, EX1 2LU, Exeter, UK*
[c] *LSE Health, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK*
[d] *Care Policy and Evaluation Centre, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK*
[e] *Imperial College, London, UK*
[f] *MRC/NIHR Trials Methodology Research Partnership, Department of Health Data Science, University of Liverpool, Liverpool L69 3BX, UK*
[g] *Department of Social and Political Science, Bocconi University, via Sarfatti 36, 20136, Milan, Italy*

Accepted 2 March 2021; Available online 6 March 2021

Abstract

Objective: There is no comprehensive assessment of which patient-reported outcomes (PROs) are recommended in core outcome sets (COS), and how they should be measured. The aims of this study are to review COS that include patient-reported outcomes measures (PROMs), identify their target health domains, main characteristics, and their overlap within and across different disease areas.

Study design and setting: We selected COS studies collected in a publicly available database that included at least one recommended PROM. We gathered information on study setting, disease area, and targeted outcome domains. Full-text of recommended instruments were obtained, and an analysis of their characteristics and content performed. We classified targeted domains according to a predefined 38-item taxonomy.

Results: Overall, we identified 94 COS studies that recommended 323 unique instruments, of which: 87% were included in only one COS; 77% were disease-specific; 1.5% preference-based; and 61% corresponded to a full questionnaire. Most of the instruments covered broad health-related constructs, such as global quality of life (25%), physical functioning (22%), emotional functioning and wellbeing (7%).

Conclusion: The wealth of recommended instruments observed even within disease areas does not fit with a vision of systematic, harmonized collection of PROM data in COS within and across disease areas.

Keywords: Core outcome set; COS; Patient-reported outcome measure; PRO; PROM; Outcomes research

1. Introduction

The last years have witnessed an increased commitment by the health services research and professional community to involving patients and the public in the development, delivery and evaluation of health care services [1]. In parallel, the assessment of patient-reported outcomes (PROs) has gained international traction as one of the enablers of patient-centered healthcare. Regulatory agencies define a PRO as any report of the status of a patient's health condition that comes directly from the patient, without interpretation by a practitioner or anyone else [2,3]. PROs complement existing information and physical examinations by providing standardized assessments of how patients function or feel with respect to their health, quality of life, mental well-being, or satisfaction with the healthcare process [4]. Hence, whether in clinical research or in the context

Conflict of interest: The authors do not declare any conflict of interest in relation to the present study.

Authors' contributions: O.C., M.S.K., M.M. contributed to the conceptualization and methodology. O.C., M.S.K., M.M., K.S., S.G., S.D. performed the analyses. O.C., M.S.K., M.M. wrote the original draft. K.S., S.G., S.D., P.W., G.F. reviewed and edited the draft. G.F. acquired financial support for the project.

Funding: This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 116055 (Big Data for Better Outcomes DO-IT). This Joint Undertaking receives support from the European Union's Horizon 2020 Research and Innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA). This work was also supported by a National Institute for Health Research (NIHR) Senior Investigator Award (NF-SI_0513-10025) and the MRC Trials Methodology Research Partnership (grant reference MR/S014357/1).

* Corresponding author. Tel.: +39-02-5836-5248.
E-mail address: oriana.ciani@unibocconi.it (O. Ciani).

https://doi.org/10.1016/j.jclinepi.2021.03.003

of routine clinical encounter, PROs are typically collected using formally designed and validated questionnaires. A patient-reported outcome measure (PROM) is defined as an instrument, scale, or single-item measure used to assess a range of relevant health domains as perceived and self-reported by the patient [5].

However, several issues related with the collection (e.g., length, assessment schedule), analysis (e.g., missing data, multiple testing), reporting (e.g., cherry-picking of results) and interpretation (e.g., cut-off scores, clinically meaningful thresholds) of PROs may jeopardize their effective use and dissemination in clinical research [6].

One of the challenges relates to the availability of many validated PROMs, which means that different outcomes are reported from a multiplicity of items and scales, often making use of non-standardized terminology, developed by different groups and disciplines (for example, clinical versus psychological) or for differing purposes (for example, measurement of health in generic populations versus disease-specific patient groups) [7]. This problem hinders comparison of PROMs and data synthesis across research studies.

A potential solution to this challenge lies in the development and use of core outcome sets (COS). A COS is a minimum set of outcomes to be measured in any trial and, increasingly, also in observational studies and clinical practice in a given disease area, with the aim of improving the efficiency of the research process and transparency in reporting of results. Agreement on a minimum set of outcomes and how these should be measured enables comparison and synthesis of results across research studies [8]. PROs are becoming increasingly common components of COS. In the Core Outcome Measures in Effectiveness Trials (COMET) database, which is a public online database collecting COS development studies across all disease areas, outcomes relating to 'life impact' were recommended in 43% of 299 COS published before 2016 [9].

When selecting instruments for measuring recommended PROs, COS developers should examine the validity and quality of PROMs, their content and similarities across instruments, scales and items, in order to reduce duplication and promote efficiency whilst preserving patient involvement throughout the entire process [7]. Whilst there is some evidence about methods used to select outcome measurement instruments generally [10], there is no comprehensive assessment of which PROs are proposed and how they are recommended to be measured in COS.

Therefore, the aims of this study are: (1) to review COS for research that include PROMs, and identify their target health domains, (2) to describe the main characteristics of recommended PROMs, (3) to assess the amount of overlap of recommended PROMs across COS for research, and (4) to explore approaches used for selecting PROMs and specific considerations made by COS developers when recommending PROMs.

2. Materials and methods

This study is based on a cross-sectional analysis of COS studies included in a comprehensive database initiated and maintained by the University of Liverpool as part of the COMET Initiative, which promotes the development and uptake of COS across a wide range of disease areas (http://www.comet-initiative.org). This review includes all COS studies published up to the end of 2018 [11,8]. It is important to point out that studies do not have to be explicitly identified as "core outcome set" to be eligible for inclusion in the COMET database, rather they are included if they have developed or applied methodology for determining which outcome domains or outcomes should be measured in clinical trials or other forms of health research.

3. Inclusion and exclusion criteria

Out of the list of COS development studies identified from the original COMET systematic review and annual updates [11], we selected those studies that included at least one PRO amongst the recommended outcomes and made a recommendation on how to measure it (i.e., specified a PROM).

We excluded COS development studies that only provided a recommendation on the core outcome domains (what to measure) without clarifying how to measure them, or studies that discussed how to measure PROs without a clear endorsement for one of the measures reviewed. For the purpose of this analysis, we considered the definition of PRO given above [2,3], hence we excluded caregiver-, surrogate- or proxy-reported outcomes.

4. Study selection

All identified articles were then screened independently by two researchers (M.S.K., K.S.) to identify those including PROs. Disagreements were discussed to reach consensus between the two reviewers on whether the study fulfilled the eligibility criteria for this analysis.

5. Data extraction

Data extraction was initially piloted on five randomly selected COS studies [12-16] by four reviewers (OC, MM, MSK, KS) to finalize an Excel spreadsheet for standardized collection of relevant information. The template included descriptive characteristics of each COS development study (setting; disease area – as defined by COMET; target population), names of recommended PROMs with accompanying information (full questionnaire, stand-alone question, subscale of an existing questionnaire; generic or disease-specific; administration mode; specific PROMs selection methods; target domain as reported by COS developers). We considered generic PROMs as those designed to be

applicable across a wide range of populations and interventions, whilst specific measures are designed to be relevant to particular interventions or in certain subpopulations [17]. Moreover, we classified PROMs as preference-based when yielding preference weights for quality-adjusted life years calculations [18,19]. Because terminology is not universally agreed in this field, we extracted verbatim the target domain of the PROM as reported by COS developers, and then matched this to a 38-item taxonomy for outcome classification [9]. This taxonomy was developed for the classification of outcomes included in all trials, COS, systematic reviews, and trial registries. It is based on top level ''core areas'' common to other outcome hierarchies but provides a more detailed taxonomy appropriate for all potential outcomes, from physiological/clinical (organ-, system- or disease- specific) outcome domains to life impact (including all forms of functioning), mortality, adverse events and resource use outcome domains.

For each unique PROM recommended, we conducted online searches to obtain the full text (unless already included in the COS development studies). We recorded whether full-texts could not be found or could not be obtained free of charge. Additional detailed information was collected based on full-texts of recommended PROMs (structure of the PROM; number of items; verbatim items). Data extraction was double-checked by at least one reviewer.

6. Data analysis

We summarized occurrence of PROMs recommended and their characteristics across COS development studies by means of descriptive statistics and bar charts. We used cross-tabulation to investigate what target domains are recommended across COS development studies and across disease areas. Bubble charts were used as a generalization of the scatter plot to display the relationship between the outcome taxonomy classification, disease area, and the number of PROMs recommended. We reported results for characteristics of all recommended PROMs. Additional thematic analysis on the content of individual items of recommended PROMs was conducted to explore further overlap by outcome domains within disease areas.

7. Results

7.1. Characteristics of COS studies

Out of 337 COS development studies screened, we included 94 COS development studies that included PROs with recommendations on both what (i.e., which target domains) and how (i.e., which PROM) to measure target domains (Fig. 1, list available in Supplementary Material). Cohen's kappa for agreement about inclusion was 0.83. Included COS studies spanned 26 different disease areas, with some more frequently investigated than others:

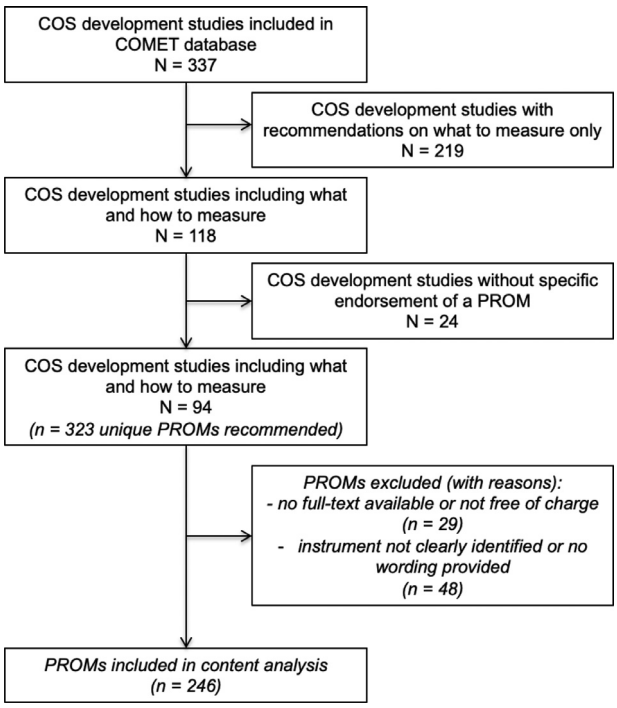

Fig. 1. Flow chart of COS studies and recommended instruments selection process.

there were 16 COS in Neurology, 11 in Rheumatology, 10 in Orthopedics & Trauma, 9 in Lungs & Airways, 8 in Heart & Circulation, and 7 in Cancer. Included COS studies were published in the 1990s (n = 11), first decade of 2000 (n = 33) and up to 2018 (n = 50).

7.2. Outcome domains targeted by COS developers for recommended PROMs

'Global quality of life' was found to be the most frequently targeted domain across disease areas, followed by physical functioning, and emotional functioning and well-being (Fig. 2). Recommended PROMs were mostly used to capture life impact and functioning domains, rather than physiological/clinical (organ-, system- or disease- specific) outcome domains [9], with some variation across disease areas. Fig. 3 shows how often each domain was targeted by recommended instruments in each of the 26 disease areas examined in the COS studies.

7.3. Characteristics of PROMs recommended

The total number of unique PROMs recommended across these 94 COS development studies was 323. These include variations of existing PROMs, such as the Short Form-12 (SF-12) as a shortened version of the Short Form-36 (SF-36)). When these variations were considered as if related to the same measure, the total number of unique PROMs recommended was 243.

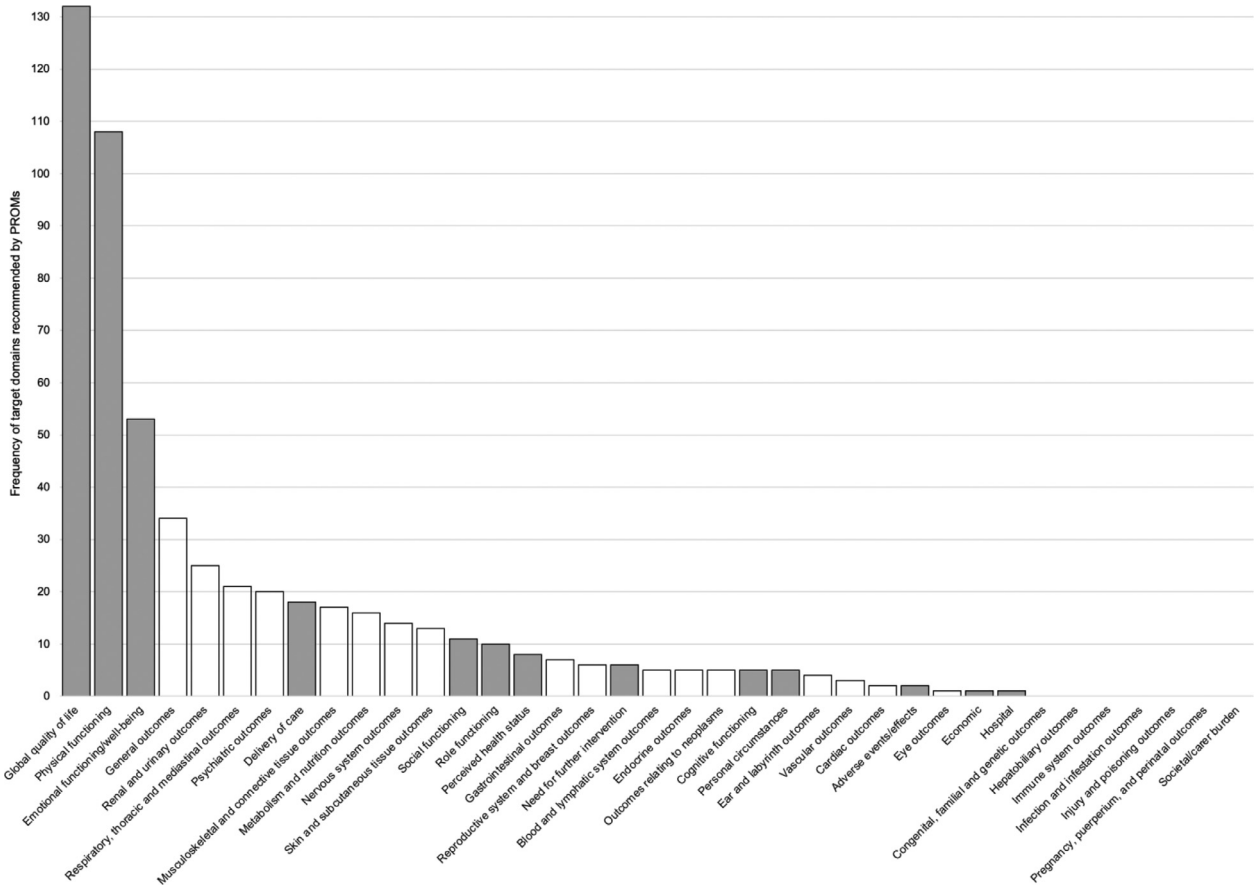

Fig. 2. Outcome domains targeted by recommended PROMs across disease areas. White bars relate to physiological/clinical (organ-, system-, or disease- specific) outcome domains. Shaded bars relate to life impact, functioning, resource use outcome domains.

Overall, the majority of recommended PROMs were disease-specific (248, 77%), that is, they were designed to be relevant to a specific disease, population, function or condition [19] (Table 1). Overall, five preference-based PROMs (1.5%) were identified.

The majority of recommended PROMs were full questionnaires (61%). By contrast, a significant minority (18%) were questions or subscales of existing measures (e.g., the bodily pain subscale of the SF-36, or the cognition subscale of the Functional Independence Measure [FIM]). In most cases, no information was given by COS developers regarding the validation of subscales to be used instead of the full questionnaire. In one study [12], it was explicitly mentioned that the bodily pain subscale of SF-36 had a strong psychometric support and extensive normative data. In other studies, there was a generic reference to subscale validation [13] or a statement that single items were believed to be more informative, and even more statistically sensitive, than the overall questionnaires [20]. Lastly, one study [21] provided a list of referenced studies that developed and tested individual components of a full PROM (i.e. the PROMIS-29 questionnaire).

More than one in five recommended PROMs (21%) were single questions, i.e. either individual questions taken out of existing questionnaires, or stand-alone questions that COS developers deemed relevant. These were often numerical rating scales (NRS) or visual analogue scales (VAS) for pain (35 out of 68 single questions). Stand-alone questions or questions forming part of larger questionnaires were not always reported to be validated. In one study, the authors acknowledged missing validation, but stated that the committee still deemed PROMs "extremely important" and therefore recommended them [13]. Some stand-alone questions were recommended without clear indication of wording to use (e.g., recommended to self-report overall uveitis-related disability assessment by children [22], patient global score [23] or use of rescue analgesics [24]). Content analysis was not possible for these questions.

7.4. Overlap of recommended PROMs

The vast majority of the 323 PROMs was recommended in only one COS study (280, 87%). Twenty-three PROMs (7%) were recommended in 2 COS studies, 8 (2.5%) were recommended in 3 COS studies, and 6 (2%) were recommended in 4 COS studies (Table 1). Half of the 20 PROMs recommended in 3 or more COS were generic.

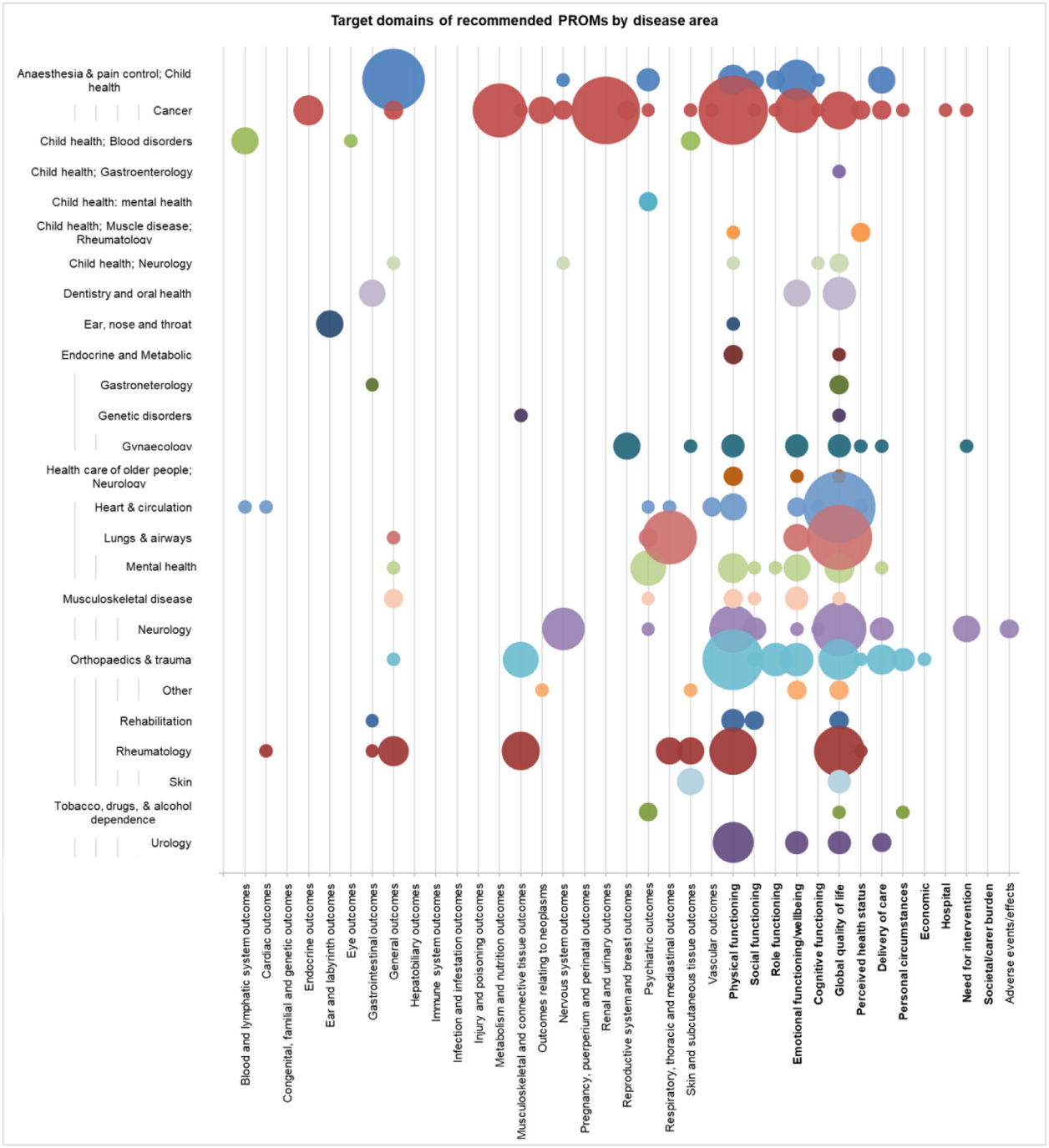

Fig. 3. Outcome domains targeted in recommend PROMs by disease area. Figure shows 37 outcome domains (excluding mortality) on the horizontal axis targeted by 323 unique recommended instruments by disease area (vertical axis). Life impact, functioning, and resource use outcome domains are highlighted in bold. Bubble sizes indicate frequency of an individual domain being the target domain of a recommended PROM in the disease area. For example, in COS for urology, physical functioning was the target domain of 9 recommended PROMs, emotional functioning/wellbeing was the target domain of 3 recommended PROMs, global quality life was the target domain of 3 recommended PROMs, and delivery of care was the target domain of 2 recommended PROMs. Details provided in Supplementary Material.

The average number of unique PROMs recommended per COS was 4.5 (median 3, minimum 1, maximum 17). Most frequently, only one PROM was recommended (n = 18 COS), but on the other end, there were 9 COS studies including recommendations for 10 or more PROMs to be used (Fig. 4).

There were six disease areas with five or more COS studies recommending PROMs (i.e., neurology, rheumatology, orthopedics & trauma, lungs & airways, cancer, and heart & circulation). Fig. 5 shows the number and names of PROMs that were recommended in more than one COS for each of those disease areas. For example, there were

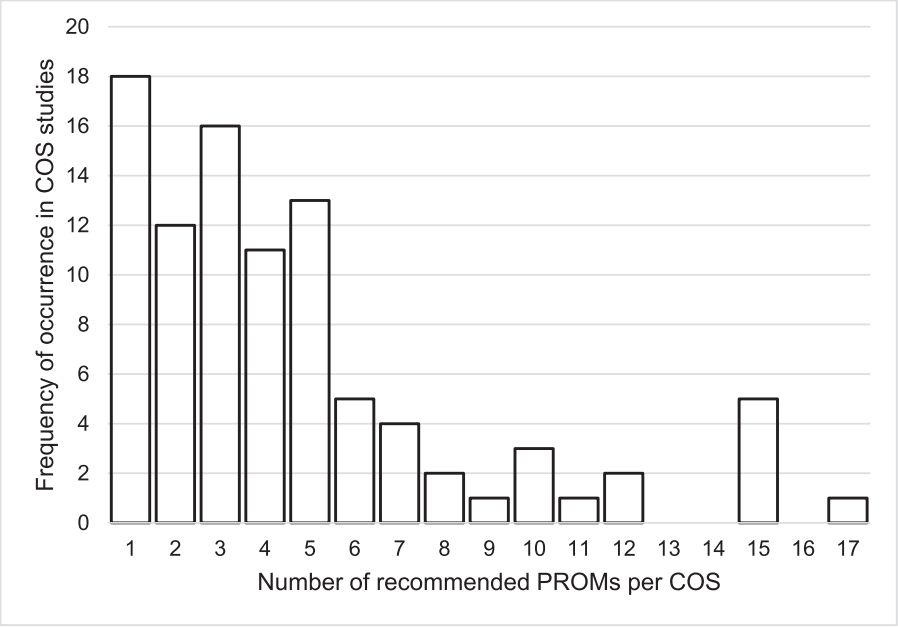

Fig. 4. Distribution of number of PROMs recommended in a COS study.

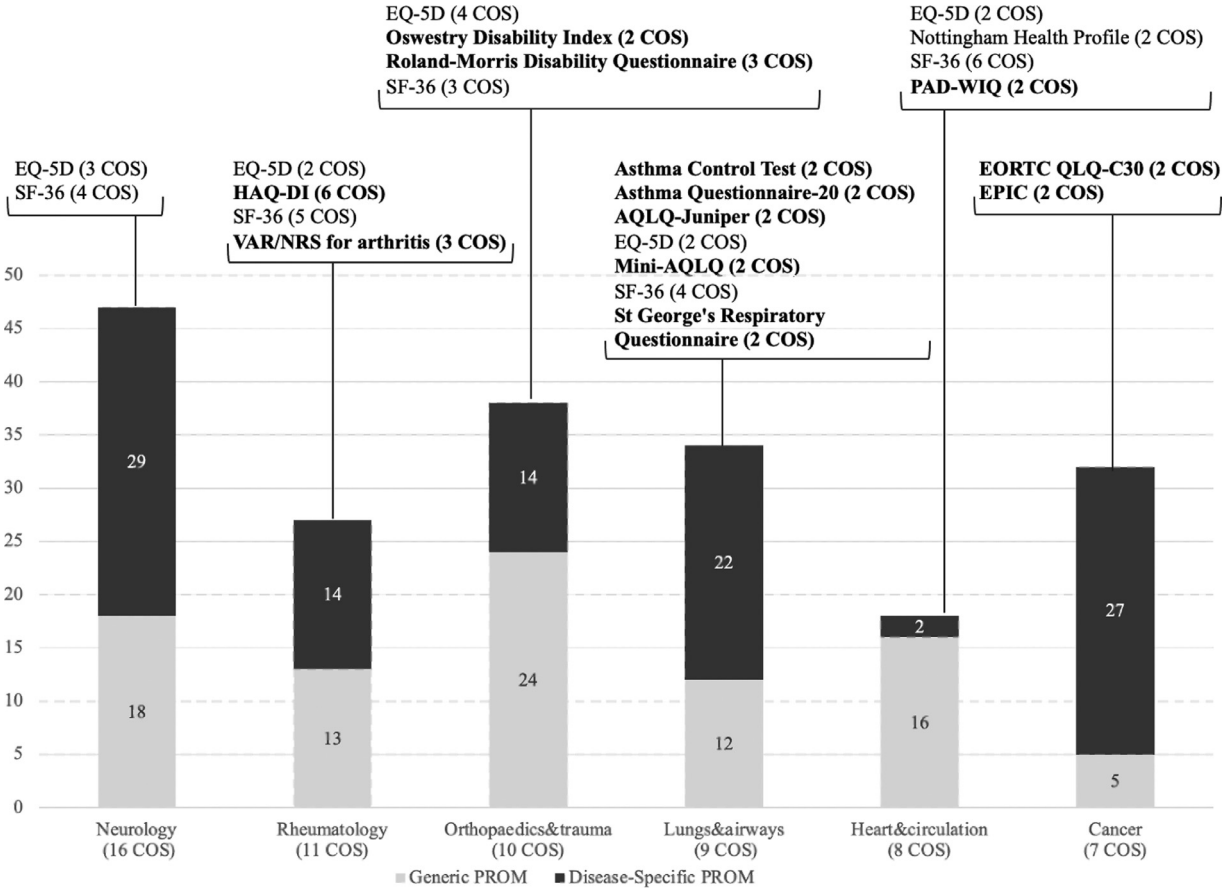

Fig. 5. Combination of generic and disease-specific PROMs in disease areas with five or more COS studies, with an indication of PROMs recommended in more than one COS (bold for disease-specific PROMs).

Table 1. Characteristics of recommended PROMs

	No. of PROMs (% of total)
Unique PROMs recommended *of which total no. of PROMs excluding variants*	323 (100%) *243 (75%)*
PROMs recommended in 1 COS	280 (86.5%)
PROMs recommended in 2 COS	23 (7%)
PROMs recommended in 3 COS	8 (2.5%)
PROMs recommended in 4 COS	6 (2%)
PROMs recommended in > 4 COS	6 (2%)
- SF-12 (in 5 COS)	
- HAQ-DI (in 6 COS)	
- NRS or NRS for pain intensity (in 11 COS)	
- EQ-5D-3L (in 15 COS)	
- SF-36 (in 27 COS)	
Generic PROMs	75 (23%)
- of which preference-based (i.e. EQ-5D, HUI3, QWB, SF-6D)	*4 (1.2%)*
Disease-specific PROMs	248 (77%)
- of which preference-based (i.e. PORPUS)	*1 (0.3%)*
Full questionnaires	196 (61%)
Questions (or subscales) of existing PROMs	59 (18%)
Single questions	68 (21%)
- of which NRS or VAS	*35 (11%)*
- of which other single questions	*33 (10%)*

COS = Core Outcome Set; EQ-5D: EuroQol-5 Dimension; HAQ-DI = Health Assessment Questionnaire Disability Index; HUI = Health Utility Index; NRS = Numeric Rating Scale; QWB = Quality of Wellbeing; PORPUS = Patient-Oriented Prostate Utility Scale; SF = Short Form; VAS = Visual Analogue Scale

16 COS for a neurological condition, including four COS for headaches and migraine [25-28], two for amyotrophic lateral sclerosis [29,30], and one each for ischemic stroke [31], cerebral palsy [32], insomnia [33], peripheral neuropathy [34], Charcot-Marie-Tooth disease [35], traumatic brain injury [36], multiple sclerosis [37], post-stroke aphasia [38], mild or moderate dementia [39], and sensorimotor recovery after stroke [40]. In total, 47 unique PROMs were recommended, but only two PROMs were recommended more than once: SF-36 (in 4 COS) and EQ-5D (in 3 COS).

Content analysis of recommended PROMs within individual disease areas showed thematic overlap by outcome domains. Table 2 illustrates how items relating to the physical functioning domain were worded in PROMs recommended in neurology COS. Within the physical functioning domain, recommended PROMs asked about the ability of the patient to wash themselves (4 unique PROMs), the ability to walk (5 unique PROMs), and other, disease-specific functioning aspects.

7.5. Methods used for selecting PROMs and specific considerations by COS developers

Occasionally, authors provided details specifically on the methodology adopted in selecting PROMs. An OMERACT group reported that PROMs were discussed separately from other outcomes included in the COS by organizing a dedicated break-out session attended by patient research partners, in addition to researchers and clinicians [41], a group composition that differed from those of other break-out sessions which did not involve patients. Among the available techniques for generating consensus on PROM selection, the Delphi technique by using mailed or online surveys was reported in few cases [13,32,42]; in another study [43], Delphi was used for gaining consensus from healthcare professionals, while patients were separately involved in a focus group.

The justification given to support the choice of the recommended PROMs was generally in line with the principle of "standard practice" being the measure commonly adopted in clinical studies of a specific condition. Sometimes COS developers acknowledged lack of "superior tools." In a few studies, the authors made reference to some forms of validation and/or reliability, described as "internal consistency," "discrimination," "test-retest," or "concurrent/convergent validity," "divergent validity," "discriminant validity," "content validity" [44-46]. Other criteria mentioned for PROMs selection, were appropriateness; responsiveness; comprehensiveness; interpretability; precision of scores; acceptability; burden and feasibility; availability and equivalence of alternate forms and

Table 2. PROMs items mapped to physical functioning domain in neurology

Physical functioning	Example items
Washing oneself	- Does your health now limit you in these activities? If so, how much? - Bathing or dressing yourself (SF-36) - Does Patient need help when washing, rinsing or drying the body? (FIM, Functional Independence Measure - Motor Subscale) - Self-care (have no problems with self-care/some problems washing or dressing myself/ I am unable to wash or dress myself) (EQ-5D-3L) - During the past week, how much trouble did you have taking a bath or shower? (Stroke and Aphasia Quality of Life Scale, SAQOL-39)
Walking	- Do you have any problem with your walking? (ODSS, Overall disability sum score) - How do you usually get around for about 10 metres? (Without aid /With one stick or crutch or holding to someone's arm/ With two sticks or crutches or one stick or crutch and holding to someone's arm/ With a wheelchair) (ODSS, Overall disability sum score) - Does your health now limit you in these activities? If so, how much? - i. Walking one block (SF-36) - Mobility (I have no problems/some problems in walking about/I am confined to bed) (EQ-5D-3L) - Does Patient need help to walk 150 feet (50 m)/go 150 feet (50m) in a wheelchair? (FIM, Functional Independence Measure – Motor Subscale) - During the past week, how much trouble did you have walking? (Stroke and Aphasia Quality of Life Scale, SAQOL-39) - During the past week, how much trouble did you have walking without stopping to rest, or using a wheelchair without stopping to rest? (Stroke and Aphasia Quality of Life Scale, SAQOL-39)
Other disease-specific aspects	- Compared with before the accident, do you now (i.e., over the last 24 hours) suffer from sleep disturbance? (RPQ, Rivermead Post Concussion Questionnaire) - I restrict my recreational activities because of my headache (HDI, Headache Disability Inventory) - Your ability to keep up physically with your peers? (CP QOL-Child, Cerebral Palsy Quality of Life Questionnaire for Children) - In the past two weeks, how much has your MS limited your ability to grip things tightly (e.g. turning on taps)? (MSIS-29, Multiple Sclerosis Impact Scale)

methods of administration (e.g., self-report, interviewer); and availability and equivalence of versions for different cultures and languages [47-51].

8. Discussion

In this study we mapped recommendation of PROMs as part of standardized outcome collection in clinical research across various conditions. By searching a comprehensive database of COS studies, we included 94 studies that spanned 26 different disease areas, with some more frequently represented in the sample (e.g., neurology, rheumatology, lungs & airways, cancer). This might reflect different attitudes and familiarity across specialties in dealing with PROs type of outcome measures.

Most of the PROMs (25%) covered "'global quality of life," physical functioning (22%), followed by emotional functioning and wellbeing (7%). These are broad health-related constructs that appear consistently targeted across disease areas, yet a broad range of different PROMs was recommended to measure them. Despite the intended aim of striving for harmonization and alignment of outcomes, we found a fragmented landscape of recommended PROMs in COS for clinical trials. A total of 323 unique PROMs were recommended for use. The vast majority (87%) of those were recommended in only one COS, and individual COS recommended a median of more than 4 instruments (range, 1-17) each. This finding may reflect long-standing use of PROs in some conditions or a more holistic view of the impact a disease has on patients' lives in some disease areas rather than others.

Although some of our findings may be explained by the heterogeneity of disease areas, which encourages development of disease-specific measures that are preferred to generic ones, there is still scope for increased harmonization of outcome measurement even within disease areas. An example of such efforts is the Patient-Reported Outcomes Measurement Information System (PROMIS) by the US National Institutes of Health; PROMIS uses item response theory to create standardized PROMs on the basis of existing items [52], which can be used across disease areas. However, PROMIS measures do not appear to be commonly used in COS: in our sample of 94 COS only 3 recommended standardized PROMIS measures [21,53,54].

Almost one in four recommended PROMs were generic, meaning that they were, in principle, applicable to a wider population, regardless of any existing condition or state, and therefore of potential relevance across several COS. Indeed, we found that generic measures were more likely to be recommended in 3 or more COS, and therefore contributed to the harmonization of outcome measurement across disease areas. Only five (1.5%) PROMs were preference-based, of which four were generic, and therefore useful for cost-utility analyses.

Finally, we identified some issues with methodological quality of selecting PROMs. Firstly, the total number of questions to be answered by a single patient was more than 100 for more than a quarter of all COS. Whilst some

of the recommended PROMs might be relatively quick to administer, these figures suggest that COS developers might not have routinely considered the burden for patients to complete multiple questionnaires. While standards have been published that call for the inclusion of patients in the COS development process [55], these are still relatively recent and would not have been available during the development of COS included in our sample. Furthermore, we found that evidence of empirical performance (e.g., responsiveness, discriminant validity and other validations) of recommended PROMs was not commonly reported. Considerations about ceiling effect or floor effect in the distribution of scores were not reported, although it should be noted instruments' performance is more often assessed with generic preference-based measures (e.g., EQ-5D, SF-6D), and even among them lack of head-to-head comparisons and poor reporting impede thorough comparative evaluation [56].

More than 60% of recommended PROMs were full-questionnaires, 18% were subscales or items from full-questionnaires and 21% were single-item or stand-alone questions, mostly NRS or VAS. Evidence on validation was mostly lacking for stand-alone questions, despite establishing sufficient psychometric validation should be one of the standard criteria for including/dropping/adding PROMs from a COS. Moreover, we were not able to track the exact wording of the chosen question and answer in a significant number of cases (n = 48). This is an interesting finding, because without a clear statement on how the question and possible answers should be framed with instructions to give to the patient, the COS might fail to achieve its intended purpose, that is to standardize outcome measurement and ensure comparability.

This study has limitations. We did not assess the methodological rigor, as Gorst et al did in a previous publication [10], with which COS were developed and whether COS developers followed the COMET and COSMIN guidance [55,57] to first identify the domains to be measured, and in the next step identify and evaluate potential measures for these.

Striving for efficient PRO data collection is an imperative in the context of rising use of e-health solutions for clinical research, audit and quality improvement. Patients, as well as health care professionals administering PROMs, are unlikely or unable to spend a considerable amount of time filling out questionnaires. In a recent mapping exercise from COS studies in prostate cancer to existing real world data sources, we found that self-reported outcome measures are a dimension not typically covered in routinely collected data sources [58]. However, the current technological landscape would allow for a wide scale, standardized, continuous collection of PROMs [2]. Of course, issues of interoperability, data governance, security, privacy, logistics and ethics must be addressed in advance but incorporation in routinely collected data of the voice, preferences, and experience of the patient is theoretically possible locally, regionally, and even nationally. The wealth of recommended PROMs observed, even within disease areas, the use of single questions, often developed ad hoc and without proper validation, does not fit with a vision of systematic, harmonized collection of PROM data and reveal lack of a much-needed effort to agree on standardized measurement tools across key target domains. This aspect goes beyond the trial setting, and applies to a range of applications (e.g., general population surveys, disease registries) whose results include norms and other evidence useful in interpreting health status and outcomes.

The community of COS developers and methodologists is called upon to solve the problems unveiled by this study, by improving methods for making recommendations about how to measure the core outcomes, and leveraging on the consensus-based standards for the selection of health measurement instruments (COSMIN) initiative.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jclinepi.2021.03.003.

References

[1] Weldring T, Smith SM. Patient-reported outcomes (PROS) and patient-reported outcome measures (PROMs). Health Serv Insights 2013;6:61–8. doi:10.4137/HSI.S11093.

[2] U.S. Food And Drug Administration. Patient-reported outcome measures: use in medical product development to support labeling claims 2009 Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims], 17/11/2020.

[3] European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man - the use of patient-reported outcome (PRO) measures in oncology studies 2016 Available at: https://www.ema.europa.eu/en/appendix-2-guideline-evaluation-anticancer-medicinal-products-man-use-patient-reported-outcome-pro, 17/11/2020.

[4] Lavallee DC, Chenok KE, Love RM, Petersen C, Holve E, Segal CD, et al. Incorporating patient-reported outcomes into health care to engage patients and enhance care. Health Aff (Millwood). 2016;35(4):575–82. doi:10.1377/hlthaff.2015.1362.

[5] Basch E, Torda P, Adams K. Standards for patient-reported outcome-based performance measures. JAMA 2013;310(2):139–40. doi:10.1001/jama.2013.6855.

[6] Ciani O, Federici CB. value lies in the eye of the patients: the why, what, and how of patient-reported outcomes measures. Clin Ther 2020;42(1):25–33 Epub 2020 Jan 10. doi:10.1016/j.clinthera.2019.11.016.

[7] Macefield RC, Jacobs M, Korfage IJ, Nicklin J, Whistance RN, Brookes ST, et al. Developing core outcomes sets: methods for identifying and including patient-reported outcomes (PROs). Trials 2014;15:49. doi:10.1186/1745-6215-15-49.

[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. doi:10.1371/journal.pone.0099111.

[9] Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. J Clin Epidemiol 2018;96:84–92 Epub 2017 Dec 28. doi:10.1016/j.jclinepi.2017.12.020.

[10] Gorst SL, Prinsen CAC, Salcher-Konrad M, Matvienko-Sikar K, Williamson PR, Terwee CB. Methods used in the selection of instruments for outcomes included in core outcome sets have improved since the publication of the COSMIN/COMET guideline. J Clin Epidemiol 2020;125:64–75 Epub 2020 May 26. doi:10.1016/j.jclinepi.2020.05.021.

[11] Gargon E, Gorst SL, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 5th annual update to a systematic review of core outcome sets for research. PLoS One 2019;14(12):e0225980. doi:10.1371/journal.pone.0225980.

[12] Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine (Phila Pa 1976) 2000;25(24):3100–3. doi:10.1097/00007632-200012150-00003.

[13] Khanna D, Lovell DJ, Giannini E, Clements PJ, Merkel PA, Seibold JR, et al. Development of a provisional core set of response measures for clinical trials of systemic sclerosis. Ann Rheum Dis 2008;67(5):703–9 Epub 2007 Sep 24. doi:10.1136/ard.2007.078923.

[14] Clements PJ, Allanore Y, Khanna D, Singh M, Furst DE. Arthritis in systemic sclerosis: systematic review of the literature and suggestions for the performance of future clinical trials in systemic sclerosis arthritis. Semin Arthritis Rheum 2012;41(6):801–14 Epub 2011 Dec 15. doi:10.1016/j.semarthrit.2011.10.003.

[15] Merkel PA, Aydin SZ, Boers M, Direskeneli H, Herlyn K, Seo P, et al. The OMERACT core set of outcome measures for use in clinical trials of ANCA-associated vasculitis. J Rheumatol 2011;38(7):1480–6. doi:10.3899/jrheum.110276.

[16] Hellmich B, Flossmann O, Gross WL, Bacon P, Cohen-Tervaert JW, Guillevin L, et al. EULAR recommendations for conducting clinical studies and/or clinical trials in systemic vasculitis: focus on anti-neutrophil cytoplasm antibody-associated vasculitis. Ann Rheum Dis 2007;66(5):605–17 Epub 2006 Dec 14. doi:10.1136/ard.2006.062711.

[17] Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. Pharmacoeconomics 2000;17(1):13–35. doi:10.2165/00019053-200017010-00002.

[18] Brazier J, Ara R, Rowen D, Chevrou-Severac H. A review of generic preference-based measures for use in cost-effectiveness models. Pharmacoeconomics 2017;35(Suppl 1):21–31. doi:10.1007/s40273-017-0545-x.

[19] Rowen D, Brazier J, Ara R, Azzabi Zouraq I. The role of condition-specific preference-based measures in health technology assessment. Pharmacoeconomics 2017;35(Suppl 1):33–41. doi:10.1007/s40273-017-0546-9.

[20] Vocci F, de Wit H. Consensus statement on evaluation of outcome of pharmacotherapy for substance abuse/dependence: report from a NIDA/CPDD meeting, Bethesda, MD: National Institute on Drug Abuse Medications Development Division; 1999. Available at: http://www.webcitation.org/5zXE1Axw7 17/11/2020.

[21] Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, et al. Report of the NIH Task Force on research standards for chronic low back pain. J Pain 2014;15(6):569–85 Epub 2014 Apr 29. doi:10.1016/j.jpain.2014.03.005.

[22] Heiligenhaus A, Foeldvari I, Edelsten C, Smith JR, Saurenmann RK, Bodaghi B, et al. Multinational Interdisciplinary Working Group for Uveitis in Childhood. Proposed outcome measures for prospective clinical trials in juvenile idiopathic arthritis-associated uveitis: a consensus effort from the multinational interdisciplinary working group for uveitis in childhood. Arthritis Care Res (Hoboken) 2012;64(9):1365–72. doi:10.1002/acr.21674.

[23] Saketkoo LA, Mittoo S, Frankel S, LeSage D, Sarver C, Phillips K, et al. Reconciling healthcare professional and patient perspectives in the development of disease activity and response criteria in connective tissue disease-related interstitial lung diseases. J Rheumatol 2014;41(4):792–8 Epub 2014 Feb 1. doi:10.3899/jrheum.131251.

[24] Vincent K, Kennedy S, Stratton P. Pain scoring in endometriosis: entry criteria and outcome measures for clinical trials. Report from the Art and Science of Endometriosis meeting. Fertil Steril 2010;93(1):62–7 Epub 2008 Nov 5. doi:10.1016/j.fertnstert.2008.09.056.

[25] Lipton RB, Micieli G, Russell D, Solomon S, Tfelt-Hansen P, Waldenlind E. Guidelines for controlled trials of drugs in cluster headache. Cephalalgia 1995;15(6):452–62.

[26] Tfelt-Hansen P, Block G, Dahlöf C, Diener HC, Ferrari MD, Goadsby PJ, et al. Guidelines for controlled trials of drugs in migraine: second edition. Cephalalgia 2000;20(9):765–86. doi:10.1046/j.1468-2982.2000.00117.x.

[27] Penzien DB. Guidelines for trials of behavioral treatments for recurrent headache: purpose, process, and product. Headache 2005;45(Suppl 2):S87–9. doi:10.1111/j.1526-4610.2005.4502001.x.

[28] Bendtsen L, Bigal ME, Cerbo R, Diener HC, Holroyd K, Lampl C, et al. Guidelines for controlled trials of drugs in tension-type headache: second edition. Cephalalgia 2010;30(1):1–16. doi:10.1111/j.1468-2982.2009.01948.x.

[29] Miller RG, Munsat TL, Swash M, Brooks BR. Consensus guidelines for the design and implementation of clinical trials in ALS. World Federation of Neurology committee on Research. J Neurol Sci 1999;169(1-2):2–12. doi:10.1016/s0022-510x(99)00209-9.

[30] Leigh PN, Swash M, Iwasaki Y, Ludolph A, Meininger V, Miller RG, et al. Amyotrophic lateral sclerosis: a consensus viewpoint on designing and implementing a clinical trial. Amyotroph Lateral Scler Other Motor Neuron Disord 2004;5(2):84–98. doi:10.1080/14660820410020187.

[31] Higashida RT, Furlan AJ, Roberts H, Tomsick T, Connors B, Barr J, et al. Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke. Stroke 2003;34(8):e109–37 Epub 2003 Jul 17. doi:10.1161/01.STR.0000082721.62796.09.

[32] Vargus-Adams JN, Martin LK. Measuring what matters in cerebral palsy: a breadth of important domains and outcome measures. Arch Phys Med Rehabil 2009;90(12):2089–95. doi:10.1016/j.apmr.2009.06.018.

[33] Mindell JA, Emslie G, Blumer J, Genel M, Glaze D, Ivanenko A, et al. Pharmacologic management of insomnia in children and adolescents: consensus statement. Pediatrics 2006;117(6):e1223–32. doi:10.1542/peds.2005-1693.

[34] Merkies IS, Lauria G. 131st ENMC international workshop: selection of outcome measures for peripheral neuropathy clinical trials 10-12 December 2004, Naarden, The Netherlands. Neuromuscul Disord 2006;16(2):149–56 Epub 2006 Jan 23. doi:10.1016/j.nmd.2005.12.003.

[35] Reilly MM, de Jonghe P, Pareyson D. 136th ENMC International Workshop: Charcot-Marie-Tooth disease type 1A (CMT1A)8-10 April 2005, Naarden, The Netherlands. Neuromuscul Disord 2006;16(6):396–402 Epub 2006 May 8. doi:10.1016/j.nmd.2006.03.008.

[36] Wilde EA, Whiteneck GG, Bogner J, Bushnik T, Cifu DX, Dikmen S, et al. Recommendations for the use of common outcome measures in traumatic brain injury research. Arch Phys Med Rehabil 2010;91(11)):1650–1660.e17. doi:10.1016/j.apmr.2010.06.033.

[37] Paul L, Coote S, Crosbie J, Dixon D, Hale L, Holloway E, et al. Core outcome measures for exercise studies in people with multiple sclerosis: recommendations from a multidisciplinary consensus meeting. Mult Scler 2014;20(12):1641–50 Epub 2014 Mar 17. doi:10.1177/1352458514526944.

[38] Wallace SJ, Worrall L, Rose T, Le Dorze G, Breitenstein C, Hilari K, et al. A core outcome set for aphasia treatment research: the ROMA consensus statement. Int J Stroke 2019;14(2):180–5 Epub 2018 Oct 10. doi:10.1177/1747493018806200.

[39] Webster L, Groskreutz D, Grinbergs-Saull A, Howard R, O'Brien JT, Mountain G, et al. Core outcome measures for interventions to prevent or slow the progress of dementia for people living with mild to moderate dementia: Systematic review and consensus recommendations. PLoS One 2017;12(6):e0179521. doi:10.1371/journal.pone.0179521.

[40] Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, et al. Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. Int J Stroke 2017;12(5):451–61. doi:10.1177/1747493017711813.

[41] Kloppenburg M, Bøyesen P, Visser AW, Haugen IK, Boers M, Boonen A, et al. Report from the OMERACT Hand Osteoarthritis Working Group: Set of Core Domains and Preliminary Set of Instruments for Use in Clinical Trials and Observational Studies. J Rheumatol 2015;42(11):2190–7 NovEpub 2015 Jul 1. doi:10.3899/jrheum.141017.

[42] Distler O, Behrens F, Pittrow D, Huscher D, Denton CP, Foeldvari I, et al. Defining appropriate outcome measures in pulmonary arterial hypertension related to systemic sclerosis: a Delphi consensus study with cluster analysis. Arthritis Rheum 2008;59(6):867–75. doi:10.1002/art.23718.

[43] Khanna D, Mittoo S, Aggarwal R, Proudman SM, Dalbeth N, Matteson EL, et al. Connective tissue disease-associated interstitial lung diseases (CTD-ILD) - report from OMERACT CTD-ILD working group. J Rheumatol 2015;42(11):2168–71 Epub 2015 Mar 1. doi:10.3899/jrheum.141182.

[44] Barlow GD, Lamping DL, Davey PG, Nathwani D. Evaluation of outcomes in community-acquired pneumonia: a guide for patients, physicians, and policy-makers. Lancet Infect Dis 2003;3(8):476–88. doi:10.1016/s1473-3099(03)00721-7.

[45] Dorman S, Jolley C, Abernethy A, Currow D, Johnson M, Farquhar M, et al. Researching breathlessness in palliative care: consensus statement of the National Cancer Research Institute Palliative Care Breathlessness Subgroup. Palliat Med 2009;23(3):213–27 Epub 2009 Feb 27. doi:10.1177/0269216309102520.

[46] Fitzpatrick R, Chambers J, Burns T, Doll H, Fazel S, Jenkinson C, et al. A systematic review of outcome measures used in forensic mental health research with consensus panel opinion. Health Technol Assess 2010;14(18):1–94. doi:10.3310/hta14180.

[47] Merkel PA, Herlyn K, Mahr AD, Neogi T, Seo P, Walsh M, et al. Progress towards a core set of outcome measures in small-vessel vasculitis. Report from OMERACT 9. J Rheumatol. 2009;36(10):2362–8. doi:10.3899/jrheum.090373.

[48] McGrath PJ, Walco GA, Turk DC, Dworkin RH, Brown MT, Davidson K, et al. Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: PedIMMPACT recommendations. J Pain 2008;9(9):771–83 Epub 2008 Jun 17. doi:10.1016/j.jpain.2008.04.007.

[49] Miller FW, Rider LG, Chung YL, Cooper R, Danko K, Farewell V, et al. Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies. Rheumatology (Oxford) 2001;40(11):1262–73. doi:10.1093/rheumatology/40.11.1262.

[50] Goldhahn J, Beaton D, Ladd A, Macdermid J, Hoang-Kim ADistal Radius Working Group of the International Society for Fracture Repair (ISFR); International Osteoporosis Foundation (IOF). Recommendation for measuring clinical outcome in distal radius fractures: a core set of domains for standardized reporting in clinical practice and research. Arch Orthop Trauma Surg. 2014;134(2):197–205 Epub 2013 Jun 1. doi:10.1007/s00402-013-1767-9.

[51] Becker LB, Aufderheide TP, Geocadin RG, Callaway CW, Lazar RM, Donnino MW, et al. Primary outcomes for resuscitation science studies: a consensus statement from the American Heart Association. Circulation. 2011;124(19):2158–77 Epub 2011 Oct 3. doi:10.1161/CIR.0b013e3182340239.

[52] Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. Clin Exp Rheumatol 2005;23(5 Suppl 39):S53–7.

[53] Grieve S, Perez RSGM, Birklein F, Brunner F, Bruehl S, Harden RN, et al. Recommendations for a first core outcome measurement set for complex regional pain syndrome clinical studies (COMPACT). Pain 2017;158(6):1083–90. doi:10.1097/j.pain.0000000000000866.

[54] Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LOP, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. Pain 2018;159(3):481–95. doi:10.1097/j.pain.0000000000001117.

[55] Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. PLoS Med 2017;14(11):e1002447. doi:10.1371/journal.pmed.1002447.

[56] Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. Eur J Health Econ 2018;19(4):557–70 Epub 2017 May 30. doi:10.1007/s10198-017-0902-x.

[57] Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. Trials 2016;17(1):449. doi:10.1186/s13063-016-1555-2.

[58] Meregaglia M, Ciani O, Banks H, Salcher-Konrad M, Carney C, Jayawardana S, et al. A scoping review of core outcome sets and their 'mapping' onto real-world data using prostate cancer as a case study. BMC Med Res Methodol 2020;20(1):41. doi:10.1186/s12874-020-00928-w.