Journal of Clinical Epidemiology

A cross-sectional analysis of Patient-Reported Outcome Measures in Core Outcome Sets reveals huge overlap within and across disease areas but lack of harmonization in recommended instruments

--Manuscript Draft--

Manuscript Number:	
Article Type:	Original article
Keywords:	core outcome set; COS; patient-reported outcome measure; PRO; PROM; outcomes research
Corresponding Author:	Oriana Ciani SDA Bocconi School of Management Milano, County (optional) ITALY
First Author:	Oriana Ciani
Order of Authors:	Oriana Ciani
	Maximilian Salcher-Konrad
	Michela Meregaglia
	Kathrine Smith
	Sarah L Gorst
	Susanna Dodd
	Paula R Williamson
	Giovanni Fattore
Manuscript Region of Origin:	Europe
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 analyze the health domains targeted by PROs, the specific patient-reported outcomes measures (PROMs) recommended, 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.
Suggested Reviewers:	CAROLINE TERWEE cb.terwee@vumc.nl Author in 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 Sep;125:64-75. doi: 10.1016/j.jclinepi.2020.05.021. Epub 2020 May 26. PMID: 32470621.
	Natasha Roberts

	N6.roberts@hdr.qut.edu.au Author of Roberts NA, Alexander K, Wyld D, Janda M. Statistical process control assessed implementation fidelity of patient-reported outcome measures (PROMs) in routine care. J Clin Epidemiol. 2020 Nov;127:76-86. doi: 10.1016/j.jclinepi.2020.06.022. Epub 2020 Jun 20. PMID: 32569618.
	Lidwine B Mokkink w.mokkink@vumc.nl Author of Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018 May;27(5):1171-1179. doi: 10.1007/s11136-017-1765-4. Epub 2017 Dec 19. PMID: 29260445; PMCID: PMC5891552.
Opposed Reviewers:	

Milan, 30th November 2020

Dear Editor,

I hope this letter finds you well despite the unprecedented circumstances we are going through.

I am delighted to submit our manuscript "A cross-sectional analysis of Patient-Reported Outcome Measures in Core Outcome Sets reveals huge overlap within and across disease areas but lack of harmonization in recommended instruments" for consideration in Journal of Clinical Epidemiology.

We mapped 94 Core Outcome Sets (COS) for research and practice across 26 disease areas and identified a total of 323 unique Patient-Reported Outcome Measures (PROMs) were recommended for use. Of these, 87% were recommended in only one COS, with each COS including a median of 4.5 instruments. We investigated what domains these PROMs targeted based on a previously published taxonomy (J Clin Epidemiol 2018 Apr;96:84-92). Both disease-specific and general health domains were covered, however global quality of life (25%) and physical functioning (22%) were the most frequently targeted. Overall, a fragmented landscape of recommended PROMs in COS emerges, calling for better harmonization of PRO selection and measurement.

We believe our study will be an important contribution to expand knowledge on the effective and efficient use of PROMs in clinical research and practice.

We look forward to hearing back from you,

kind regards

Oriana Ciani, PhD

On behalf of all coauthors

A cross-sectional analysis of Patient-Reported Outcome Measures in Core Outcome Sets reveals huge overlap within and across disease areas but lack of harmonization in recommended instruments

Oriana Ciani[1,2], Maximilian Salcher-Konrad[3,4], Michela Meregaglia[1], Kathrine Smith[5], Sarah L Gorst[6], Susanna Dodd[6], Paula R Williamson[6], Giovanni Fattore[1,7]

[1] Centre for Research on Health and Social Care Management, SDA Bocconi, via Sarfatti 10, 20136, Milan, Italy

[2] Evidence Synthesis and Modeling for Health Improvement, College of Medicine and Health, University of Exeter, EX1 2LU, Exeter, UK

[3] LSE Health, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

[4] Care Policy and Evaluation Centre, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

[5] Imperial College, London, UK

[6] MRC/NIHR Trials Methodology Research Partnership, Department of Health Data Science, University of Liverpool, Liverpool L69 3BX, UK

[7] Department of Social and Political Science, Bocconi University, via Sarfatti 36, 20136, Milan, Italy

Corresponding author
Oriana Ciani, PhD
Centre for Research on Health and Social Care Management, SDA Bocconi
via Sarfatti 10
20136, Milan, Italy
Ph: +39 02 5836 5248
oriana.ciani@unibocconi.it

Email addresses:
Maximilian Salcher-Konrad: m.salcher@lse.ac.uk
Michela Meregaglia: michela.meregaglia@unibocconi.it
Kathrine Smith: smith.katherine.a@gmail.com
Sarah Gorst: sgorst@liverpool.ac.uk
Susanna Dodd: s.r.dodd@liverpool.ac.uk
Paula Williamson: prw@liv.ac.uk
Giovanni Fattore: giovanni.fattore@unibocconi.it

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 analyze the health domains targeted by PROs, the specific patient-reported outcomes measures (PROMs) recommended, 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

Highlights

- A total of 323 unique PROMs were recommended for use across 94 COS and 26 disease areas.
- 87% of instruments were recommended in only one COS, and each COS included a median of 4.5 instruments.
- Targeted outcomes ranged from disease-specific to general health domains, with variability across disease areas.
- Overall, global quality of life (25%) and physical functioning (22%) were the most frequently targeted health outcome domains.
- A fragmented landscape of recommended PROMs in COS calls for better harmonization of PRO selection and measurement.

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. Combining clinical, genomic, and proteomic with PRO data provides the most complete picture of a patient's health status and may fuel, in the context of the clinical encounter and beyond, shared decision making and individualized care (2). 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 (3, 4). 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 (5). Hence, 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 (6).

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 and practice (7).

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) (8). This problem hinders comparison of PROMs and data synthesis across 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 studies or sites (9). 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 (10).

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 (8). Whilst there is some evidence about methods used to select outcome measurement instruments generally (11), there is no comprehensive assessment of which PROs are proposed and how they are recommended to be measured in COS.

The aims of this study are: (1) to review COS that include PROs, and identify their target health domains, (2) to identify which specific PROMs are recommended to measure PROs, and describe their main characteristics, (3) to assess the amount of overlap of target health domains across individual PROMs, and (4) the amount of overlap of recommended PROMs across different disease areas.

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). The database is regularly updated (12) (9), in order to ensure the currency of its content.

Inclusion and exclusion criteria

Out of the list of COS development studies identified from the original COMET systematic review and annual updates (12), 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 (3, 4), hence we excluded caregiver-, surrogate- or proxy-reported outcomes.

Study Selection

All identified articles were then screened independently by two researchers (MSK, KS) 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.

Data extraction

Data extraction was initially piloted on five randomly selected COS studies (13-17) 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 instruments 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 instruments 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 (18). Moreover, we classified the instruments as preference-based when yielding preference weights for quality-adjusted life years calculations (19, 20). 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 (10).

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, could not be obtained free of charge, or were not available in one of the languages spoken by the four data extractors (English, German, Italian). Additional detailed information were collected based on full-texts of recommended PROMs (structure of the instrument; number of items; verbatim items) to conduct content analysis. Each individual item included in the instruments was then matched to the same 38-domain outcome taxonomy used to classify target domains. Data extraction was double-checked by at least one reviewer.

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 instruments recommended. We reported results for characteristics of all recommended PROMs (full sample of unique PROMs). We used the same methods to conduct additional analyses on the contents of individual items of recommended PROMs in selected disease areas where full texts were available ('content analysis' sample).

Results

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 (Figure **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: there were 16 COS in Neurology, 11 in Rheumatology, 10 in Orthopaedics & Trauma, 9 in Lung&Airways, 8 in Heart & Circulation, and 7 in Cancer.

Characteristics of PROMs recommended

The total number of unique instruments recommended across these 94 COS development studies was 323. These include variations of existing PROMs, such as shorter versions (e.g. 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 instrument, the total number of unique PROMs recommended was 243.

The vast majority of the 323 instruments was recommended in only one COS study (280, 87%). 23 instruments (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 (**Error! Reference source not found.**). Lastly, there were few instruments (2%) that were recommended in more than 4 COS studies.

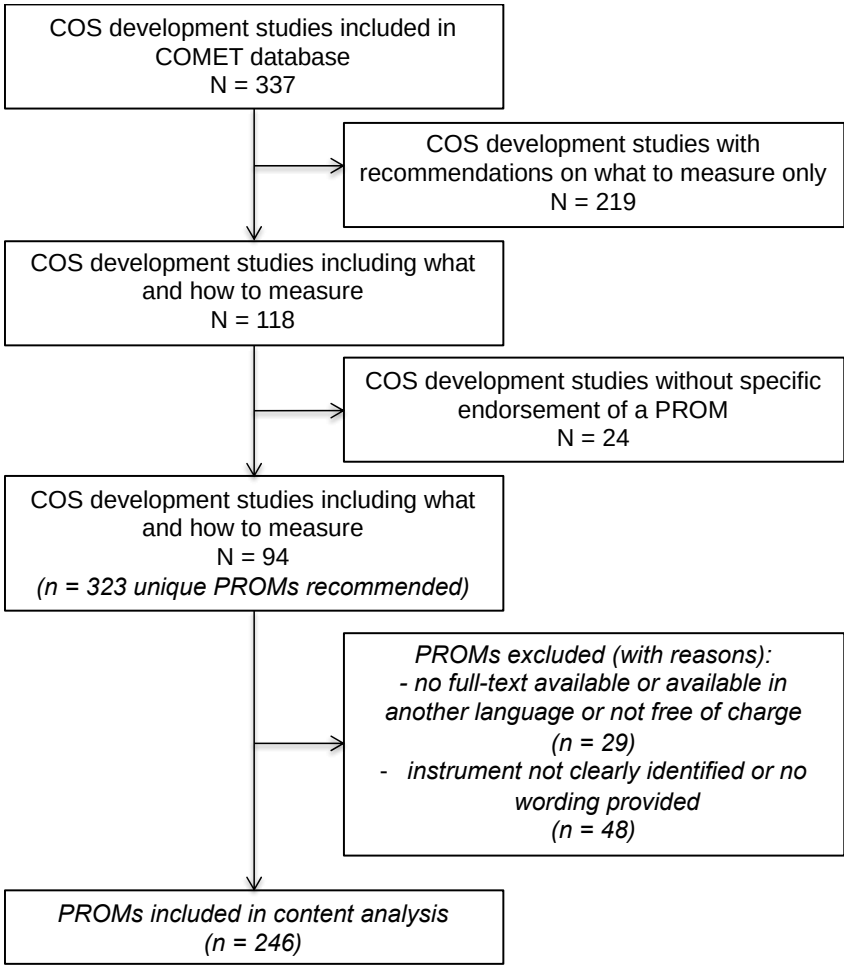


Figure 1 Flow chart of COS studies and recommended instruments selection process

The average number of unique instruments recommended per COS was 4.5 (median 3, minimum 1, maximum 17). Most frequently, only one instrument 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 (Figure **2**).

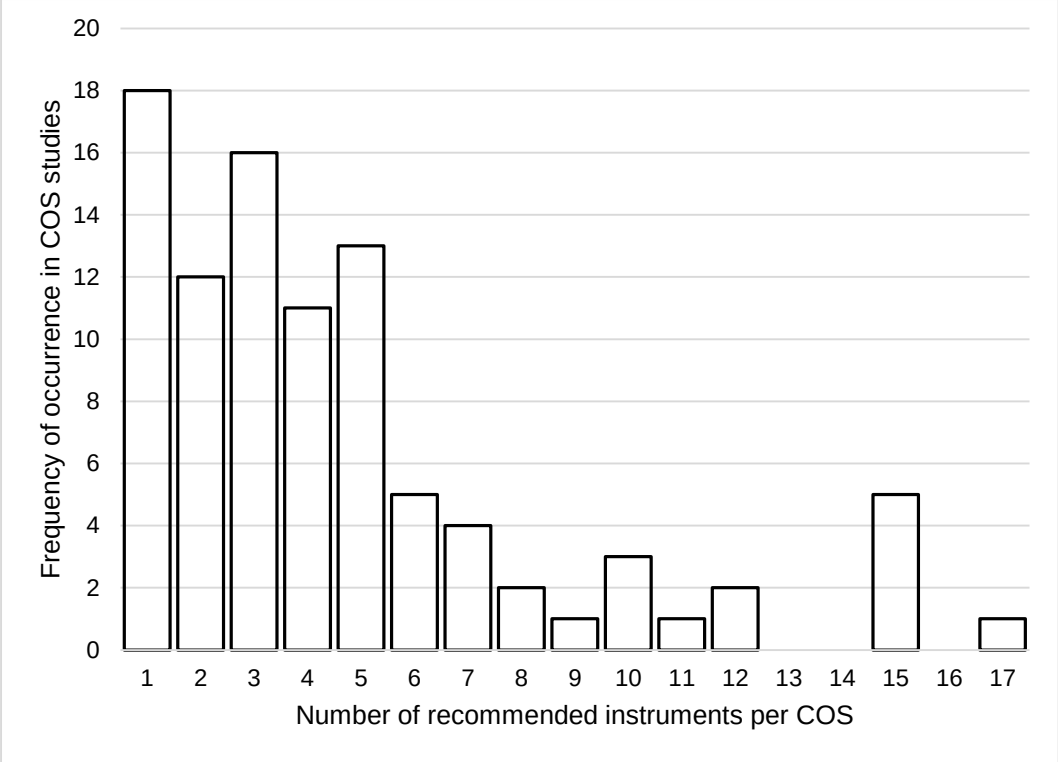


Figure 2 Distribution of number of instruments recommended in a COS study

Overall, the majority of the instruments recommended were disease-specific (248, 77%), that is, they were designed to be relevant to a specific disease, population, function or condition (20). However, half of the 20 instruments recommended in 3 or more COS were generic (**Error! Reference source not found.**). Overall, five preference-based instruments (1.5%) were identified.

The majority of recommended instruments were full questionnaires (61%). A significant minority (18%) were questions or subscales of existing instruments (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 (13), 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 (14) or a statement that single items were believed to be more informative, and even more

statistically sensitive, than the overall questionnaires (21). Lastly, one study (22) provided a list of referenced studies that developed and tested individual components of a full instrument (i.e. the PROMIS-29 questionnaire).

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

Table 1 Characteristics of recommended PROMs

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

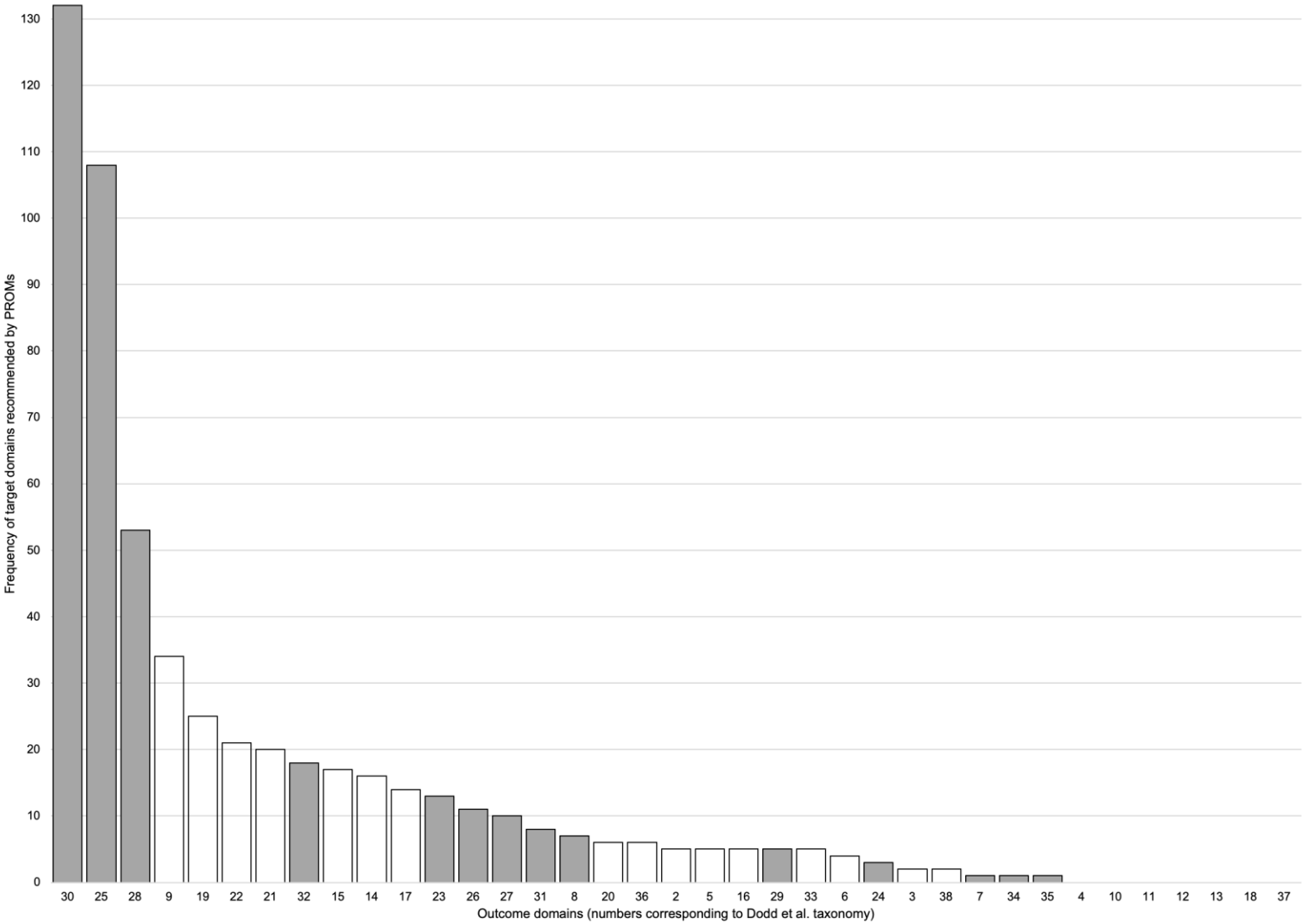
More than one in five recommended instruments (21%) were single questions, i.e. either individual questions taken out of existing questionnaires, or stand-alone questions that COS developers deemed relevant. These were often numerical rating

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 instruments 'extremely important' and therefore recommended them (14). 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 (23), patient global score (24) or use of rescue analgesics (25)). Content analysis was not possible for these questions.

Outcome domains targeted by COS developers for recommended PROMs

'Global quality of life' (domain 30) was found to be the most frequently targeted domain across disease areas, followed by physical functioning (domain 25) and emotional functioning and wellbeing (domain 28) (Figure **3**). Recommended PROMs were mostly used to capture general health related outcome domains, rather than organ-, system- or disease- specific outcome domains (10), with some variation across disease areas. Figure 4 shows how often each domain was targeted by recommended instruments in each of the 26 disease areas examined in the COS studies.

Figure 3 Outcome domains targeted by recommended PROMs across disease areas



White bars relate to organ-, system- or disease- specific outcome domains. Shaded bars relate to generic health-related outcome domains.

Figure 4 Outcome domains targeted in recommend PROMs by disease area

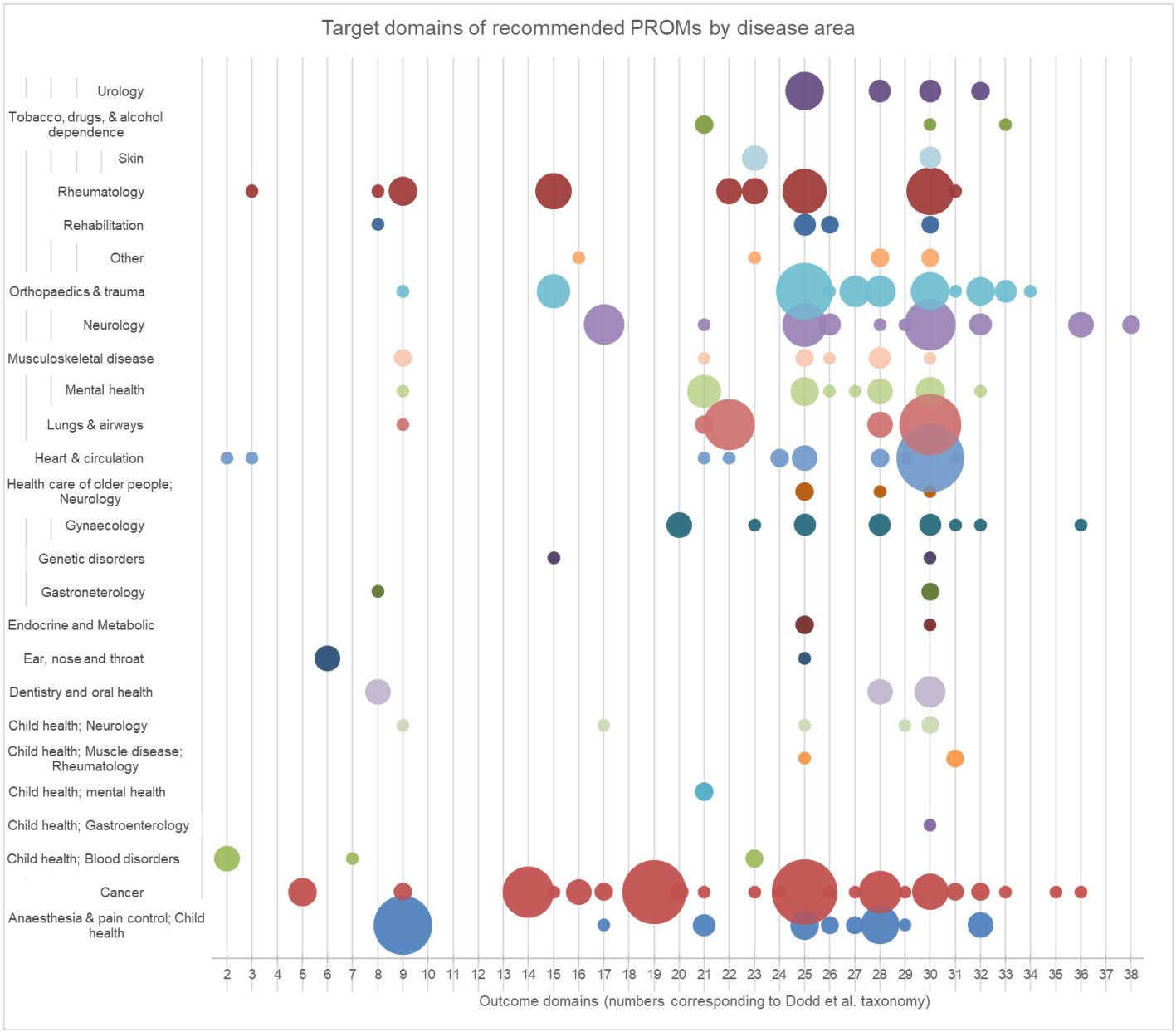


Figure shows the 38 outcome domains (horizontal axis) targeted by 323 unique recommended instruments by disease area (vertical axis). The list of outcome domains is provided below (reproduced under CC BY license from Dodd et al., 2018). Bubble sizes indicate frequency of an individual domain being the target domain of a recommended PROM in the disease area. For example, in COS for urology, domain 25 (physical

functioning) was the target domain of 9 recommended PROMs, domain 28 (emotional functioning/wellbeing) was the target domain of 3 recommended PROMs, domain 30 (global quality life) was the target domain of 3 recommended PROMs, and domain 32 (delivery of care) was the target domain of 2 recommended PROMs. Details provided in Supplementary Material.

1: Mortality/survival	9: General outcomes	17: Nervous system outcomes	25: Physical functioning	33: Personal circumstances
2: Blood and lymphatic system outcomes	10: Hepatobiliary outcomes	18: Pregnancy, puerperium and perinatal outcomes	26: Social functioning	34: Economic
3: Cardiac outcomes	11: Immune system outcomes	19: Renal and urinary outcomes	27: Role functioning	35: Hospital
4: Congenital, familial and genetic outcomes	12: Infection and infestation outcomes	20: Reproductive system and breast outcomes	28: Emotional functioning/wellbeing	36: Need for intervention
5: Endocrine outcomes	13: Injury and poisoning outcomes	21: Psychiatric outcomes	29: Cognitive functioning	37: Societal/carer burden
6: Ear and labyrinth outcomes	14: Metabolism and nutrition outcomes	22: Respiratory, thoracic and mediastinal outcomes	30: Global quality of life	38: Adverse events/effects
7: Eye outcomes	15: Musculoskeletal and connective tissue outcomes	23: Skin and subcutaneous tissue outcomes	31: Perceived health status	
8: Gastrointestinal outcomes	16: Outcomes relating to neoplasms: benign, malignant and unspecified (including cysts and polyps)	24: Vascular outcomes	32: Delivery of care	

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 (26), 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 (14, 27, 28); in another study (29), 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 instrument 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' (30-32). 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 methods of administration (e.g., self-report, interviewer); and availability and equivalence of versions for different cultures and languages (33-37).

Content analysis of recommended PROMs by disease area

For the purpose of content analysis, the number of included unique instruments decreased from 323 to 246 (Figure 1, list available in the Supplementary Material).

This is because we were not able to locate the full text, could not obtain it free of charge, or could not find it in one of covered languages in the research team (n=29), but most frequently because COS developers did not provide sufficient details in terms of proposed language for the item, or response options (n=48).

There were six disease areas with five or more COS studies recommending PROMs (i.e. neurology, rheumatology, orthopaedics and trauma, lungs and airways, cancer, and heart and circulation). For these disease areas, we conducted content analysis to identify outcome domains targeted by individual items or subscales within each recommended instrument.

Among included COS studies, 16 referred to a neurological condition, including four COS for headaches and migraine (38-41), two for amyotrophic lateral sclerosis (42, 43), and one each for ischaemic stroke (44), cerebral palsy (28), insomnia (45), peripheral neuropathy (46), Charcot-Marie-Tooth disease (47), traumatic brain injury (48), multiple sclerosis (49), post-stroke aphasia (50), mild or moderate dementia (51), and sensorimotor recovery after stroke (52). In total, 47 unique instruments were recommended in neurology, but only two instruments were recommended more than once: SF-36 (in 4 COS) and EQ-5D (in 3 COS).

Among the 24 instruments for which full texts or clear wording were available, the median number of items per recommended PROM was 17.5, ranging from 5 to 54 items. Figure **5**5 provides an overview of the domains targeted within each recommended PROM across the neurology COS studies, with an indication of the frequency with which each outcome domain is targeted.

Figure 5 Outcome domains targeted by recommended PROMs across COS in Neurology

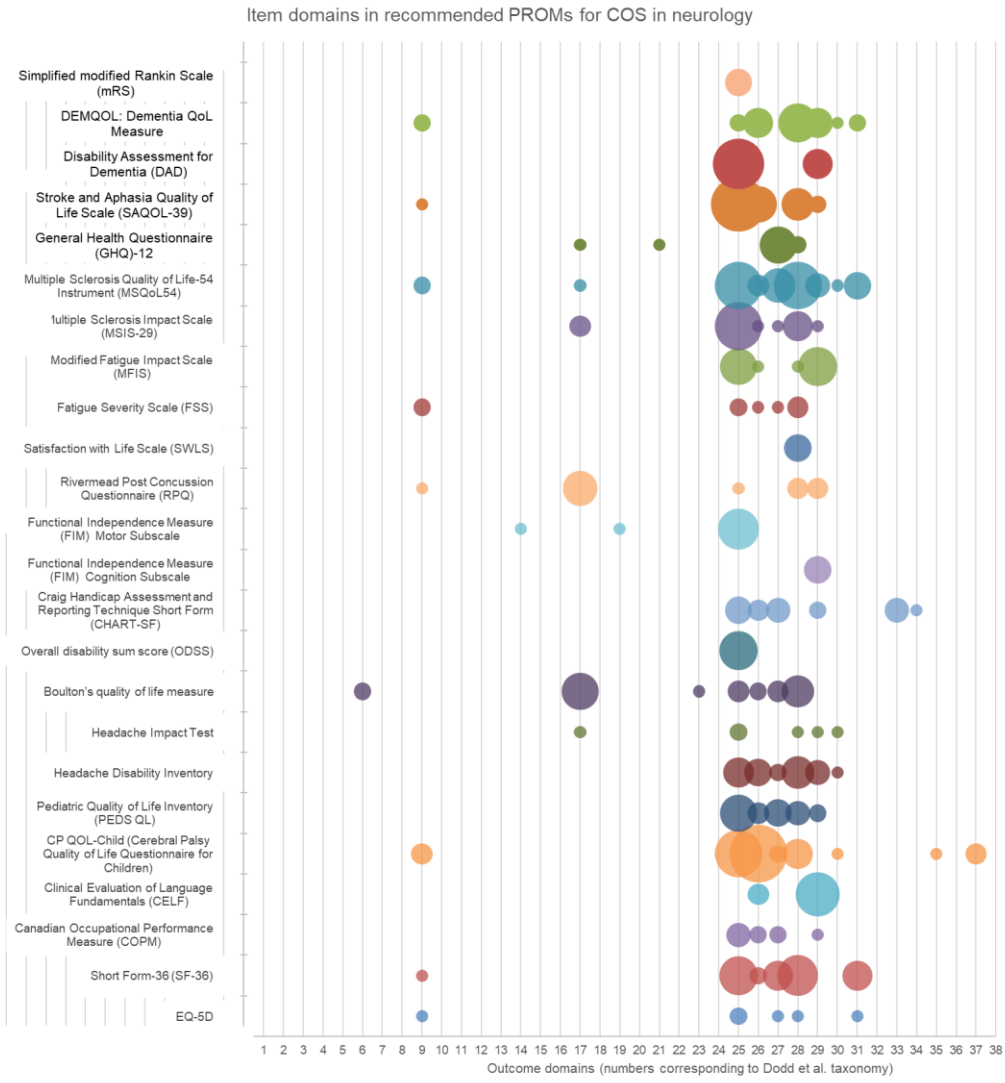


Figure shows the 38 outcome domains (horizontal axis) targeted by 24 recommended instruments (vertical axis) in neurology COS. The list of outcome domains is provided Figure 4 (reproduced under CC BY license from Dodd et al., 2018). Bubble sizes indicate frequency of individual items in recommended PROMs being categorized in the domains listed below. For example, in the SF-36, 10 items were categorized in domain 25 (physical functioning), 2 items in domain 26 (social functioning), 6 items in domain 27 (role functioning) etc.

Further content analysis of the individual items showed thematic overlap by outcome domains categorized, with an illustration related to the physical functioning domain in Table **2**.

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)

Table 2 PROMs items mapped to Physical Functioning domain in Neurology

Similar patterns are seen for the content analyses performed across PROMs recommended in other disease areas (Supplementary Material).

Discussion

In this study we mapped recommendation of PROMs as part of standardized outcome collection in clinical research and practice 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.

Despite the intended aim of striving for harmonization and alignment of outcomes, we found a fragmented landscape of recommended PROMs in COS. A total of 323 unique PROMs were recommended for use. The vast majority (87%) of instruments were recommended in only one COS, and individual COS recommended a median of more than 4 instruments (range, 1-17) each. We found that, by recommending PROMs, COS developers targeted a variety of outcome domains from disease activity or severity of disease to personal circumstances or need for intervention. 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. However, most of the instruments (25%) covered 'global quality of life' (domain 30), physical functioning (22%) (domain 25), followed by emotional functioning and wellbeing (7%) (domain 28). These are broad health-related constructs that appear consistently targeted across disease areas, yet a broad range of different instruments was recommended to measure them. This was true even within disease areas, where our in-depth content analysis of individual PROMs revealed significant overlap among specific items of recommended instruments. Although some of our findings may be explained by the heterogeneity of disease areas, which encourages development of disease-specific instruments that are preferred to generic ones, there is still scope for increased harmonization of outcome measurement even within disease areas.

Almost one in four recommended instruments 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%) instruments were preference-based, and therefore useful for

cost-utility analyses; four of these were generic measures, with potential use in cost-utility analyses across disease areas.

Finally, while not the focus of our study, 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 instruments 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 (53), 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 for validation of recommended instruments was not commonly reported.

More than 60% of recommended instruments 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, whilst 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 with which COS were developed and whether COS developers followed the COMET and COSMIN guidance (53, 54) to first identify the domains to be measured, and in the next step identify and evaluate potential instruments for these.

Striving for efficient PRO data collection is an imperative in the context of rising use of real-world data 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 (55). However, the current technological landscape would allow for a wide-scale, standardized, continuous collection of PROMs that is integrated in clinical practice and everyday care (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 instruments 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.

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).

References

1. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). (1178-6329 (Print)).
2. Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. (1756-1833 (Electronic)).
3. U.S. Food And Drug Administration. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims Accessed 9 September 2019.
4. 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. https://www.ema.europa.eu/en/appendix-2-guideline-evaluation-anticancer-medicinal-products-man-use-patient-reported-outcome-pro Accessed 9 September 2019.
5. Lavallee DC, Chenok KE, Love RM, et al. Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. (1544-5208 (Electronic)).
6. Basch E, Torda P Fau - Adams K, Adams K. Standards for patient-reported outcome-based performance measures. (1538-3598 (Electronic)).
7. Ciani O, Federici CB. Value Lies in the Eye of the Patients: The Why, What, and How of Patient-Reported Outcomes Measures. LID - S0149-2918(19)30581-8 [pii] LID - 10.1016/j.clinthera.2019.11.016 [doi]. (1879-114X (Electronic)).
8. Macefield RC, Jacobs M Fau - Korfage IJ, Korfage Ij Fau - Nicklin J, et al. Developing core outcomes sets: methods for identifying and including patient-reported outcomes (PROs). (1745-6215 (Electronic)).
9. Gargon E, Gurung B, Medley N, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. (1932-6203 (Electronic)).
10. 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. (1878-5921 (Electronic)).
11. 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. (1878-5921 (Electronic)).
12. Gargon EA-O, 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. (1932-6203 (Electronic)).
13. Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. (0362-2436 (Print)).
14. Khanna D, Lovell Dj Fau - Giannini E, Giannini E Fau - Clements PJ, et al. Development of a provisional core set of response measures for clinical trials of systemic sclerosis. (1468-2060 (Electronic)).
15. Clements PJ, Allanore Y Fau - Khanna D, Khanna D Fau - Singh M, Singh M Fau - Furst DE, Furst DE. Arthritis in systemic sclerosis: systematic review of the literature and suggestions for the performance of future clinical trials in systemic sclerosis arthritis. (1532-866X (Electronic)).

16. Merkel PA, Aydin Sz Fau - Boers M, Boers M Fau - Direskeneli H, et al. The OMERACT core set of outcome measures for use in clinical trials of ANCA-associated vasculitis. (0315-162X (Print)).
17. Hellmich B, Flossmann O Fau - Gross WL, Gross Wl Fau - Bacon P, et al. EULAR recommendations for conducting clinical studies and/or clinical trials in systemic vasculitis: focus on anti-neutrophil cytoplasm antibody-associated vasculitis. (0003-4967 (Print)).
18. Coons SJ, Rao S Fau - Keininger DL, Keininger Dl Fau - Hays RD, Hays RD. A comparative review of generic quality-of-life instruments. (1170-7690 (Print)).
19. Brazier J, Ara R, Rowen D, Chevrou-Severac H. A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. (1179-2027 (Electronic)).
20. Rowen D, Brazier J, Ara R, Azzabi Zouraq I. The Role of Condition-Specific Preference-Based Measures in Health Technology Assessment. (1179-2027 (Electronic)).
21. Vocci, de Wit H. Consensus statement on evaluation of outcome of pharmacotherapy for substance abuse/ dependence: report from a NIDA/CPDD meeting. Betheseda, MD: National Institute on Drug Abuse Medications Development Division1999.
22. Deyo RA, Dworkin SF, Amtmann D, et al. Report of the NIH Task Force on research standards for chronic low back pain. (1528-8447 (Electronic)).
23. Heiligenhaus A, Foeldvari I Fau - Edelsten C, Edelsten C Fau - Smith JR, et al. 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. (2151-4658 (Electronic)).
24. Saketkoo LA, Mittoo S Fau - Frankel S, Frankel S Fau - LeSage D, 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. (0315-162X (Print)).
25. Vincent K, Kennedy S Fau - Stratton P, Stratton P. Pain scoring in endometriosis: entry criteria and outcome measures for clinical trials. Report from the Art and Science of Endometriosis meeting. (1556-5653 (Electronic)).
26. Kloppenburg M, Boyesen P, Visser AW, 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. (0315-162X (Print)).
27. Distler O, Behrens F Fau - Pittrow D, Pittrow D Fau - Huscher D, et al. Defining appropriate outcome measures in pulmonary arterial hypertension related to systemic sclerosis: a Delphi consensus study with cluster analysis. (0004-3591 (Print)).
28. Vargus-Adams JN, Martin LK. Measuring what matters in cerebral palsy: a breadth of important domains and outcome measures. (1532-821X (Electronic)).
29. Khanna D, Mittoo S, Aggarwal R, et al. Connective Tissue Disease-associated Interstitial Lung Diseases (CTD-ILD) - Report from OMERACT CTD-ILD Working Group. (0315-162X (Print)).
30. Barlow GD, Lamping Dl Fau - Davey PG, Davey Pg Fau - Nathwani D, Nathwani D. Evaluation of outcomes in community-acquired pneumonia: a guide for patients, physicians, and policy-makers. (1473-3099 (Print)).
31. Dorman S, Jolley C Fau - Abernethy A, Abernethy A Fau - Currow D, et al. Researching breathlessness in palliative care: consensus statement of the National Cancer Research Institute Palliative Care Breathlessness Subgroup. (1477-030X (Electronic)).
32. Fitzpatrick R, Chambers J Fau - Burns T, Burns T Fau - Doll H, et al. A systematic review of outcome measures used in forensic mental health research with consensus panel opinion. (2046-4924 (Electronic)).

33. Merkel PA, Herlyn K Fau - Mahr AD, Mahr Ad Fau - Neogi T, et al. Progress towards a core set of outcome measures in small-vessel vasculitis. Report from OMERACT 9. (0315-162X (Print)).
34. McGrath PJ, Walco Ga Fau - Turk DC, Turk Dc Fau - Dworkin RH, et al. Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: PedIMMPACT recommendations. (1528-8447 (Electronic)).
35. Miller FW, Rider Lg Fau - Chung YL, Chung Yl Fau - Cooper R, et al. Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies. (1462-0324 (Print)).
36. Goldhahn J, Beaton D Fau - Ladd A, Ladd A Fau - Macdermid J, Macdermid J Fau - Hoang-Kim A, Hoang-Kim A. Recommendation for measuring clinical outcome in distal radius fractures: a core set of domains for standardized reporting in clinical practice and research. (1434-3916 (Electronic)).
37. Becker Lb Fau - Aufderheide TP, Aufderheide Tp Fau - Geocadin RG, Geocadin Rg Fau - Callaway CW, et al. Primary outcomes for resuscitation science studies: a consensus statement from the American Heart Association. (1524-4539 (Electronic)).
38. Lipton RB, Micieli G Fau - Russell D, Russell D Fau - Solomon S, Solomon S Fau - Tfelt-Hansen P, Tfelt-Hansen P Fau - Waldenlind E, Waldenlind E. Guidelines for controlled trials of drugs in cluster headache. (0333-1024 (Print)).
39. Tfelt-Hansen P, Block G Fau - Dahlof C, Dahlof C Fau - Diener HC, et al. Guidelines for controlled trials of drugs in migraine: second edition. (0333-1024 (Print)).
40. Penzien DB. Guidelines for trials of behavioral treatments for recurrent headache: purpose, process, and product. (0017-8748 (Print)).
41. Bendtsen L, Bigal Me Fau - Cerbo R, Cerbo R Fau - Diener HC, et al. Guidelines for controlled trials of drugs in tension-type headache: second edition. (1468-2982 (Electronic)).
42. Miller RG, Munsat Tl Fau - Swash M, Swash M Fau - Brooks BR, Brooks BR. Consensus guidelines for the design and implementation of clinical trials in ALS. World Federation of Neurology committee on Research. (0022-510X (Print)). doi:D - KIE: 126469
D - NRCBL: VF 18.2 OTO - KIE
43. Leigh PN, Swash M Fau - Iwasaki Y, Iwasaki Y Fau - Ludolph A, et al. Amyotrophic lateral sclerosis: a consensus viewpoint on designing and implementing a clinical trial. (1466-0822 (Print)).
44. Higashida RT, Furlan Aj Fau - Roberts H, Roberts H Fau - Tomsick T, et al. Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke. (1524-4628 (Electronic)).
45. Mindell JA, Emslie G Fau - Blumer J, Blumer J Fau - Genel M, et al. Pharmacologic management of insomnia in children and adolescents: consensus statement. (1098-4275 (Electronic)).
46. Merkies IS, Lauria G. 131st ENMC international workshop: selection of outcome measures for peripheral neuropathy clinical trials 10-12 December 2004, Naarden, The Netherlands. (0960-8966 (Print)).
47. Reilly MM, de Jonghe P Fau - Pareyson D, Pareyson D. 136th ENMC International Workshop: Charcot-Marie-Tooth disease type 1A (CMT1A)8-10 April 2005, Naarden, The Netherlands. (0960-8966 (Print)).
48. Wilde EA, Whiteneck Gg Fau - Bogner J, Bogner J Fau - Bushnik T, et al. Recommendations for the use of common outcome measures in traumatic brain injury research. (1532-821X (Electronic)).
49. Paul L, Coote S, Crosbie J, et al. Core outcome measures for exercise studies in people with multiple sclerosis: recommendations from a multidisciplinary consensus meeting. (1477-0970 (Electronic)).

50. Wallace SA-O, Worrall L, Rose T, et al. A core outcome set for aphasia treatment research: The ROMA consensus statement. (1747-4949 (Electronic)).
51. Webster LA-O, Groskreutz D, Grinbergs-Saull A, 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. (1932-6203 (Electronic)).
52. Kwakkel G, Lannin NA, Borschmann K, et al. Standardized measurement of sensorimotor recovery in stroke trials: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. (1747-4949 (Electronic)).
53. Kirkham JJ, Davis KA-Ohoo, Altman DG, et al. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. (1549-1676 (Electronic)).
54. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. (1745-6215 (Electronic)).
55. Meregaglia MA-O, Ciani O, Banks H, et al. A scoping review of core outcome sets and their 'mapping' onto real-world data using prostate cancer as a case study. (1471-2288 (Electronic)).

Conflict of interest

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

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) [OC, MS, MM, KS, GF]. This work was also supported by a National Institute for Health Research (NIHR) Senior Investigator Award (NF-SI_0513-10025) [PW] and the MRC Trials Methodology Research Partnership (grant reference MR/S014357/1) [SD, SG, PW].