**Improvement needed in the standards of development for cancer core outcome sets**

Elizabeth Gargon1\*, Paula R. Williamson1, Jane M. Blazeby2, Jamie J. Kirkham1

1 MRC North West Hub for Trials Methodology Research, Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom

2 MRC ConDuCT II Hub for Trials Methodology Research and National Institute for Health Research Bristol Biomedical Research Centre, Population Health Sciences, Bristol Medical School, University of Bristol

\*Corresponding author

e.gargon@liverpool.ac.uk

+44(0) 151 794 9727

**Abstract**

**Objective**

The Core Outcome Set-STAndards for Development (COS-STAD) contains 11 standards (12 criteria) that are deemed to be the minimum design recommendations for all core outcome set (COS) development projects. Cancer is currently the disease area with the highest number of published COS and is a major cause of worldwide morbidity and mortality. The aim of this study was to provide a baseline of cancer COS standards.

**Study design**

Systematic reviews of COS have identified 307 published COS studies. Cancer COS were eligible for inclusion. Two reviewers independently assessed each of the COS against the 12 criteria.

**Results**

Forty-nine cancer COS were included; none met all 12 criteria representing the 11 minimum standards assessed in this study (range 4-11 criteria, median 6 criteria). All studies met the four scope standards, eight (16%) met all three standards for stakeholders involved, and two (4%) met all four standards for consensus process standards.

**Conclusion**

With the exception of Scope specification, there is much need for improvement. Poor reporting often made it challenging to assess whether minimum standards were met. The consensus process criteria were most difficult to assess, particularly those that required an assessment of being a priori. This is the first application of COS-STAD criteria to studies that have developed COS, and provides a baseline of cancer COS standards of development.

**Keywords** Core outcome set; Minimum standards; Cancer; Research methodology

**Word count: 3440**

**What is new?**

* This is the first application of COS-STAD criteria to studies that have developed COS, and provides a baseline of cancer COS standards of development. No COS met all of the 12 criteria representing the 11 minimum standards assessed in this study (range 4-11 (criteria), median 6 (criteria)).
* COS-STAD was not published until 2017; all included COS studies were carried out before this publication date. Therefore, this assessment is not a criticism of these studies or the study authors, but rather a baseline against which future comparisons of Cancer COS can be made.
* This current review provides guidance on how to compare a published COS to the standards (Table 2). This will further facilitate users to assess whether a COS has been well developed.
* Poor reporting often made it challenging to assess whether the minimum standards had been met.
* This study identified the need to separate standard number 9 into two criteria, considering scoring process and consensus definition separately. We recommend this separation for future users of COS-STAD.

1. **Introduction**

To make well-informed decisions about healthcare, we need to be able to compare and contrast research findings on the basis of the same outcomes. Core outcome sets (COS) represent the minimum important outcomes that should be measured and reported in all research studies for a specific condition. The use of COS will improve the quality of evidence used in health care decision-making, ultimately translating to improved health care for patients. For COS to be successfully implemented, they need to be easily accessible to researchers and other key groups, developed using rigorous methods, and reported clearly. The Core Outcome Measures in Effectiveness Trials (COMET) Initiative was set up to help achieve this. The development and continued maintenance of the COMET database [1, 2], through rigorous systematic reviews [3-7], means that COS are now easily accessible to users of COS. The COMET Handbook advocates the use of rigorous methods through an accumulation of current knowledge in the area of COS development, and will be updated periodically to continue to inform good methodological practice in this area [8]. Furthermore, the Core Outcome Set-STAndards for Development (COS-STAD) recommendations have been established to improve the methodological approach for developing COS and help users to assess whether a particular COS has been developed using a reasonable approach [9].

Research in the area of COS development is becoming more prevalent but it is still quite new; we would therefore expect improvements in methodological standards in the coming years. To be able to assess this, however, a baseline, against which we can compare future quality of COS development and measure improvements in methodological standards, is necessary. COS are developed across a wide range of disease areas. Cancer is currently the disease area with the highest number of published COS. Furthermore, cancer is a major cause of worldwide morbidity and mortality, has substantial variability in populations and treatments, and covers a wide range of diverse clinical areas. Treatments aim to cure the disease but are associated with multiple side effects; trials are commonly performed and major clinical oncology organisations exist worldwide. It is essential to ensure that outcomes in cancer trials reflect issues of importance to patients and health professionals through the use of well-developed COS. The aim of this study was therefore to provide a baseline of cancer COS standards of development against which future comparisons can be made.

1. **Methods**

**2.1 Identification of COS studies**

As of the end of 2017, a regularly updated systematic review of COS [3-7] had identified 307 published COS for research studies covering a wide range of different health areas. Only COS developed for cancer were eligible for inclusion in this evaluation.

**2.2 Pilot study**

COS-STAD contains 11 standards that are deemed to be the minimum design recommendations for all COS development projects [9]. The recommendations focus on three key domains: scope, stakeholders, and consensus process.

A pilot study was carried out to inform the process of assessment used in this study. One COS was randomly selected. Three of the authors (EG, PRW and JJK) independently read the COS and made an assessment against each of the 11 criteria listed in Table 1. Assessments were compared and a discussion occurred between the three authors to define how the process should be applied. This included the criteria used for assessment as well as the sources of information that would be used to identify supporting information. Following this, a further five randomly selected COS were independently assessed by two of the authors (EG and JJK), and the process was further refined. Of note, it was agreed that Criteria number 9 (“A scoring process and consensus definition is described a priori”) would be split into two for assessment purposes. It became apparent that in the five COS assessed in the pilot study that a study team could describe an a priori definition for one of these only (e.g. provide a description of a scoring process but not provide a consensus definition or vice versa), resulting in different assessments for each part of this criteria, and should therefore be assessed separately. Hereafter they will be referred to as 9a (scoring process) and 9b (consensus definition). A total of 12 criteria were therefore assessed in this study, representing the 11 minimum standards.

**2.3 Process**

Two reviewers (EG and JJK) independently assessed each of the cancer COS against the 12 criteria of development. Each criteria was assessed as yes (meeting that standard), no (not meeting that standard) or unsure (it was unclear whether the criteria had been met). Further details of the assessment are described in Table 2.

Verbatim text was extracted to support the assessment being made and to aid discussion. Assessments were compared, and a third reviewer consulted where there was uncertainty. PRW and JMB were consulted for methodological queries, and JMB (surgical oncologist) for clinical queries. Where the development process was described across multiple papers for an individual COS, a global assessment was made for each of the standards.

***2.3.1 Sources of information***

Articles describing the development of the COS were eligible for providing supporting information for each of the standards. This included main study publications as well as protocols. The COMET database was not used as a source of information in this study as it is a secondary source of information (populated by the systematic reviews used in this study to identify cancer COS) and does not contain any additional details necessary for this study.

**2.4 Data analysis and presentation of results**

The median and range were used to summarise the 12 criteria (representing the 11 minimum standards) met across all of the included Cancer COS studies. Percentage frequencies were used to report the number of studies that met each standard.

1. **Results**

Forty-nine cancer COS were included. An overview of the minimum standards assessments is provided in Table 3, and by study in Appendix 1. No COS met all of the 12 criteria representing the 11 minimum standards assessed in this study (range 4-11 (criteria), median 6 (criteria)).

**3.1 Scope specification**

All studies met all four standards for scope.

***3.1.1 Standard number 1: the research or practice setting(s) in which the COS is to be applied***

All studies met this criteria: 40/49 (82%) stated that the COS was intended for use in clinical trials, 4/49 (8%) stated that the COS was intended for use in clinical research, two (4%) stated that it was intended for use in clinical trials and routine clinical practice, and three (6%) for research and routine clinical practice.

***3.1.2 Standard number 2: the health condition(s) covered by the COS***

All studies met this criteria. Five COS were developed for cancer (non-specific), and four specifically for complications of the treatments of cancer. Nine COS were developed for prostate cancer, five for cancer of the colon and/or rectum and three for ovarian cancer. Two were developed for non-small cell lung cancer, two for Hodgkin’s and non-Hodgkin’s lymphomas, and two COS were developed for acute myeloid leukaemia. The remaining seventeen unique cancer COS areas are listed in Appendix 2.

***3.1.3 Standard number 3: the population(s) covered by the COS***

All studies met this criteria. In ten of the cancer COS, it was not explicitly stated that the COS was intended for an adult population only, but this was assumed because the cancer covered by the COS was an adult only cancer (e.g. prostate cancer). Further assumptions were made for an additional thirteen COS that related to cancers that may occur in children as well as adults, but did not make any statement about the population that the COS was intended. These thirteen COS were deemed to be implicitly for adults only because they did not refer to children specifically.

***3.1.4 Standard number 4: the intervention(s) covered by the COS***

All studies met this criteria. Where the authors referred to ‘evaluating treatments’, ‘cancer treatments’, or a COS for ‘all trials’, we took this to mean all interventions. Twenty-six (53%) COS were developed for all/any interventions, eight (16%) for drug interventions, four (8%) for procedures, three (6%) for surgical interventions, two (4%) for screening and the remaining six (12%) were for other specific interventions (pre-operative treatment strategies, adjuvant treatment, larynx preservation strategies, vaccination, exercise and compression interventions).

**3.2 Stakeholders involved**

Eight (16%) studies met all three standards for stakeholders involved.

***3.2.1 Standard number 5: those who will use the COS in research***

Thirty-six studies (74 %) met this standard. Assumptions were made based on the author affiliations (where it was clear that the authors contributed to the COS development process), or from the participant list affiliations when these were provided, for ten COS. These ten did not explicitly state that researchers were included as participants, but we assumed their inclusion when research institutions were listed as affiliations.

Seven studies (14%) did not state that those who will use the COS in research were involved in the COS development, and there was not enough evidence to support an assumption being made therefore these were categorised as unclear. Examples include where it was not clear that the authors were the panel/participants involved, (and no other statement was made in relation to stakeholders), participants were described as conference participants but their backgrounds were not described, or they were described as ‘experts’ without any further description.

Six studies (12%) did not meet this standard. In these instances, the stakeholders involved were clearly described and did not include those who will use the COS in research.

***3.2.2 Standard number 6: healthcare professionals (HCPs) with experience of patients with the condition***

Thirty-five studies (71%)clearlymet this standard. Assumptions were made for eight COS based on the author affiliations (where it was clear that the authors are the group who developed the COS), or from the participant list affiliations when these were provided. These eight did not explicitly state that HCPs were included as participants, but we assumed their inclusion when clinical care settings were listed as affiliations.

Eight (16%) did not state explicitly that HCPs were involved in the COS development and there was ambiguity in the description provided; these were, therefore, categorised as unclear. Examples echoed those given for Standard number 5.

Six studies (12%) did not meet this standard. In these instances, the stakeholders involved were clearly described and did not include HCPs in the process.

***3.2.3 Standard number 7: patients with the condition or their representatives***

Thirteen studies (27%) met this standard. Eight studies described including patient advocates or representatives, and five included patients themselves. The number of patients in six of these studies was not reported. The percentage of patients with the condition or their representatives in the remaining seven studies ranged from 9% to 68% in studies with mixed participants.

Thirty-five studies (71%) did not meet this standard. In these instances, the stakeholders involved were clearly described and did not include patients in the process. It was unclear in one study whether patients were included, as participants were described as conference delegates without any further description.

**3.3 Consensus process**

Two studies (4%) met all four standards [five criteria] for the consensus process.

***3.3.1 Standard number 8: the initial list of outcomes considered both healthcare professionals’ and patients’ views***

Eight studies (16%)met this standard. Two of these five included patient reported outcome (PRO) data (i.e. studies that have collected data from patients using a patient reported outcome measure (PROM)) as well as trial data, and therefore on face value have included the patient perspective in the process.

Five studies (10%) did not clearly state whose views were considered when generating the initial list, and so therefore have been categorised as unclear. One of these studies included PROM items in the review (PROM items only not PRO data). Patients may or may not have been involved in the PRO development, but it was beyond the scope of this study to research PROM development; we cannot therefore say if they considered patients’ views.

Thirty-six studies (74%) did not meet this standard and did not consider both HCPs ***and*** patients’ views when generating the initial list of outcomes used in the COS development. These studies considered trial data, clinical trials literature or clinical guidelines only (hence did not consider patients’ views).

***3.3.2 Standard number 9a: a scoring process was described a priori***

Five studies (10%) met this standard.

It was unclear whether the remaining 44 (90%) met this standard. Ten studies described a scoring process but it was not clear whether this process was determined a priori; 28 studies did not describe specific methods relating to scoring, and six studies did not describe the methods used at all.

***3.3.3 Standard number 9b: a consensus definition was described a priori***

Six studies (12%) met this standard.

It was unclear whether this standard was met for the remaining 43 (88%). In 26 studies it was unclear whether the consensus definition was defined a priori; eleven studies did not describe a process of consensus and therefore did not provide a definition, and the remaining six studies did not describe their consensus methods at all.

***3.3.4 Standard number 10: criteria for including/dropping/adding outcomes were described a priori***

Two studies (4%) met this standard.

It was unclear whether this standard was met for the remaining 47 studies (96%). The detail of all three elements of this standard was not clearly described in six studies, and for two studies that did describe this process, it was not possible to assess the a priori element of this criteria. Thirty-three studies had no description of this process, and six did not describe their methods overall, making it impossible to assess whether this criteria had been considered.

***3.3.5 Standard number 11: care was taken to avoid ambiguity of language used in the list of outcomes***

Five studies (10%) met this criteria. In one study, the outcomes were formulated with examples in parentheses for some patient reported outcomes; and in another study, the patient representatives received a glossary of terms prior to completing the survey. Although this information is limited, as COS-STAD is being applied retrospectively here, we have interpreted this as some consideration of the language used to describe outcomes. This criteria was clearly met in the other three studies, where the questionnaires were assessed, by patients, for face validity and comprehension before use.

For 44/49 studies (90%) there was no evidence that care was taken to avoid ambiguity of language, and therefore it was unclear whether they met this criteria.

1. **Discussion**

This is the first application of the COS-STAD criteria to studies that have developed COS. Forty-nine cancer COS were included. No COS met all of the minimum standards, with most studies meeting half of the standards. However, COS-STAD was not published until 2017, which was after all of these studies started. Therefore, this assessment is not a criticism of these studies or the study authors, but rather a baseline against which future comparisons can be made. Furthermore, this current review provides guidance on how to compare a published COS to the standards (Table 2). This will further facilitate users to assess whether a COS has been well developed.

Standards in the Scope domain were well met. One explanation for this could be that these scope criteria are synonymous with the ‘PIC’ part of the PICO format that is often used to formulate a research question [10]. For the purposes of this study, there was an assumption that a COS was developed for adults only unless stated otherwise. This was deemed to be clinically relevant for cancer COS, and consideration should be given to other disease areas and whether the same assumption is correct. When assessing standards in the scope domain, we observed that multiple COS were developed for some cancers, for example nine COS have been developed for prostate cancer alone. At times this may reflect unnecessary duplication, but often they had relevant and appropriate differences in scope; for example differences in stage or type such as advanced or localised cancer. COS-STAD can be used as a tool to help users of COS to assess whether a COS has been well developed. Users will need to use their own judgement regarding the applicability of the COS (scope) for the purpose they require [9].

The majority of COS studies did include those who will use the COS in research and health care professionals in the development process, while just over a quarter included patients or patient representatives in the process. This is reflective of COS in general, and is not a methodological problem specific to cancer. Recent research has shown an improvement in patient participation in more recently published COS [6, 11]. Although thirteen studies did include patients in COS development, there is still great variability in the level of participation of patients in COS development. This also needs to be taken into account when deciding if a COS was well developed.

The consensus process standards were the most difficult to compare against, particularly those that required an assessment of being specified a priori. Only four studies stated that they had a protocol, of which two were published and one we were able to obtain from study authors. In the remaining studies, because we were applying this retrospectively, we took it at face-value when the COS report stated that something was specified a priori because we were unable to verify this in a protocol. While making the protocol publicly available for a COS development study was not agreed upon as a COS-STAD minimum standard, it was suggested that the availability of a protocol would ensure that the methods are explicitly documented before the COS development project starts, thus promoting research integrity and transparency of the finalised COS.  This will make it easier for users to assess the COS against COS-STAD for future studies.

This study identified the need to consider the scoring process and consensus definition separately. We recommend this separation for future users of COS-STAD. With regards to criteria 9a (scoring), an observation was that in those that did not meet this standard, the method of development used for consensus in all of these studies was some form of meeting or workshop to decide on the COS, but detailed description of what was done at those meetings was missing. Further consideration is needed about the applicability and suitability of this criterion for all methods that might be used for COS development.

Standard number 10, criteria for including/dropping/adding outcomes, was only met by two studies. The a priori part of the assessment was unclear for a further two studies, but the majority did not mention anything about this process, or did not describe it in sufficient detail to be able to assess whether this criteria was met. This suggests that perhaps COS developers do not fully understand the implications or importance of this aspect of the development process, and detail might be lacking as it might be considered too much information for publication. It should be noted that a lack of reporting of methods does not necessarily mean that these methodological aspects were not considered. This standard is also reflected in COS-STAR as a reporting requirement [12]. However, COS-STAR was not published until 2016 which is after most of these included COS studies were published. It is therefore unrealistic to expect these reporting standards to be implemented in these studies, so it is to be hoped that evidence of consideration of this criteria (and indeed the other standards) will be included in study reports in the future.

COS-STAD focusses on the main design principles for COS development, while COS-STAR is exclusive to the reporting of COS studies. As already highlighted, issues with reporting were one of the main limitations of this study as it made it difficult, sometimes impossible, to assess consideration of the COS-STAD standards. We did obtain protocols when it was stated that a protocol was available, but this was very few. As such we were unable to distinguish development plans from poor final reporting. Furthermore, this meant that assumptions were sometimes made to enable a judgement to be made. Another limitation was that the studies being assessed were published prior to COS-STAD and therefore could not have been informed by the development standards. Another potential limitation is that we identified cancer COS from existing systematic reviews of COS, and did not conduct a separate search. To address this potential limitation, the list of COS was reviewed by a relevant expert (JMB) for completeness. The inclusion of cancer/COS methodology experts was a further strength of this study, as was the independent dual extraction and assessment of data.

This study aimed to provide a baseline of cancer COS standards of development against which future comparisons can be made. With the exception of Scope specification, there is much need for improvement. Poor reporting made it challenging to assess whether the minimum standards had been met for all Stakeholder involvement and Consensus process standards. With the publication of methodological evidence (for example [13, 14]) and guidance [8, 9, 15], as well as reporting standards [12], improvement is expected over time.

**Competing interests**

EG, JMB and PRW are members of the COMET Management Group and co-applicants on grants to support COMET and related work. JMB, JJK and PRW were involved in the development of COS-STAD. JMB, PRW and JJK have been involved in the development of COS.

**Funding**

This work was supported by the NIHR Senior Investigator Award (NF-SI\_0513-10025). Professor Paula Williamson is a National Institute for Health Research (NIHR) Senior Investigator. The views expressed in this article are those of the author(s) and not necessarily those of the NIHR, or the Department of Health and Social Care.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. JMB is supported by the MRC ConDuCT-II Hub and the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. JMB is also a NIHR Senior Investigator.

**Authors’ contributions**

EG, JJK and PRW conceived the idea for the study. EG, JJK and PRW designed the study. EG and JJK performed data extraction assisted by JMB. All authors carried out the analysis and interpreted data. EG wrote the first draft of the manuscript, with subsequent input from all authors. All authors commented on and approved the final manuscript.

**References**

1. Gargon, E., et al., *The COMET Initiative database: progress and activities from 2011 to 2013.* Trials, 2014. **15**(1): p. 279.

2. Gargon, E., et al., *The COMET initiative database: progress and activities update (2014).* Trials, 2015. **16**(1): p. 515.

3. Gargon, E., et al., *Choosing important health outcomes for comparative effectiveness research: a systematic review.* PLoS ONE, 2014. **9**(6): p. e99111.

4. Gorst, S.L., et al., *Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and User Survey.* PLoS ONE, 2016. **11**(1): p. e0146444.

5. Gorst, S.L., et al., *Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and Identification of Gaps.* PLoS One, 2016. **11**(12): p. e0168403.

6. Davis, K., et al., *Choosing important health outcomes for comparative effectiveness research: An updated systematic review and involvement of low and middle income countries.* PLoS One, 2018. **13**(2): p. e0190695.

7. Gargon, E., et al., *Choosing important health outcomes for comparative effectiveness research: 4th annual update to a systematic review of core outcome sets for research* PLoS One Submitted

8. Williamson, P.R., et al., *The COMET Handbook: version 1.0.* Trials, 2017. **18**(Suppl 3): p. 280.

9. Kirkham, J.J., et al., *Core Outcome Set-STAndards for Development: The COS-STAD recommendations.* PLoS Med, 2017. **14**(11): p. e1002447.

10. Sackett, D., et al., *Evidence-based medicine: How to practice and teach EBM*, ed. C. Livingston. 1997, New York.

11. Biggane, A.M., et al., *Survey indicated that core outcome set development is increasingly including patients, being conducted internationally and using Delphi surveys.* Trials, 2018. **19**(1): p. 113.

12. Kirkham, J.J., et al., *Core Outcome Set-STAndards for Reporting: The COS-STAR Statement.* PLoS Med, 2016. **13**(10): p. e1002148.

13. Gargon, E., et al., *Higher number of items associated with significantly lower response rates in COS Delphi surveys.* J Clin Epidemiol, 2018.

14. MacLennan, S., et al., *A randomized trial comparing three Delphi feedback strategies found no evidence of a difference in a setting with high initial agreement.* J Clin Epidemiol, 2018. **93**: p. 1-8.

15. Dodd, S., et al., *A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery.* Journal of Clinical Epidemiology, 2018. **96**: p. 84-92.

16. Glynne-Jones, R., et al., *Alternative clinical end points in rectal cancer--are we getting closer?* Ann Oncol, 2006. **17**(8): p. 1239-48.

17. Auvinen, A., et al., *Prospective evaluation plan for randomised trials of prostate cancer screening. The International Prostate Cancer Screening Trial Evaluation Group.* J Med Screen, 1996. **3**(2): p. 97-104.

18. Denis, L., et al., *Planning controlled clinical trials. Prostatic cancer.* Urology, 1997. **49**(4A Suppl): p. 15-26.

19. Dawson, N.A., *Apples and oranges: building a consensus for standardized eligibility criteria and end points in prostate cancer clinical trials.* J Clin Oncol, 1998. **16**(10): p. 3398-405.

20. Rajkumar, S.V., et al., *Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Workshop Consensus Panel 1.* Blood, 2011. **117**(18): p. 4691-5.

21. Chow, E., et al., *International consensus on palliative radiotherapy endpoints for future clinical trials in bone metastases.* Radiother Oncol, 2002. **64**(3): p. 275-80.

22. *International bone metastases consensus on endpoint measurements for future clinical trials: proceedings of the first survey and meeting (work in progress) International Bone Metastases Consensus Working Party.* Clin Oncol (R Coll Radiol), 2001. **13**(2): p. 82-4.

23. Hesketh, P.J., et al., *Methodology of antiemetic trials: response assessment, evaluation of new agents and definition of chemotherapy emetogenicity.* Support Care Cancer, 1998. **6**(3): p. 221-7.

24. Pallis, A.G., et al., *EORTC workshop on clinical trial methodology in older individuals with a diagnosis of solid tumors.* Ann Oncol, 2011. **22**(8): p. 1922-6.

25. Prorok, P.C. and P.M. Marcus, *Cancer screening trials: nuts and bolts.* Semin Oncol, 2010. **37**(3): p. 216-23.

26. Punt, C.J., et al., *Endpoints in adjuvant treatment trials: a systematic review of the literature in colon cancer and proposed definitions for future trials.* J Natl Cancer Inst, 2007. **99**(13): p. 998-1003.

27. Wils, J., et al., *Evaluation of clinical efficacy of new medical treatments in advanced colorectal cancer. Results of a workshop organized by the EORTC GITCCG. European Organization for Research and Treatment of Cancer. Gastrointestinal Tract Cancer Cooperative Group.* Tumori, 1998. **84**(3): p. 335-47.

28. McVie, J.G. and K.M. de Bruijn, *Methodology of antiemetic trials.* Drugs, 1992. **43 Suppl 3**: p. 1-5.

29. Miller, A.B., et al., *Reporting results of cancer treatment.* Cancer, 1981. **47**(1): p. 207-14.

30. Llovet, J.M., et al., *Design and endpoints of clinical trials in hepatocellular carcinoma.* J Natl Cancer Inst, 2008. **100**(10): p. 698-711.

31. Pagliusi, S.R. and M. Teresa Aguado, *Efficacy and other milestones for human papillomavirus vaccine introduction.* Vaccine, 2004. **23**(5): p. 569-78.

32. Lefebvre, J.L. and K.K. Ang, *Larynx preservation clinical trial design: key issues and recommendations-a consensus panel summary.* Int J Radiat Oncol Biol Phys, 2009. **73**(5): p. 1293-303.

33. Middleton, R.G., et al., *Prostate Cancer Clinical Guidelines Panel Summary report on the management of clinically localized prostate cancer. The American Urological Association.* J Urol, 1995. **154**(6): p. 2144-8.

34. Schellhammer, P., et al., *Assessment of endpoints for clinical trials for localized prostate cancer.* Urology, 1997. **49**(4A Suppl): p. 27-38.

35. Gridelli, C., et al., *Treatment of advanced non-small-cell lung cancer patients with ECOG performance status 2: results of an European Experts Panel.* Ann Oncol, 2004. **15**(3): p. 419-26.

36. Gridelli, C., et al., *Maintenance treatment of advanced non-small-cell lung cancer: results of an international expert panel meeting of the Italian association of thoracic oncology.* Lung Cancer, 2012. **76**(3): p. 269-79.

37. Anderson, K.C., et al., *Clinically relevant end points and new drug approvals for myeloma.* Leukemia, 2008. **22**(2): p. 231-9.

38. Kulke, M.H., et al., *Future directions in the treatment of neuroendocrine tumors: consensus report of the National Cancer Institute Neuroendocrine Tumor clinical trials planning meeting.* J Clin Oncol, 2011. **29**(7): p. 934-43.

39. Bellm, L.A., et al., *Defining clinically meaningful outcomes in the evaluation of new treatments for oral mucositis: oral mucositis patient provider advisory board.* Cancer Invest, 2002. **20**(5-6): p. 793-800.

40. du Bois, A., et al., *2004 consensus statements on the management of ovarian cancer: final document of the 3rd International Gynecologic Cancer Intergroup Ovarian Cancer Consensus Conference (GCIG OCCC 2004).* Ann Oncol, 2005. **16 Suppl 8**: p. viii7-viii12.

41. Stuart, G.C., et al., *2010 Gynecologic Cancer InterGroup (GCIG) consensus statement on clinical trials in ovarian cancer: report from the Fourth Ovarian Cancer Consensus Conference.* Int J Gynecol Cancer, 2011. **21**(4): p. 750-5.

42. Thigpen, T., et al., *First-line therapy in ovarian cancer trials.* Int J Gynecol Cancer, 2011. **21**(4): p. 756-62.

43. Comenzo, R.L., et al., *Consensus guidelines for the conduct and reporting of clinical trials in systemic light-chain amyloidosis.* Leukemia, 2012. **26**(11): p. 2317-25.

44. Dixon, D.O., et al., *Reporting outcomes in Hodgkin's disease and lymphoma.* J Clin Oncol, 1987. **5**(10): p. 1670-2.

45. Cheson, B.D., et al., *Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia.* J Clin Oncol, 2003. **21**(24): p. 4642-9.

46. Cheson, B.D., et al., *Revised response criteria for malignant lymphoma.* J Clin Oncol, 2007. **25**(5): p. 579-86.

47. Cheson, B.D., et al., *Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group.* J Clin Oncol, 1999. **17**(4): p. 1244.

48. Adelstein, D.J., et al., *Transoral resection of pharyngeal cancer: summary of a National Cancer Institute Head and Neck Cancer Steering Committee Clinical Trials Planning Meeting, November 6-7, 2011, Arlington, Virginia.* Head Neck, 2012. **34**(12): p. 1681-703.

49. Chen, R.C., et al., *Recommended patient-reported core set of symptoms to measure in prostate cancer treatment trials.* J Natl Cancer Inst, 2014. **106**(7).

50. Chera, B.S., et al., *Recommended patient-reported core set of symptoms to measure in head and neck cancer treatment trials.* J Natl Cancer Inst, 2014. **106**(7).

51. Donovan, K.A., et al., *Recommended patient-reported core set of symptoms and quality-of-life domains to measure in ovarian cancer treatment trials.* J Natl Cancer Inst, 2014. **106**(7).

52. Fraser, J.F., et al., *Reporting standards for endovascular chemotherapy of head, neck and CNS tumors.* J Neurointerv Surg, 2013. **5**(5): p. 396-9.

53. Glynne-Jones, R., et al., *End points in anal cancer: hopes for a common language.* J Clin Oncol, 2014. **32**(12): p. 1281-2.

54. Reeve, B.B., et al., *Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials.* J Natl Cancer Inst, 2014. **106**(7).

55. Wildiers, H., et al., *End points and trial design in geriatric oncology research: a joint European organisation for research and treatment of cancer--Alliance for Clinical Trials in Oncology--International Society Of Geriatric Oncology position article.* J Clin Oncol, 2013. **31**(29): p. 3711-8.

56. Haeusler, G.M., et al., *Core outcomes and definitions for pediatric fever and neutropenia research: A consensus statement from an international panel.* Pediatric Blood & Cancer, 2015. **62**(3): p. 483-489.

57. van den Bos, W., et al., *Focal therapy in prostate cancer: international multidisciplinary consensus on trial design.* Eur Urol, 2014. **65**(6): p. 1078-83.

58. van den Bos, W., et al., *Salvage ablative therapy in prostate cancer: international multidisciplinary consensus on trial design.* Urol Oncol, 2015. **33**(11): p. 495.e1-7.

59. Gerritsen, A., et al., *Developing a core set of patient-reported outcomes in pancreatic cancer: A Delphi survey.* European journal of cancer (Oxford, England : 1990), 2016. **57**(arv, 9005373): p. 68-77.

60. Partsch, H., et al., *Clinical trials needed to evaluate compression therapy in breast cancer related lymphedema (BCRL). Proposals from an expert group.* Int Angiol, 2010. **29**(5): p. 442-53.

61. McNair, A.G.K., et al., *Core Outcomes for Colorectal Cancer Surgery: A Consensus Study.* PLoS Medicine, 2016. **13**(8).

62. Avery, K.N.L., et al., *Development of a Core Outcome Set for Clinical Effectiveness Trials in Esophageal Cancer Resection Surgery.* Annals of Surgery, 2017.

63. Kamat, A.M., et al., *Definitions, End Points, and Clinical Trial Designs for Non-Muscle-Invasive Bladder Cancer: Recommendations From the International Bladder Cancer Group.* Journal of Clinical Oncology, 2016. **34**(16): p. 1935-44.

64. Karam, A., et al., *Fifth Ovarian Cancer Consensus Conference of the Gynecologic Cancer InterGroup: first-line interventions.* Annals of Oncology, 2017. **28**(4): p. 711-717.

65. Kilari, D., et al., *Designing exercise clinical trials for older adults with cancer: Recommendations from 2015 Cancer and Aging Research Group NCI U13 Meeting.* Journal of Geriatric Oncology, 2016. **7**(4): p. 293-304.

66. MacLennan, S., et al., *A core outcome set for localised prostate cancer effectiveness trials.* BJU Int, 2017. **120**(5b): p. E64-e79.

67. Ong, W.L., et al., *A Standard Set of Value-Based Patient-Centered Outcomes for Breast Cancer: The International Consortium for Health Outcomes Measurement (ICHOM) Initiative.* JAMA Oncology, 2017. **3**(5): p. 677-685.

68. van der Poel, H.G., et al., *Sentinel node biopsy for prostate cancer: report from a consensus panel meeting.* BJU International, 2017. **120**(2): p. 204-211.

69. Dohner, H., et al., *Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel.* Blood, 2017. **129**(4): p. 424-447.

**Table 1: Original COS-STAD criteria**

|  |  |  |
| --- | --- | --- |
| **DOMAIN** | **STANDARD NUMBER** | **STANDARD** |
| Scope specification | 1 | The research or practice setting(s) in which the COS is to be applied  |
| 2 | The health condition(s) covered by the COS  |
| 3 | The population(s) covered by the COS  |
| 4 | The intervention(s) covered by the COS |
| Stakeholders involved | 5 | Those who will use the COS in research  |
| 6 | Healthcare professionals with experience of patients with the condition  |
| 7 | Patients with the condition or their representatives  |
| Consensus process | 8 | Initial list of outcomes considered both healthcare professionals’ and patients’ views |
| 9 | A scoring process and consensus definition was described a priori |
| 10 | Criteria for including/dropping/adding outcomes were described a priori |
| 11 | Care was taken to avoid ambiguity of language used in the list of outcomes |

**Table 2: Assessment criteria: guidance on how to compare a published COS to the standards**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **DOMAIN** | **STANDARD NUMBER** | **STANDARD** | **FEATURES TO BE CONSIDERED** | **STANDARD MET** | **UNCLEAR IF STANDARD IS MET** | **STANDARD NOT MET** |
| Scope specification | 1 | The research or practice setting(s) in which the COS is to be applied should be considered  |  Descriptions include (but not limited to) health research, specific study types (e.g. trials, observational studies, and longitudinal studies), routine care, audit, registries. The developed COS might apply to a single setting or a combination of settings. | Evidence that the setting(s) (or context of use) in which the COS is to be applied has been considered.  | The setting(s) (or context of use) in which the COS is to be applied is unclear or is not described.  | An explicit statement that this was not considered. |
| 2 | The health condition(s) covered by the COS should be considered | This could be general (e.g. cancer) or specific (e.g. lung cancer).  | Evidence that he health condition(s) has been considered. | The health condition(s) is unclear or is not described. |  An explicit statement that this was not considered. |
| 3 | The population(s) covered by the COS should be considered | Characteristics should be relevant to the population. The following were considered for cancer COS: 1. Disease stage
2. Age of population
 | Evidence that the population has been considered. | Population characteristics are discussed (e.g. in the introduction or discussion of the paper) but not in the context of the COS being developed. Population is unclear or not described. | An explicit statement that this was not considered. |
| 4 | The intervention(s) covered by the COS should be considered | 1. This could be all/any intervention or specific intervention types (e.g. drug, surgery)
 | Evidence that the intervention has been considered. | Interventions are discussed (e.g. in the introduction or discussion of the paper) but not in the context of the COS being developed. Intervention is unclear or not described. | An explicit statement that this was not considered. |
| Stakeholders involved | 5 | Those who will use the COS in research  | Descriptions include (but not limited to): 1. Clinical trialists (including trial investigators, members of trial groups)
2. Researchers (including methodologists, experts with expertise in the conduct of studies, scientists)
3. Industry representatives
 | COS paper includes description of i. ***OR*** ii. ***OR*** iii.  | It was not clear whether those who will use the COS in research were included.  | It was clear that those who will use the COS in research were not included. |
| 6 | Healthcare professionals with experience of patients with the condition  | Health professionals (including all disciplines such as oncologists, urologists, pharmacists, surgeons, nurses etc.) ^ | COS paper includes description of  | It was not clear whether healthcare professionals with experience of patients with the condition were included.  | It was clear that healthcare professionals with experience of patients were not included. |
| 7 | Patients with the condition or their representatives  | Descriptions include (but not limited to):1. Patients with the condition
2. Patient representatives
3. Patient advocates
4. Parents
5. Carers
 | COS paper includes description of i. ***OR*** ii. ***OR*** iii. ***OR*** iv. ***OR*** v. | It was not clear whether patients/patient representatives were included.  | It was clear that patients/ patient representatives were not included. |
| Consensus process | 8 | Initial list of outcomes considered both healthcare professionals’ and patients’ views | 1. Health care professionals (see #6 descriptors)
2. Patient (see #7 descriptors)
 | Initial list clearly included i. ***AND*** ii.  | It was not clear whose views’ were considered.  | The initial list only included i. ***OR*** ii. The initial list did not include either view.  |
| 9a | A scoring process was described a priori\* | Any description of a scoring system to rate outcomes [8], including (but not limited to) 1. A Likert scale or similar
2. Ranking of outcomes
3. Allocation of points.
 | It does describe a scoring process, and it is clear that this ***was*** a priori\* | It does describe a scoring process but does ***not*** state whether this was a priori (or when it was described/defined).It does ***not*** mention a scoring process.  | It clearly has not used a scoring process. It does mention a scoring process, BUT states this ***was* *not*** a priori, or text about methods confirms it.  |
| 9b | A consensus definition was described a priori\* | There are numerous ways to define the consensus criteria, commonly these relate to a mean or median value for each outcome or a percentage of participants scoring an outcome as important [8].  | It does provide a definition of consensus, and it is clear that this ***was*** a priori\*. | It does mention consensus (or a synonym) or provides a description of definition, but it is ***not*** clear whether this was a priori/when it was described/defined.It does ***not*** mention consensus.  | It clearly has not used consensus methods.It does mention consensus (or a synonym), does provide further description of definition, BUT states this ***was* *not*** a priori, or text about methods confirms it. |
| 10 | Criteria for including/dropping/adding outcomes were described a priori\* | A description of this process (e.g. a description of including AND dropping AND adding outcomes).  | It does include a description of including AND dropping AND adding outcomes, as well as stating this ***was*** a priori\*. | It does include a description of including AND dropping AND adding outcomesbut **does not** state whether this was a priori/when it was described/defined.It does ***not*** include any description about including/adding/dropping outcomes.It does ***not*** describe ***all*** 3 elements of this criterion (including AND dropping AND adding outcomes). | It does include a description of including/ dropping/ adding outcomes, BUT states this ***was* *not*** a priori, or text about methods confirms it. |
| 11 | Care was taken to avoid ambiguity of language used in the list of outcomes | Consideration for language should be taken into account when describing outcomes to different stakeholder groups. An example might be the use of both plain language descriptions and medical terms, with these pilot tested for understanding. [9]  | A clear description of methods/steps taken to avoid ambiguity of language.  | Some suggestion that this may have been done, but not clearly described. No evidence of consideration given to ambiguity of language. | An explicit statement that this was not considered. |

**\*** ”a priori” as assessed by inclusion in a protocol, or when stated ‘a priori’ in the study report without a protocol to verify this (we have taken this at face value).

^Clinician, physicians, clinical investigators, and medical faculty were all different descriptors used for this in the cancer COS papers.

**Table 3: Cancer COS minimum standards assessments (N=49)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **DOMAIN** | **STANDARD NUMBER** | **STANDARD** | **STANDARD MET****N (%)**  | **STANDARD UNCLEAR****N (%)** | **STANDARD NOT MET****N (%)** |
| Scope specification | 1 | The research or practice setting(s) in which the COS is to be applied  | 49 (100) | 0 | 0 |
| 2 | The health condition(s) covered by the COS  | 49 (100) | 0 | 0 |
| 3 | The population(s) covered by the COS  | 49 (100) | 0 | 0 |
| 4 | The intervention(s) covered by the COS | 49 (100) | 0 | 0 |
| Stakeholders involved | 5 | Those who will use the COS in research  | 36 (74%) | 7 (14) | 6 (12) |
| 6 | Healthcare professionals with experience of patients with the condition  | 35 (71) | 8 (16) | 6 (12) |
| 7 | Patients with the condition or their representatives  | 13 (27) | 1 (2) | 35 (71) |
| Consensus process | 8 | Initial list of outcomes considered both healthcare professionals’ and patients’ views | 8 (16) | 5 (10) | 36 (74) |
| 9a | A scoring process was described a priori | 5 (10) | 44 (90) | 0  |
| 9b | A consensus definition was described a priori | 6 (12) | 43 (88) | 0  |
| 10 | Criteria for including/dropping/adding outcomes were described a priori | 2 (4) | 47 (96) | 0  |
| 11 | Care was taken to avoid ambiguity of language used in the list of outcomes | 5 (10) | 44 (90) | 0  |

**Appendix 1: Cancer COS minimum standards: assessment by study**

|  |
| --- |
| **Key **= standard met **O**  = unclear whether standard is met  **X** = Standard not met  |
|  | **Scope** | **Stakeholders** | **Consensus Process** |
| **Standard number**  | **1** | **2** | **3** | **4** | 5 | **6** | **7** | **8** | **9a** | **9b** | **10** | **11** |
| Glynne-Jones et al. (2006). [16] |  |  |  |  | x | x | x | x | o  | o  | o  | o  |
| Auvinen et al. (1996). [17] |  |  |  |  |  | x | X | x | o  | o  | o  | o  |
| Denis et al. (1997). [18] |  |  |  |  | o | o | X | x | o  | o  | o  | o  |
| Dawson (1998). [19] |  |  |  |  |  |  | X | x | o | o | o  | o  |
| Rajkumar et al. (2011). [20] |  |  |  |  | o | o | X | x | o  | o | o  | o  |
| Chow et al. (2002). [21]International Bone Metastases Consensus Working Party. (2001). [22] |  |  |  |  |  | o | X | x | o | o | o  | o  |
| Hesketh et al. (1998). [23] |  |  |  |  | o | o | X | x | o  | o | o  | o |
| Pallis et al. (2011). [24] |  |  |  |  |  |  | X | x | o  | o | o  | o  |
| Prorok and Marcus (2010). [25] |  |  |  |  |  | x | X | x | o  | o  | o  | o  |
| Punt et al. (2007). [26] |  |  |  |  |  |  | X | x | o  | o | o  | o  |
| Wils et al. (1998). [27] |  |  |  |  | o | o | X | x | o  | o  |  o | o  |
| McVie and de Bruijn (1992).[28] |  |  |  |  |  | x | X | x | o  | o  | o  | o  |
| Miller et al. (1981). [29] |  |  |  |  |  |  | X | x |  o | o  |  o | o  |
| Llovet et al. (2008). [30] |  |  |  |  | o | o | X | x | o  | o  | o  | o  |
| Pagliusi and Aguado (2004). [31] |  |  |  |  |  | x | X | x |  o | o  | o  | o  |
| Lefebvre et al. (2009). [32] |  |  |  |  |  |  | X | x | o  | o |  o | o  |
|  Middleton et al. (1995). [33]Schellhammer et al. (1997). [34] |  |  |  |  |  | o | X | x | o  | o | o  | o  |
| Gridelli et al. (2004). [35] |  |  |  |  |  |  | X | x | o  | o | o  | o  |
| Gridelli et al. (2012). [36] |  |  |  |  |  |  | X | x | o  | o | o  | o  |

|  |
| --- |
| **Key **= standard met **O**  = unclear whether standard is met  **X** = Standard not met |
|  | **Scope** | **Stakeholders** | **Consensus Process** |
| **Standard number**  | **1** | **2** | **3** | **4** | 5 | **6** | **7** | **8** | **9a** | **9b** | **10** | **11** |
| Anderson et al. (2008). [37] |  |  |  |  |  |  |  | x | o  | o | o  | o  |
| Kulke et al. (2011). [38] |  |  |  |  |  |  |  | x | o  | o | o  | o  |
| Bellm et al. (2002). [39] |  |  |  |  |  |  |  | x | o | o | o  | o  |
| du Bois et al. (2005) [40]Stuart et al. (2011). [41]Thigpen et al. (2011). [42] |  |  |  |  |  | x | x | o | o | o | o | o  |
| Comenzo et al. (2012). [43] |  |  |  |  |  |  | x | x | o  | o | o  | o  |
| Dixon et al. (1987). [44] |  |  |  |  |  |  | x | x |  o | o  | o  | o  |
| Cheson et al. (2003). [45]  |  |  |  |  |  |  | x | x | o  | o  | o  | o  |
| Cheson et al (2007) [46]Cheson et al (1999) [47] |  |  |  |  |  |  | x | x | o  | o  | o  | o  |
| Adelstein et al. (2012). [48] |  |  |  |  |  |  | x | x | o  | o | o  | o  |
| Chen et al. (2014). [49] |  |  |  |  |  |  |  |  | o | o | o  | o  |
| Chera et al. (2014). [50] |  |  |  |  |  |  |  |  | o | o | o |  o |
| Donovan et al. (2014). [51] |  |  |  |  | x |  |  |  | o  | o | o  | o  |
| Fraser et al. (2013). [52] |  |  |  |  |  |  | x | o |  o | o | o  | o  |
| Glynne-Jones et al. (2014). [53] |  |  |  |  |  |  | x | x | o  | o  | o  | o  |
| Reeve et al. (2014). [54] |  |  |  |  |  |  |  |  | o | o | o  | o  |
| Wildiers et al. (2013). [55] |  |  |  |  | o |  | x | x | o  | o  | o  | o  |
| Haeusler et al. (2015). [56] |  |  |  |  |  |  |  | x | o |  | o |  |
| van den Bos et al. (2014). [57] |  |  |  |  |  |  | x | o | o | o | o  | o  |
| Van den Bos et al. (2015). [58] |  |  |  |  |  |  | x | x | o  | o | o | o  |
| Gerritsen et al. (2016). [59] |  |  |  |  | x |  |  | o |  |  |  |  |
| Partsch et al. (2010). [60] |  |  |  |  | o |  | x | x | o  | o  | o  | o  |

|  |
| --- |
| **Key **= standard met **O**  = unclear whether standard is met  **X** = Standard not met |
|  | **Scope** | **Stakeholders** | **Consensus Process** |
| **Standard number**  | **1** | **2** | **3** | **4** | 5 | **6** | **7** | **8** | **9a** | **9b** | **10** | **11** |
| McNair et al. (2016). [61] |  |  |  |  | x |  |  |  |  |  | o |  |
| Avery et al. (2017). [62] |  |  |  |  | X |  |  |  |  |  | o |  |
| Kamat et al. (2016). [63] |  |  |  |  |  |  | X | X | o | o | o | o |
| Karam et al. (2017). [64] |  |  |  |  |  | o | o | o | o | o | o | o |
| Kilari et al. (2016). [65] |  |  |  |  |  |  | X | X |  | o | o | o |
| MacLennan et al. (2017). [66] |  |  |  |  | X |  |  |  |  |  |  |  |
| Ong et al. (2017). [67] |  |  |  |  |  |  |  |  |  |  | o | o |
| van der Poel et al. (2017). [68] |  |  |  |  |  |  | X | X | o | o | o | o |
| Dohner et al. (2017). [69] |  |  |  |  |  |  | x | x | o | o | o | o |

**Appendix 2: Cancer types covered by 49 COS reviewed**

|  |  |
| --- | --- |
| **Disease name**  | **n** |
| Prostate cancer | 9 |
| Cancer of the colon and/or rectum | 5 |
| Cancer | 5 |
| Ovarian cancer | 3 |
| Non-small-cell lung cancer | 2 |
| Acute Myeloid Leukaemia | 2 |
| Chemotherapy induced nausea and vomiting  | 2 |
| Hodgkin’s and non-Hodgkin’s lymphomas | 2 |
| Cancer: solid tumors | 1 |
| Myeloma  | 1 |
| Multiple myeloma | 1 |
| Bone metastases  | 1 |
| Hepatocellular Carcinoma | 1 |
| Human papillomavirus (Cervical cancer) | 1 |
| Locally advanced laryngeal and hypopharyngeal cancer. | 1 |
| Neuroendocrine tumors (NETs) | 1 |
| Oral mucositis (OM) as a complication of chemotherapy and radiotherapy | 1 |
| Systemic immunoglobulin light-chain amyloidosis (AL) | 1 |
| Pharyngeal cancers | 1 |
| Head and neck cancer | 1 |
| Head, neck and CNS tumors | 1 |
| Fever and neutropenia (FN) as a complication of the treatment of childhood cancers | 1 |
| Pancreatic cancer  | 1 |
| Non–muscle-invasive bladder cancer | 1 |
| Esophageal cancer | 1 |
| Breast cancer  | 1 |
| Breast cancer related lymphedema  | 1 |
| **TOTAL**  | **49** |