



## Original Article

# The standards of obstetrics and gynecology core outcome sets: A scoping review



Jiyuan Shi<sup>a,b,1</sup>, Ya Gao<sup>a,c,1</sup>, Shuang Wu<sup>d</sup>, MingMing Niu<sup>a,b</sup>, Yamin Chen<sup>a,b</sup>, Meili Yan<sup>a,b</sup>, Ziwei Song<sup>a,b</sup>, Hui Feng<sup>d</sup>, Junhua Zhang<sup>e,\*</sup>, Jinhui Tian<sup>b,c,\*</sup>

<sup>a</sup> School of Nursing, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

<sup>b</sup> Evidence-Based Medicine Center, Lanzhou University, Lanzhou, Gansu, China

<sup>c</sup> WHO Collaborating Center for Guideline Implementation and Knowledge Translation, Lanzhou, Gansu, China

<sup>d</sup> Xiangya school of nursing, Central South University, Changsha, Hunan, China

<sup>e</sup> Chinese Clinical Trials Core Outcome Set Research Center, Evidence-Based Medicine Center, Tianjin University of Traditional Chinese Medicine, Tianjin, China

## ARTICLE INFO

## Article history:

Received 22 February 2021

Revised 7 September 2021

Accepted 9 September 2021

Available online 28 September 2021

## Keywords:

Core outcome set

Obstetrics

Gynecology

Research methodology

COS

Report methodology

## ABSTRACT

**Background:** Core outcome sets (COSs) are the minimum outcomes which should be measured and reported by researchers investigating a specific condition. The definition of standards of COSs vary across different health-related areas. This investigated the characteristics of COSs regarding obstetrics and gynecology (OG) and examined the reports and designs of standards of OG COSs.

**Methods:** A comprehensive search was conducted on the COMET database on December 20, 2019 to identify systematic reviews on COSs. Two reviewers independently evaluated whether the reported OG COS met the reporting requirements as stipulated in the Core Outcome Set-STAndards for Reporting (COS-STAR) statement checklist and the minimum design recommendations as outlined in the Core Outcome Set-STAndards for Development (COS-STAD) checklist.

**Results:** Forty-four OG COSs related to 26 topics were identified. None of them met all the 25 standards of COS-STAR statement which representing 18 items considered essential for transparent and complete reporting list for all COS studies (range: 6.0–24.0, median: 14.0). The compliance rates to 16 standards of methods and result sections ranged from 27.3%–68.2%. Total COS-STAR compliance items for OG COSs with the prior protocol was significantly higher than without prior protocol (MD = 3.846, 95% CI: 0.835–6.858,  $P = 0.012$ ). None of the OG COSs met all the 12 criteria in the COS-STAD minimum standards (range: 3.0–11.0, median: 5.0). The compliance rates for all three standards of stakeholders involved and all four standards of the consensus process were lower than 60%.

**Conclusions:** Methodological and reporting standards of OG COSs should be improved.

© 2021 Korea Institute of Oriental Medicine. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## 1. Introduction

Selection of outcomes that adequately reflect issues relevant to patients and health care professionals is essential when designing clinical trials to directly compare the effects of different interventions.<sup>1,2</sup> Core outcome sets (COSs) represent the minimum im-

portant outcomes that should be measured and reported in all research studies on a specific condition.<sup>1</sup> Use of high-quality COSs in clinical trials can improve comparability between similar trials, reduces selective reporting of results, increases the relevance of trial and systematic review results, and enhances the quality of evidence used in health care decision-making.<sup>3</sup> This ultimately improves delivery of health care support to patients.<sup>4</sup> COSs are important for the design, implementation, and reporting of randomized trials, systematic reviews, and other forms of research.<sup>5–7</sup>

Rapid increase in the number of COSs covering a wide range of different health-related areas has resulted in significant variation of the standards of COSs definition.<sup>6,8–13</sup> Previous studies report that high-quality set of COSs should include a comprehensive scop-

\* Corresponding authors at: Chinese Clinical Trials Core Outcome Set Research Center, Evidence-Based Medicine Center, Tianjin University of Traditional Chinese Medicine, Tianjin, China (J. Zhang); Evidence-Based Medicine Center, Lanzhou University, Lanzhou, Gansu, China (J. Tian).

E-mail addresses: [zjhtcm@foxmail.com](mailto:zjhtcm@foxmail.com) (J. Zhang), [15038059737@163.com](mailto:15038059737@163.com) (J. Tian).

<sup>1</sup> These authors contributed equally to this work.

ing process and a consensus process that involves multi-stakeholder groups.<sup>14–16</sup> However, studies have not explored whether the published COSs undertook a systematic review of existing outcomes or fully considered the views of different stakeholder groups on the COSs.<sup>17</sup> The Core Outcome Set-STAndards for Development (COS-STAD) project explored aspects of COSs development to establish minimum standards and apply them regardless of the consensus method chosen.<sup>15</sup> Recommendations have been established to improve the methodological approach for developing COSs and help users assess applicability of a particular COSs.<sup>15</sup> In addition, the Core Outcome Set-STAndards for Reporting (COS-STAR) statement was established using the recommended approach for developing medical reporting guidelines.<sup>18</sup> COS-STAD focuses on the principles of design associated with COSs development, whereas the COS-STAR focuses on reporting of COSs development studies. A recent pilot study used COS-STAD standards to assess the design quality of cancer COSs and reported that the scoping process and consensus process of cancer COSs were not reliable.<sup>19</sup> Moreover, none of the studies met all the 12 standards representing the 11 minimum standards assessed. Therefore, it is important to explore ways of improving the existing standards of COSs development.

Obstetrics and gynecology (OG) is a field in clinical medicine that focuses on physiological and pathological changes of the female reproductive system, and fertility regulation.<sup>20–22</sup> Accurately-developed COSs should ensure that outcomes in OG trials prioritize relevant issues for both patients and health care professionals over other disciplines.<sup>21–25</sup> The primary aim of the current study was to use COS-STAD and COS-STAR checklists to assess the methodological and reporting standards of OG COSs. Secondary objectives of the study included exploring factors that affect quality of developed COSs standards and exploring approaches for improving baseline standards for obstetrics and gynecology COSs development.

## 2. Methods

### 2.1. Inclusion criteria

OG COSs that had developed or applied methodologies to determine which outcome areas or outcomes should be measured, or OG COSs important in clinical trials or other forms of health research were included in the current study. We identified the OG diseases according to the ICD-11 criteria. The latest updated version of COSs was used. Studies that; only reported use of a COS, were systematic reviews of clinical trials, focused on systematic reviews and surveys of outcomes measured in clinical trials or quantitative descriptions (such as frequency) of outcomes, or protocol of COS, were excluded from the study. The inclusion criteria did not include limitations on the year of publication, language, age of participants, and types of interventions.

### 2.2. Identification and selection of COSs

COMET is a repository for international COS literature (<http://www.comet-initiative.org/studies/search>).<sup>8</sup> A comprehensive search of COMET database was conducted on December 20, 2019. Two reviewers (Y.G. and J.Y.S.) independently exported all registered COS study records from the COMET database. Full-text publications that met the inclusion criteria were then filtered. A regularly updated systematic review of COSs and the search result of other published articles were screened to identify OG COSs as a supplement (selected by Y.G. and J.Y.S. independently).<sup>8,9</sup> Reference lists from eligible COSs and related reviews were reviewed for additional eligible studies. Any disagreement of the final list of COSs was solved by a COS/methodology expert (J.H.Z.). Eligible studies published in any language were included.

### 2.3. Data extraction and management

All authors involved in this study performed a pilot study on a random sample of five included COSs to ensure the agreement among interpretation of data items, prior to the actual study. One reviewer (J.Y.S., M.L.Y., or Z.W.S.) extracted data from included COSs using a data extraction sheet from included COSs, and a second reviewer (Y.G., Y.M.C., or M.M.N) verified extracted data. Any disagreements were adjudicated by a third reviewer (J.H.T.). Extracted data included: number of authors, countries which co-authors came from, whether the COS protocol was registered, journals where the studies were published and impact factors, number of databases searched (Chinese, English, or both), funding source (industry, government, unfunded, or not reported), and impact factors of journals in which COSs included were retrieved by searching 2018 Journal Citation Reports (Thomson Reuters, 2018) through Web of Science. Details on data extraction table is shown in Supplement 1A.

### 2.4. Standards assessment

#### 2.4.1. Reporting quality assessment

COS-STAR was developed using a recommended approach for guiding minimal COS study reporting. The reporting standards of included COSs were explored using COS-STAR which contains 18 items (including 25 checklist criteria) spread over six domains including; title/abstract, introduction, methods, results, discussion, and other information. Responses for each item included; “Yes - item fully compliant, partial - item not fully compliant, or No - item not compliant. Reporting standards assessment using COS-STAR tool was independently conducted by two independent reviewers (J.Y.S. and M.L.Y.) (Supplement 1B), conflicts were adjudicated by a third reviewer (J.H.Z.).

#### 2.4.2. Methodological quality assessment

Methodological standards of included COSs was evaluated using COS-STAD, which contains 12 checklist criteria spread over three domains namely; scope specification, stakeholders involved, and consensus process (Supplement 1C). The three domains are considered the minimum design recommendations for development of all COSs. To indicate the degree of compliance, each checklist item was defined as: Yes - for fully addressed; Partial - for partially addressed; and No - for not addressed. Assessments were compared, and three authors (J.Y.S. M.L.Y., and Y.G.) deliberated on how the process should be applied. Methodological standards assessment using the COS-STAD checklist was independently conducted by two reviewers (J.Y.S and Y.G). Conflicts were adjudicated by a third reviewer (J.H.Z.).

### 2.5. Statistical analysis

Continuous data was expressed as median and range, whereas categorical data was summarized as frequency and percentages. This article is not intended to criticize the quality of the published COSs. Therefore, the study only reports compliance rates of items included in literature, but it does not score the standards of individual studies. The frequency of “Yes” response was summarized for individual items of the COS-STAD and COS-STAR for all included COSs. Chi-square test and odds ratio (OR) was calculated with 95% confidence intervals (95% CIs), and *P*-value for each item compared between group one (COSs with statistician or epidemiologist, funding, and with prior protocol) and group two (without statistician or epidemiologist, non-funding, and without protocol) was obtained. Moreover, Chi-square test was used to determine compliance for each item of COS-STAD between OG COSs and cancer COSs (the result of cancer COSs assessment was extracted from a previous

**Table 1**  
Characteristics of the included COSs.

Characteristic	Frequency	Proportion (%)
Country of correspondence author		
Developed country	40	90.9
Developing country	4	9.1
Journal impact factor: median (IQR)	5.079 (2.103, 5.357)	
Number of authors		
1 to 3 authors	6	13.6
4 to 6 authors	15	34.1
7 to 10 authors	13	30
11 to 20 authors	7	15.9
21 or more authors	3	6.8
With a priori protocol	25	56.8
With statistician, epidemiologist	13	29.5
Funding sources		
Industry	2	4.5
Non-industry	19	43.2
Industry + Non-industry	3	68.2
No funding	9	20.5
Not reported	11	25
Scoping process		
Conducted systematic/literature review	34	77.2
Reported search strategy	29	65.9
Searched 1 to 3 databases	12	27.3
Searched 4 to more databases	15	34.1
Used the interview method	3	6.8
Consensus process		
Conducted 2 rounds of Delphi survey	11	25
Conducted 3 rounds of Delphi survey	9	20.5
Used the Consensus meeting method	18	40.9
Conducted both scoping process and consensus process	16	36.4

IQR: interquartile range.

study conducted by Elizabeth G).<sup>19</sup> The mean compliance items and standard deviation (SD) were calculated for each of the COS-STAR and COS-STAD checklist items. Mean difference (MD) and 95% CI were calculated for each item to compare the overall compliance items between groups. MD value represented the difference in the mean total compliance items between group one and group two. Analyses were conducted using STATA (13.0; Stata Corporation, College Station, Texas, USA), and statistical significance was set at  $P < 0.05$ . The complete compliance rate of the methodological quality was calculated with the acquired number. Spearman correlation coefficient ( $r$ ) was estimated to determine the linear association between average citations per year (dates from Web of Science) and the total complete compliance rates of COS-STAD and COS-STAR for each OG COS. Correlation analyses were conducted using IBM SPSS Statistics v. 24.0 (IBM Corp., Armonk, NY, USA), and statistical significance was set at  $P < 0.05$ . We created bubble plots with Microsoft Excel 2016 (Microsoft Corp, Redmond, WA; [www.microsoft.com](http://www.microsoft.com)) to present the compliance rate. Considering bubble plots according to the compliance rate, the bubble size represented the number of compliance rate, the x-axis represented the COS-STAR or COS-STAD items, and the y-axis represented the compliance rate of each item.

### 3. Results

#### 3.1. Characteristics of included COSs

The first COS was published in 2000; most COSs were published from 2018–2019 (Supplement 2A). Forty-four COSs focused on 26 topics mainly with most reporting on pelvic organ prolapse (9%) and maternity care (9%) (Supplement 2B). Corresponding authors of included studies were from nine countries, with the UK (38.6%), USA (13.6%), Canada (11.4%), and Australia (9%) having the highest number (Supplement 2C). The main characteristics of included COSs are presented in Table 1 showed the main characteristics of

the included COSs, including e.g., correspondence author, impact journal factor, prior protocol, and COS development process.

#### 3.2. Results of reporting quality

Compliance rates of 8 standards were  $> 85\%$ , mainly focusing on introduction and discussion sections; among them, the compliance rates of items 2b, 3a, and 3b were 100.0%. However, compliance rates of the 11 standards on the COS-STAR assessment were less than 55%, including items 1a, 5, 7, 8, 9a, 9b, 10, 11, 12, 13a, 13b (Fig. 1). Further more, total COS-STAR compliance standards of OG COSs with the priori protocol were significantly higher compared those for OG COSs without protocol (Table 2). The compliance for each of the 25 criteria was analyzed and the findings showed significant differences between COSs with priori protocol and those without a priori protocol based on the items 1a, 4, 6a, 6b, and 17.

#### 3.3. Results of methodological quality

Compliance rates of four criteria related to “scope specification” were above 95%. However, the compliance rates of eight items related to “stakeholders involved” and “consensus process” were less than 60%. (Fig. 2). The findings showed that there was no significant difference in total COS-STAR compliance standards of OG COSs between group one and group two. Studies with a priori protocol or registered details showed higher compliance rates of the items 9a, 9b, and 10 compared with COSs without prior protocol or registered details. COSs which involved statistician, or epidemiologist authors showed higher compliance rates on item 6 compared with those COSs not involve statistician, or epidemiologist author. Moreover, funding COSs and non-funding COSs related to the item 7 showed significant differences (Table 3).

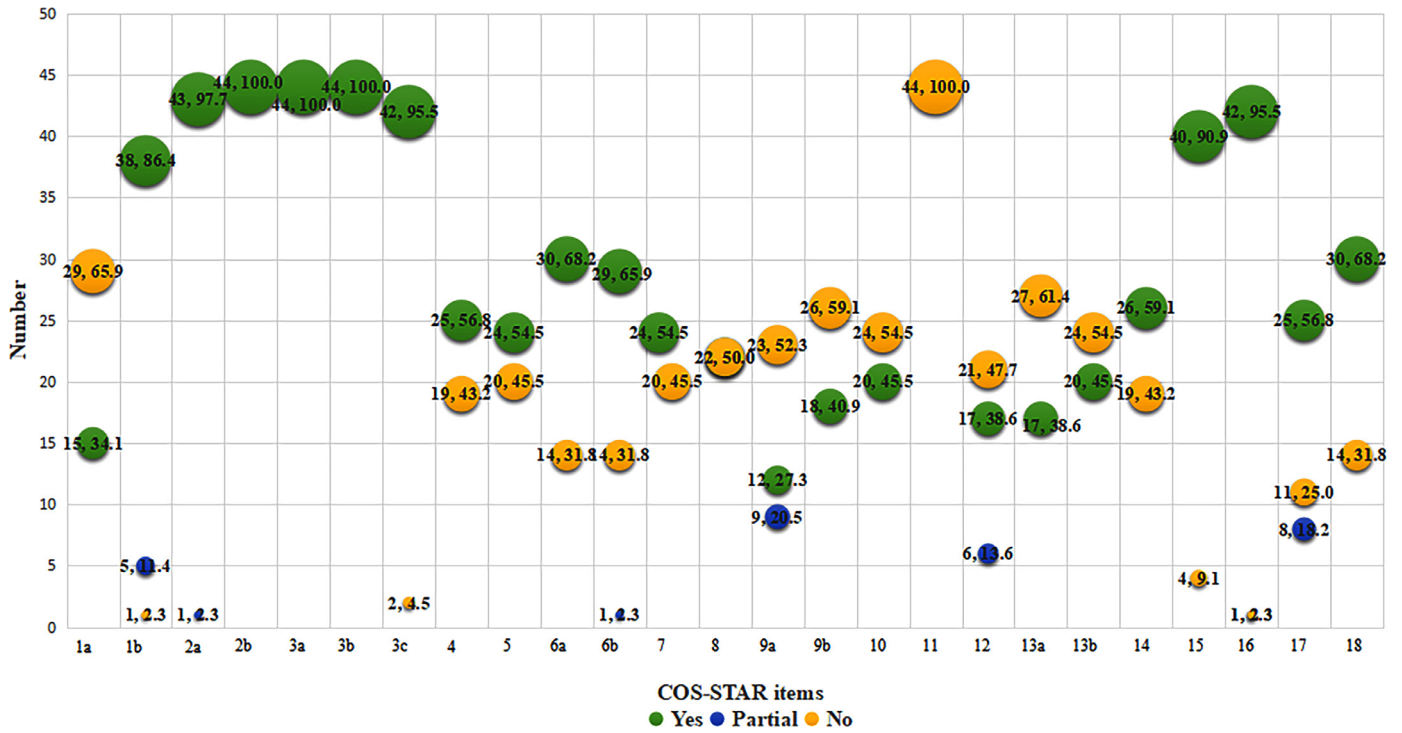


Fig. 1. Bubble plot of compliance proportions of standards on the COS-STAR assessment. bubble size: numbers/proportions of SRs; bubble colors:degree of compliance (yes, partial, and no). the x-axis represented each COS-STAR item, the y-axis represented the number and compliance rate of each item.

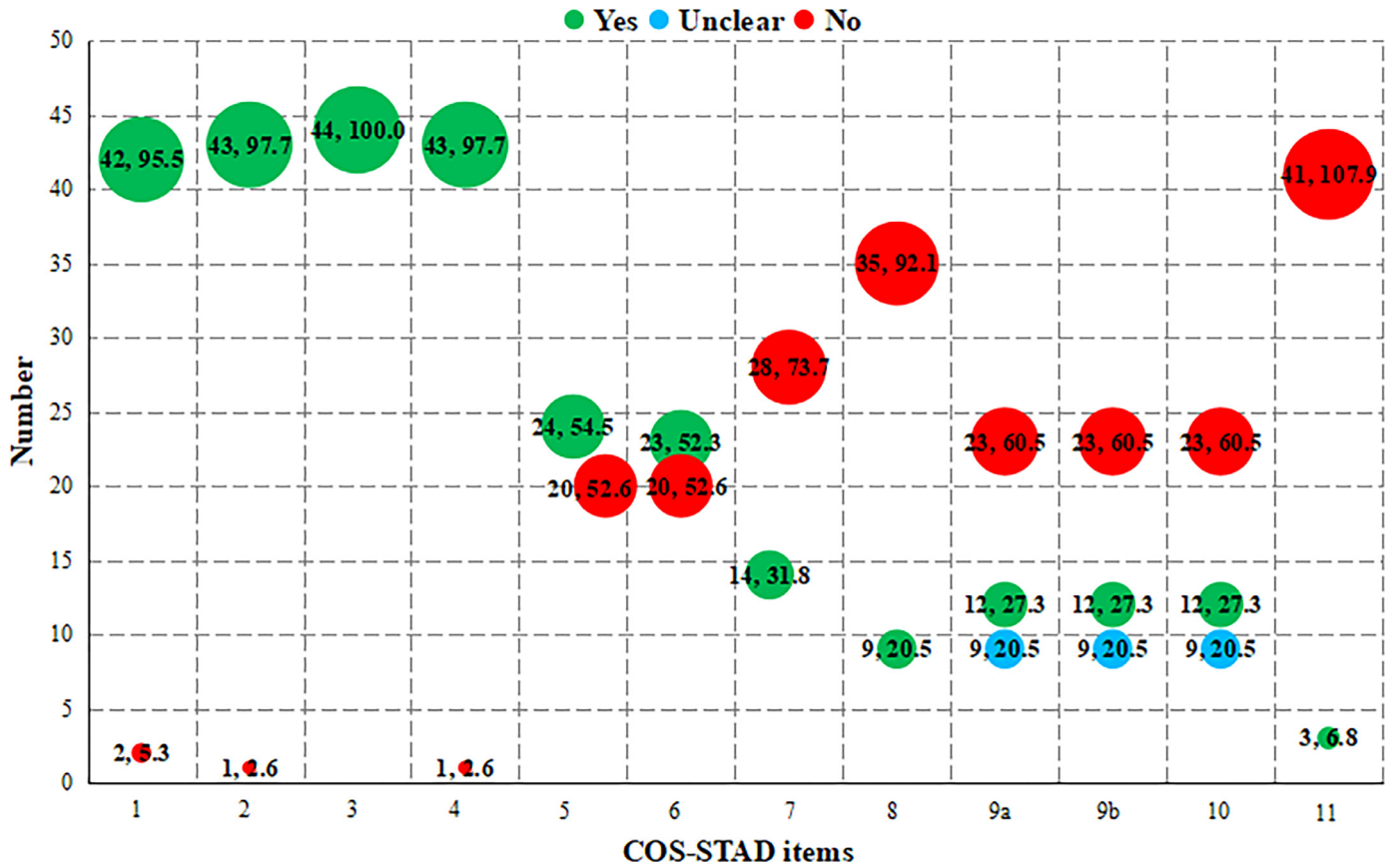


Fig. 2. Bubble plot of compliance proportions of standards on the COS-STAD assessment. bubble size: numbers/proportions of SRs; bubble colors:degree of compliance (yes, partial, and no). the x-axis represented each COS-STAD item, the y-axis represented the number and compliance rate of each item.

**Table 2**  
Stratified analyzes of reporting quality assessment in compliance of COS-STAR standards.

Section	Items/standards		With PP (n = 25)	Without PP (n = 19)	OR(95% CI)	P-value	With SE (n = 14)	Without SE (n = 30)	OR(95% CI)	P-value	Funding (n = 23)	No funding (n = 21)	OR (95% CI)	P-value
TITLE/ABSTRACT	Title	1a	13(52.0)	2(10.5)	4.94(1.26,19.32)	0.022	4(28.5)	11(36.6)	0.77(0.3,2.02)	0.608	8(34.8)	7(33.3)	1.04(0.46,2.38)	0.919
	Abstract	1b	25(100.0)	18(94.7)	1.06(0.93,1.22)	0.399	12(85.7)	26(86.6)	0.98(0.76,1.27)	0.933	19(82.6)	19(90.5)	0.91(0.72,1.15)	0.445
INTRODUCTION	Background and objectives	2a	25(100.0)	19(100.0)	-	-	13(92.8)	30(100)	0.91(0.76,1.08)	0.317	23(100.0)	20(95.2)	1.05(0.93,1.19)	0.445
	Scope	2b	25(100.0)	19(100.0)	-	-	14(100)	30(100)	-	-	23(100.0)	21(100.0)	-	-
		3a	25(100.0)	19(100.0)	-	-	14(100)	30(100)	-	-	23(100.0)	21(100.0)	-	-
		3b	25(100.0)	19(100.0)	-	-	14(100)	30(100)	-	-	23(100.0)	21(100.0)	-	-
		3c	25(100.0)	18(94.7)	1.06(0.93,1.22)	0.399	14(100)	29(96.6)	1.01(0.89,1.14)	0.803	21(91.3)	21(100.0)	0.92(0.79,1.07)	0.257
METHODS	Protocol/Registry entry	4	25(100.0)	0(0.0)	39.23(2.54,606.03)	0.009	6(42.9)	19(63.3)	0.68(0.34,1.31)	0.249	12(52.2)	13(61.9)	0.84(0.50,1.41)	0.516
	Participants	5	13(52.0)	11(57.9)	0.90(0.53,1.54)	0.695	9(64.2)	15(50)	1.28(0.75,2.18)	0.352	15(65.2)	9(42.9)	1.52(0.86,2.71)	0.154
	Information sources	6a	24(96.0)	8(42.1)	2.28(1.33, 3.89)	0.002	8(57.1)	22(73.3)	0.77(0.47,1.28)	0.33	15(65.2)	15(71.4)	0.91(0.61,1.37)	0.658
		6b	21(84.0)	8(42.1)	2.0(1.15,3.47)	0.015	7(50)	22(73.3)	0.68(0.38,1.2)	0.185	15(65.2)	14(66.7)	0.98(0.64,1.50)	0.919
	Consensus process	7	12(48.0)	12(63.2)	0.76(0.45,1.30)	0.313	10(71.4)	14(46.6)	1.53(0.92,2.53)	0.099	14(60.9)	10(47.6)	1.28(0.73,2.23)	0.386
	Outcome scoring	8	12(48.0)	10(52.6)	0.91(0.51,1.65)	0.760	8(57.1)	14(46.6)	1.22(0.67,2.21)	0.504	12(52.2)	10(47.6)	1.10(0.60,1.99)	0.764
	Consensus definition	9a	12(48.0)	8(42.1)	1.14(0.59,2.22)	0.700	3(21.4)	9(30)	0.71(0.22,2.23)	0.564	7(30.4)	5(23.8)	1.28(0.48,3.42)	0.625
		9b	11(44.0)	7(36.8)	1.19(0.57,2.49)	0.636	6(42.8)	12(40)	1.07(0.5,2.26)	0.856	11(47.8)	7(33.3)	1.44(0.68,3.01)	0.339
	Ethics and consent	10	13(52.0)	7(36.8)	1.41(0.70,2.84)	0.334	7(50)	13(43.3)	1.15(0.59,2.24)	0.673	13(56.5)	7(33.3)	1.70(0.84,3.43)	0.141
RESULTS	Protocol deviations	11	0(0.0)	0(0.0)	-	-	0(0.0)	0(0.0)	--	--	0(0.0)	0(0.0)	--	--
	Participants	12	11(44.0)	6(31.6)	1.39(0.62,3.09)	0.414	6(42.8)	11(36.6)	1.16(0.54,2.51)	0.69	11(47.8)	6(28.6)	1.67(0.75,3.72)	0.207
	Outcomes	13a	10(40.0)	6(31.6)	1.27(0.56,2.87)	0.571	6(42.8)	10(33.3)	1.28(0.58,2.82)	0.532	11(47.8)	5(23.8)	2.01(0.84,4.82)	0.119
		13b	12(48.0)	7(36.8)	1.30(0.64,2.67)	0.469	7(50)	12(40)	1.25(0.63,2.47)	0.522	12(52.2)	7(33.3)	1.57(0.76,3.22)	0.223
	COS	14	12(48.0)	13(68.4)	0.70(0.42,1.17)	0.173	10(71.4)	15(50)	1.42(0.87,2.32)	0.152	16(69.6)	9(42.9)	1.62(0.92,2.85)	0.092
DISCUSSION	Limitations	15	24(96.0)	16(84.2)	1.14(0.92,1.41)	0.222	13(92.8)	27(90)	1.03(0.85,1.24)	0.745	19(82.6)	21(100.0)	0.83(0.68,1.02)	0.074
	Conclusions	16	25(100.0)	17(89.5)	1.12(0.94,1.33)	0.199	13(92.8)	29(96.6)	0.96(0.81,1.12)	0.622	21(91.3)	21(100.0)	0.92(0.79,1.07)	0.257
OTHER INFORMATION	Funding	17	18(72.0)	6(31.6)	2.28(1.12,4.61)	0.022	9(64.2)	15(50)	1.28(0.75,2.18)	0.352	17(73.9)	7(33.3)	2.21(1.16,4.26)	0.017
	Conflicts of interest	18	19(76.0)	11(57.9)	1.31(0.84,2.04)	0.228	6(42.8)	24(80)	0.53(0.28,1)	0.052	16(69.6)	14(66.7)	1.04(0.67,1.57)	0.837
Total compliance (Mean, SD)			17.32(5.61)	13.47(4.57)	3.85(0.84,6.86)	0.012	15.92(4.82)	15.55(5.81)	0.37(-3.33,4.07)	0.845	16.17(5.99)	16.17(5.99)	1.13(-2.13,4.37)	0.5

**Table 3**  
Stratified analyzes of methodological quality assessment in compliance of COS-STAD standards.

Section	Items/standards	With a PP (n = 25)	Without PP (n = 19)	OR(95% CI)	P-value	With SE (n = 14)	Without SE (n = 30)	OR(95% CI)	P-value	Funding (n = 23)	No funding (n = 21)	OR(95% CI)	P-value
Scope specification	1.The research or practice setting(s) in which the COS is to be applied	23(92.0)	19(100)	0.92(0.8,1.07)	0.301	12(92.3)	30(96.8)	0.95(0.8,1.13)	0.585	22(91.6)	20(100)	0.92(0.79,1.06)	0.278
	2.The health condition(s) covered by the COS	25(100.0)	18(94.7)	1.06(0.92,1.21)	0.399	12(92.3)	31(100.0)	0.9(0.75,1.09)	0.305	24(100)	19(95)	1.05(0.93,1.19)	0.421
	3.The population(s) covered by the COS	25(100.0)	19(100.0)	-	-	13(100)	31(100.0)	-	-	24(100)	20(100)	-	-
	4.The intervention(s) covered by the COS	25(100.0)	18(94.7)	1.06(0.92,1.21)	0.399	12(92.3)	31(100.0)	0.9(0.75,1.09)	0.305	23(95.8)	20(100)	0.96(0.85,1.08)	0.535
Stakeholders involved	5.Those who will use the COS in research	14(56.0)	10(52.6)	1.06(0.61,1.84)	0.825	8(61.5)	16(51.6)	1.19(0.68,2.06)	0.53	15(62.5)	9(45)	1.38(0.78,2.46)	0.263
	6.Health care professionals with experience of patients with the condition	12(48.0)	11(57.8)	0.82(0.47,1.45)	0.512	10(76.9)	13(41.9)	1.83(1.1,3.05)	0.02	16(66.6)	7(35)	1.9(0.98,3.68)	0.052
	7.Patients with the condition or their representatives	10(40.0)	4(21.0)	1.9(0.7,5.13)	0.206	4(30.7)	10(32.3)	0.95(0.36,2.49)	0.923	11(45.8)	3(15)	3.35(1.08,10.38)	0.036
Consensus process	8.Initial list of outcomes considered both health care professionals' and patients' views	7(28.0)	2(10.5)	2.66(0.62,11.38)	0.187	3(23)	6(19.4)	1.19(0.35,4.06)	0.778	7(29.1)	2(10)	2.91(0.68,12.49)	0.149
	9a.A scoring process was described a priori	11(44.0)	1(5.2)	8.36(1.18,59.24)	0.034	3(23)	9(29.0)	0.79(0.25,2.47)	0.692	7(29.1)	5(25)	1.16(0.43,3.11)	0.758
	9b.A consensus definition was described a priori	11(44.0)	1(5.2)	8.36(1.18,59.24)	0.034	3(23)	9(29.0)	0.79(0.25,2.47)	0.692	7(29.1)	5(25)	1.16(0.43,3.11)	0.758
	10.Criteria for including/dropping/adding outcomes were described a priori	11(44.0)	1(5.2)	8.36(1.18,59.24)	0.034	3(23)	9(29.0)	0.79(0.25,2.47)	0.692	7(29.1)	5(25)	1.16(0.43,3.11)	0.758
	11.Care was taken to avoid ambiguity of language used in the list of outcomes	2(8.0)	1(5.2)	1.52(0.14,15.54)	0.724	1(7.6)	2(6.5)	1.19(0.11,12.02)	0.881	1(4.1)	2(10)	0.41(0.04,4.26)	0.461
Total compliance (Mean, SD)		7.2(3.25)	6(2.08)	1.2(-0.38,2.78)	0.137	7.46(2.63)	6.35(2.90)	1.11(-0.83,3.05)	0.263	7.17(2.82)	6.1(2.83)	1.07(-0.63,2.77)	0.219

## 4. Discussion

### 4.1. Summary of characteristics of included COSs

Overall, the reporting of statistical analysis methods was suboptimal, and the reporting quality and methodological quality were low. Compliance rates were insufficient for most of the items of COS-STAD and COS-STAR checklists.

A total of 44 OG COSs were identified from 26 research topics. Identification of an inclusive list of outcomes from the existing literature is important in development of a core outcome set.<sup>7</sup> Main data sources included systematic reviews of published studies, reviews of published qualitative work, and interviews with key stakeholders.<sup>16,26</sup> However, only three COSs interviewed key stakeholders to understand their views of important outcomes, whereas 16% of COSs did not conduct systematic/literature review to identify existing knowledge on outcomes. Comprehensiveness of the results of a systematic review is highly dependent on the results of the underlying data, therefore, it is important to carry out a thorough verification of reported results. Notably, only 34.1% of COSs searched more than three databases. A systematic review of the results aggregates the opinions of previous researchers. Therefore, it is important to subsequently strike a consensus with the wider community of stakeholders on outcomes to be included in a COS. In the present study, only twenty COSs (45.5%) conducted two or more rounds of Delphi survey. Notably, 40.9% of COSs used the consensus meeting method whereas 36.4% of studies used both systematic review and consensus methods.

### 4.2. Reporting standards of included COSs

For the COS-STAR checklist, only 34.1% of studies reported the title that the paper reports the development of a COS, future researcher should identify their report as a COS development study in the title. COS developers should provide a clear and transparent report of the methods they used, but the compliance rates of 9 criteria in the methods section ranged from 27.3%–68.2%, which is not conducive to improving the transparency of COS, nor is it conducive for COS users to judge whether the development process of COS is rigorous. No study met the item 11 criterion (describe any changes from the protocol), therefore, further analysis should explore whether COS not deviating from the protocol should report this item. This would improve the applicability and suitability of this criterion. Characteristics of participants involved were not presented in 61.4% of included studies, and all outcomes considered at the start of the consensus process and descriptions of any other outcomes introduced/dropped during the consensus process were absent in more than half of the studies, which makes it impossible for readers to assess whether the decisions were made in an unbiased way. Recommended outcomes in 43.2% of COSs were unclear or ambiguous, which may limit their application. Only twenty-four (54.5%) COSs met all the other two information criteria (sources of funding and conflicts of interest), future COS developer should clearly report funding information and describe any conflicts of interest within the study team. Moreover, total COS-STAR compliance criteria of OG COSs was significantly positively correlated with prior protocol, indicating that COS developer may improve the quality of COS by publishing protocol or registry details in advance.

### 4.3. Methodological standards of included COSs

Four standards for scope specification for the COS-STAD checklist were well met. All stakeholders involved in the COS development should be reported to ascertain whether the COS fully reflects the views of important outcomes for the target population in

the forthcoming research, which will improve utilization and promotion of the COS.<sup>23</sup> However, only fourteen COSs (31.8%) met all the three standards for the domain of stakeholders involved. Judging the standards of COS-STAD in the consensus process that was set in advance is challenging, analysis was performed to determine whether the studies that met this standard relied on registration information and published protocol.<sup>19</sup> But only 27.3% of COSs priori described a consensus protocol, and some registration information was inadequately reported, making it challenging to judge whether some standards were set prior. The registration platform can improve the standards for COS registration, thus promoting integrity and transparency of COS developed. Analysis showed that 92.1% of COSs did not report whether the patient representatives received a glossary of terms before completing the survey, which may cause ambiguity in language. Further more, Spearman's correlation analysis showed a statistically significant positive correlation between average citations per year and compliance rates of COS-STAD for OG COSs, which indicates that improving the quality of COS may contribute to the promotion and application of COS.

### 4.4. Comparisons between cancer COSs and OG COSs

A previous study conducted by Elizabeth G on the minimum standard assessments of cancer COSs reported that compliance rates for both four standards in scope domain were above 95%, which was consistent with that of OG COSs.<sup>19</sup> Cancer COSs showed similar low compliance rate for all five standards in consensus process standards, but had lower compliance rates for items 9a and 10 compared with OG COSs, which indicates that future researchers should pay more attention to these items (Supplement 3). Overall, OG COSs had a similar low compliance rate for COS-STAD items compared with cancer COSs, further research needs to explore whether this problem exists in COS for other fields.

### 4.5. Suggestions for future work

This study indicated that both the reporting and methodological standards of OG COSs were of low quality according to COS-STAD and COS-STAR checklists, and many items had low compliance rates. Future OG COS makers should address these identified shortcomings, and conducted COS according to the reporting and methodological guidelines. The results of this study also found that writing a protocol is of great help in improving the quality of OG COS, future OG COS developer need to increase the transparency and robustness of the methods by drafting of high-quality protocols according to the Core Outcome Set-STANDARDISED Protocol Items (COS-STAP) statement.<sup>3</sup> Moreover, future researcher should fully consider the need for COS updating, including updating the SR and consensus progress.

### 4.6. Strengths and limitations

This work had a few limitations. First, some studies assessed were published before development of COS-STAR and COS-STAD standards, therefore they could not have been informed by the development standards. However, the current study focused on investigating the current issues of COSs and explored ways of improving the COS standards instead of criticizing quality of existing COSs, thus circumventing this limitation. Second, some included COSs were published before the COMET initiative was set up, thus they were not registered. Therefore, the definition of COSs developed with the protocol was not limited to this website, but was also based on other registration platforms, and protocols published in peer-reviewed journals that met the set criteria. Third, the same COS-STAD criteria was used to compare the standards between OG

COSs with cancer COSs based on previous research. Notably, variation in standards for item compliance may exist in the assessment progress between different reviewers. COS-STAD and COS-STAR did not provide a standard to score the overall quality of COSs, therefore, analysis in the current study was limited to the compliance rate. Further studies should be conducted to explore the appropriate statistical methods for estimating an average value of the quality scores of multiple COSs.

## 5. Conclusion

Compliance of COS-STAD and COS-STAR criteria in OG COSs are not optimistic. Therefore, reporting and methodological standards of OG COSs should be further improved, mainly design of important stakeholders involved and consensus process; and reporting standards in methods, results, and other information specifications. OG COS developed with the prior protocol is reliable for improving the reporting standards.

## Acknowledgements

We would like to acknowledge: (1) the COMET Initiative team for methodological support; (2) Methodological guidance: Elizabeth Gargon, University of Liverpool, United Kingdom; (3) James M N Duffy, University of Oxford, Oxford, UK.

## Author contributions

JS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. YG: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. SW: Data curation, Investigation, Resources. MN: Data curation, Investigation, Resources. MY: Formal analysis, Data curation, Resources, Software, Validation. YC: Investigation, Resources, Data curation. ZS: Data curation, Resources. HF: Language revised. JT: Conceptualization, Formal analysis, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. JZ: Project administration, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. All authors have read and approved the manuscript.

## Conflict of interests

The authors declare that they have no competing interests.

## Funding

None.

## Ethical statement

Not applicable.

## Data availability

Not applicable.

## Supplementary materials

Supplementary material associated with this article can be found in the online version, at doi:[10.1016/j.imr.2021.100776](https://doi.org/10.1016/j.imr.2021.100776).

## References

- Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: version 1.0. *Trials*. 2017;18(Suppl 3):280.
- Li X, Ya Z, Chen Y, Yang K, Zhang Z. The reporting characteristics and methodological quality of cochrane reviews about health policy research. *Health policy*. 2015;119(4):503–510.
- Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core outcome Set-STandardised protocol items: the COS-STAP statement. *Trials*. 2019;20(1):116.
- Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*. 2010;340:c365.
- Gorst SL, Prinsen CA, 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;26(125):64–75.
- Davis K, Gorst SL, Harman N, Smith V, Gargon E, Altman DG, 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).
- Kirkham JJ, Altman DG, Chan AW, Gamble C, Dwan KM, Williamson PR. Outcome reporting bias in trials: a methodological approach for assessment and adjustment in systematic reviews. *BMJ*. 2018;362:k3802.
- Gargon E, Gorst SL, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 5th annual update to a systematic review of core outcome sets for research. *PLoS One*. 2019;14(12).
- Liu M, Gao Y, Yuan Y, Wu J, Zhang J, Tian J, et al. Inconsistency and low transparency were found between core outcome set protocol and full text publication: a comparative study. *J Clin Epidemiol*. 2020;131:59–69 Nov 21.
- Wuytack F, Smith V, Clarke M, Williamson P, Gargon E. Towards core outcome set (COS) development: a follow-up descriptive survey of outcomes in cochrane reviews. *Syst Rev*. 2015;4:73.
- Gargon E, Gurung B, Medley N, Altman DG, Blazeby JM, Clarke M, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. *PLoS One*. 2014;9(6):e99111.
- Gorst SL, Gargon E, Clarke M, Smith V, Williamson PR. Choosing important health outcomes for comparative effectiveness research: an updated review and identification of gaps. *PLoS One*. 2016;11(12).
- Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing important health outcomes for comparative effectiveness research: an updated review and user survey. *PLoS One*. 2016;11(1).
- Gorst SL, Young B, Williamson PR, Wilding JPH, Harman NL. Incorporating patients' perspectives into the initial stages of core outcome set development: a rapid review of qualitative studies of type 2 diabetes. *BMJ Open Diabetes Res Care*. 2019;7(1).
- Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core outcome Set-STAndards for development: the COS-STAD recommendations. *PLoS Med*. 2017;14(11).
- Biggane AM, Brading L, Ravaud P, Young B, Williamson PR. Survey indicated that core outcome set development is increasingly including patients, being conducted internationally and using delphi surveys. *Trials*. 2018;19(1):113.
- Kirkham JJ, Bracken M, Hind L, Pennington K, Clarke M, Williamson PR. Industry funding was associated with increased use of core outcome sets. *J Clin Epidemiol*. 2019;115:90–97.
- Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, Devane D, et al. Core outcome set-STAndards for reporting: the COS-STAR statement. *PLoS Med*. 2016;13(10).
- Gargon E, Williamson PR, Blazeby JM, Kirkham JJ. Improvement was needed in the standards of development for cancer core outcome sets. *J Clin Epidemiol*. 2019;112:36–44.
- De la Fuente-Solana EI, Suleiman-Martos N, Pradas-Hernández L, Gomez-Urquiza JL, Cañadas-De la Fuente GA, Albendín-García L. Prevalence, related factors, and levels of burnout syndrome among nurses working in gynecology and obstetrics services: a systematic review and meta-analysis. *Int J Environ Res Public Health*. 2019;16(14).
- Al Wattar BH, Tamilselvan K, Khan R, Kelso A, Sinha A, Pirie AM, et al. Development of a core outcome set for epilepsy in pregnancy (E-CORE): a national multi-stakeholder modified Delphi consensus study. *BJOG*. 2017;124(4):661–667.
- Pergialiotis V, Durnea C, Elifturi A, Duffy JMN, Doumouchtsis SK. Do we need a core outcome set for childbirth perineal trauma research? A systematic review of outcome reporting in randomised trials evaluating the management of childbirth trauma. *BJOG*. 2018;125(12):1522–1531.
- Gargon E, Williamson PR, Young B. Improving core outcome set development: qualitative interviews with developers provided pointers to inform guidance. *J Clin Epidemiol*. 2017;86:140–152.
- Tunis SR, Clarke M, Gorst SL, Gargon E, Blazeby JM, Altman DG, et al. Improving the relevance and consistency of outcomes in comparative effectiveness research. *J Comp Eff Res*. 2016;5(2):193–205.
- Duffy JMN, Thompson T, Hinton L, Salinas M, McManus RJ, Ziebland S, et al. What outcomes should researchers select, collect and report in pre-eclampsia research? A qualitative study exploring the views of women with lived experience of pre-eclampsia. *BJOG*. 2019;126(5):637–646.
- Gorst SL, Altman DG, Blazeby JM, Clarke M, Gargon E, Tunis S, et al. Proceedings of the 5th meeting of the core outcome measures in effectiveness trials (COMET) initiative. *Trials*. 2015;16(Suppl 3):A1–11 Suppl 3.