

BMJ Open Mixed-methods study to develop extensions to the SPIRIT and CONSORT statements for factorial randomised trials: the Reporting Factorial Trials (RAFT) study

Sophie S Hall ¹, Edmund Juszczak ¹, Megan Birchenall,¹ Diana Elbourne,² Elaine Beller,³ An-Wen Chan,⁴ Paul Little,⁵ Alan A Montgomery,¹ Brennan C Kahan⁶

To cite: Hall SS, Juszczak E, Birchenall M, *et al*. Mixed-methods study to develop extensions to the SPIRIT and CONSORT statements for factorial randomised trials: the Reporting Factorial Trials (RAFT) study. *BMJ Open* 2025;**15**:e082917. doi:10.1136/bmjopen-2023-082917

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2023-082917>).

Received 06 December 2023
Accepted 21 January 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to

Dr Sophie S Hall;
sophie.hall@nottingham.ac.uk

ABSTRACT

Background Extensions to Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) and Consolidated Standards of Reporting Trials (CONSORT) reporting recommendations specifically for factorial trials have been developed by the Reporting Factorial Trials (RAFT) study group. This article describes the processes and methods used to develop the extensions.

Objective To develop SPIRIT and CONSORT extensions for factorial trials.

Design and participants A four-phase, consensus-based approach was used: phase 1: scoping review, phase 2: Delphi survey (n=104 respondents in round 1), phase 3: consensus meeting (n=15 members) and phase 4: checklist finalisation.

Results In phase 1, the scoping review identified 31 reporting recommendations, which formed a long list of 50 concepts (19 applied to the SPIRIT extension and 31 applied to the CONSORT extension) to include in the guideline development. In phase 2, a three-round Delphi survey resulted in two new concepts being added and ended with 49 concepts (19 applied to SPIRIT and 30 applied to CONSORT) reaching consensus to remain, with only three concepts meeting the exclusion criteria. In phase 3, the concepts were further refined and translated into specific extension item wording, through an extensive review process conducted by the core RAFT team and leading trial experts, who attended a 2-day hybrid meeting. The resulting 9 SPIRIT items and 17 CONSORT items were further evaluated and developed through an iterative process in phase 4, to promote user acceptance and uptake.

Conclusion Uptake of the CONSORT and SPIRIT extensions will improve the conduct of factorial trials, as well as understanding and interpretation of such trials. By reporting on how these extensions were developed, we promote transparency of this process and share learning experiences to develop best practice when developing reporting guidelines.

BACKGROUND

The gold standard of methodological reporting for clinical trials involves

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Mixed-methods approach to develop Standard Protocol Items: Recommendations for Interventional Trials and Consolidated Standards of Reporting Trials extensions, encompassing the international literature in the field and expert opinion.
- ⇒ Methods adhered to Enhancing the QUALity and Transparency Of health Research principles for guideline development.
- ⇒ International experts in the field of factorial trials and guideline reporting participated in the study.
- ⇒ Alternative study methods may have produced a different set of guidelines.

following the 33-item Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT)¹ guideline and the 25-item Consolidated Standards of Reporting Trials (CONSORT)^{2,3} guideline. These are recommendations for the minimum content of reporting trial protocols and trial results, respectively. As well as providing a checklist for best practice, the guidelines are accompanied by an Explanation and Elaboration (E&E) document, which provides further details on the items. The guidelines were originally developed to improve research transparency and reduce waste in relation to two-arm parallel group trials. Extensions have since been developed to consider the methodological nuances of specific trial designs including (but not limited to) multi-arm trials, n-of-1 trials, non-inferiority trials and pilot and feasibility trials,^{4–8} enabling existing items to be modified, removed or new items to be added if required.

A factorial trial design involves testing two or more interventions within the same trial.⁹ In the simplest form of a 2 × 2 design, participants are randomised to receive neither

intervention, one or the other, nor both. This enables simultaneous assessment of two interventions without increasing the sample size, assuming that the interventions do not interact.⁹ Alternatively, sometimes the intent is to test for interactions between the interventions.⁹ With the ever-growing need to improve trial efficiency and offer value for money,^{10 11} factorial trials offer a potentially attractive trial design.

The use of a factorial design introduces additional methodological complexities, which are often inadequately understood and poorly reported.^{12–15} A review of 100 2×2 factorial trials published between January 2015 and March 2018 found a number of concerns around their design, analysis and reporting,¹² including (1) providing no clear rationale for the factorial design, (2) a lack of reporting of interactions and (3) using a preliminary test of interaction to choose the analysis method. These limitations in conduct and reporting make it difficult to assess the validity of the trial conclusions, reducing their applicability for patient care and policy recommendations. It is essential that the rationale for using a factorial design and the assumptions made, particularly if and how an assessment of the statistical interaction was planned, are reported transparently. A simple search of the ClinicalTrials.gov registry suggests that >500 new factorial trials are registered each year, and other data suggest that 3% of all trials published in high-impact medical journals employ a factorial design.¹⁶ Therefore, the impact of poor reporting is potentially wide-reaching.

To facilitate improved reporting of factorial trials, the Reporting Factorial Trial (RAFT) study developed SPIRIT and CONSORT extensions^{17 18} using methods recommended for developing reporting guidelines. This article describes the consensus-based development process used to create the final checklists and to share our learning reflections and recommendations for future studies.

METHODS

The study was registered with the Enhancing the QUALity and Transparency Of health Research (EQUATOR) network¹⁹ and followed their methods for developing healthcare reporting guidelines.²⁰ The study protocol was published before the commencement of the research activities on an Open Science Framework.²¹

Project launch

Need for the guidelines

A review of factorial trials conducted by study members highlighted key shortcomings with the reporting of factorial trials, including that the majority of published papers did not provide a rationale for using a factorial trial design, and many neglected to assess for interactions for the primary and secondary outcomes.¹² This provided motivation for developing SPIRIT and CONSORT extensions to improve the reporting of factorial trials. A review of the EQUATOR website and published literature

highlighted that no published guidelines existed to support the reporting of factorial trials.

Identification of research team members

The core RAFT study team was formed ensuring representation of important study-specific expertise relating to knowledge of factorial trials (BCK, AAM, EJ and PL) and the development of SPIRIT (A-WC and EJ) and CONSORT (DE, EB and EJ) guidelines. The project was supported by a Research Fellow (SSH) and a Research Co-ordinator (MB) to enable timely delivery of the project milestones. The core team met online each month, throughout the 24-month duration of the study.

Rationale for simultaneous development of the guidelines

The review of factorial trials¹² highlighted deficits not only in the reporting of results but also in the design and conduct of these trials. With the goal of improving the quality of submissions to research funding bodies and governance agencies, as well as improving knowledge translation and evaluation of clinical trial results, it was decided that simultaneous development of SPIRIT and CONSORT guidelines would prove an efficient use of team resources.

Funding

Funding was awarded from the Medical Research Council (MRC) Methodology Research Panel to support researcher time, purchase the Delphi software, host a consensus meeting and support with dissemination. The funder played no role in the design, analysis or reporting of the study.

Phase 1: scoping review

The objective of the scoping review was to generate a long list of concepts for consideration for inclusion in the CONSORT and SPIRIT extensions, based on reporting or methodological considerations published in the literature. Throughout the development phase, we used the term ‘concepts’ for two reasons: first, checklist ‘items’ may contain multiple reporting concepts and we wished to evaluate each of these concepts separately. Second, the actual checklist ‘item’ wording would be refined through the process reported here, and therefore, we were seeking consensus on whether the general concept was considered important as opposed to specific item wording.

Searches

MEDLINE search terms were developed by a research librarian with input from the team and guided by a recent similar review in this area.¹² Given that this was a scoping review, PROSPERO (Prospective Register of Systematic Reviews) registration was not required. In line with previous reviews searching a similar topic, searches were conducted in MEDLINE from inception to May 2019.¹² The full search strategy is presented in online supplemental material.

Selection

Articles were included in the scoping review if they discussed methodological or reporting issues relevant to factorial trials. One author (EB) screened titles and abstracts to remove articles that were not relevant (eg, those that reported the results of a randomised trial but not methodological issues). Another author (BCK) then further screened abstracts and full texts to assess whether articles were relevant to the SPIRIT/CONSORT extension. Reporting suggestions and methodological issues requiring additional reporting were recorded.

Project phase 2: Delphi survey

The objective of the Delphi survey was to obtain feedback in which concepts could be removed from the long list due to perceived lower importance.

Participants

Sample size guidelines for Delphi studies vary from as low as 10 participants²² to up to 50,²³ with other recommendations suggesting that the group characteristics are taken into consideration, ensuring there are approximately 5–10 participants representing different groups.²⁴ Based on our previous experience,²⁵ we aimed to recruit participants with a broad range of experience in factorial trials and guideline development (including representatives from funding bodies, journal editors and patient and public involvement (PPI)), with a target sample of n=100.

Recruitment strategy

We implemented two recruitment strategies concurrently. The first strategy involved sending personalised invitation letters to methodologists and trialists identified through four key sources: (1) *the RAFT team's contacts*; (2) *relevant publications* (lead authors and methodological coauthors who had published either factorial trials, methodological work related to factorial trials and/or reporting guidelines); (3) *relevant networks* (eg, MRC-NIHR Trials Methodology Research Partnership, network of UK Clinical Research Collaboration registered Clinical Trials Units) and (4) *representatives from funding bodies* (eg, National Institute of Health Research (NIHR)) and journal editors with experience in CONSORT or SPIRIT statements. The second strategy involved wider sharing of the Delphi survey link through the RAFT study website and social media accounts associated with the author's institutions, requesting individuals with experience and interest in factorial trials to participate in a Delphi survey. The opportunity to be entered into a prize draw to receive one of two £100 (or Great British Pound equivalent) online shopping vouchers was offered as an incentive to sign up. The prize draw was only available to participants who completed all three rounds.

Delphi survey process

The Delphi survey was designed and hosted online using DelphiManager software.²⁶ Participants were required to read the participant information sheet, complete a brief

demographic survey, specify if they wish to be entered in the prize draw and complete the consent items.

Three Delphi rounds were conducted to enable participants to suggest new items in the first round and still ensure that all items were rated at least twice. In each round, participants were requested to rate the importance of including each concept in the final set of guidelines for the SPIRIT extension and then for the CONSORT extension (presented as separate items), using a 9-point Likert scale: 1–3=not important, 4–6=important but not critical and 7–9=critical. In addition, an 'unable to rate option' was presented, as was space for free text to explain a given score.

After each survey round, the concept ratings were tabulated and categorised as follows:²⁷

- ▶ Consensus in = $\geq 70\%$ participants scoring 7–9 and $< 15\%$ participants scoring 1–3.
- ▶ Consensus out = $\geq 70\%$ participants scoring 1–3 and $< 15\%$ of participants scoring 7–9.
- ▶ No consensus = everything else.

For each survey round, participants were invited by email (with weekly reminders sent) to complete the survey. Each round was open for 4 weeks, with a 6-week interval between subsequent survey rounds. Only those participants who completed round 1 were invited for round 2, and only participants who completed round 2 were invited for round 3.

In all survey rounds, two linked documents^{28 29} were provided on the Delphi survey and on the email informing that the survey was open. Document 1 contained keywords and definitions relevant to factorial trials to guide participants who were less familiar with factorial trial terminology. Document 2 mapped the concept for consideration in the extension guideline against the original SPIRIT and CONSORT checklist. In round 1, participants were asked to rate the concepts (n=50 concepts) and suggest new items for inclusion in the extensions if they felt appropriate.

In round 2, an additional help document was added to provide clarification for some concepts.³⁰ This was developed after reviewing participants' comments left in round 1. In round 2, participants were presented with all the items from round 1 plus two additions based on suggestions for new items (n=52 concepts). For each survey item (each concept), participants could see how their score compared with that given by other members of the survey (group ratings for each concept were displayed as percentages). Participants were required to re-rate each concept. If their new rating crossed a boundary (eg, moved from 'important but not critical' to 'critical'), participants could provide a justification for changing their score if they wished to do so.

In round 3, a further document was provided to participants, which presented the reasons given by the Delphi members for changing their score where available. Participants were requested to rate the concepts that had not reached consensus and any additional items which were added after round 1.

Project phase 3: consensus meeting

A 2-day hybrid meeting was held at the University of Nottingham. 15 participants comprised the core RAFT team (5 attended in-person and 4 online) plus six leading experts in guideline developments and publishing as well as trial design (4 in-person and 2 online). All six experts participated in the Delphi survey and had an international reputation for improving reporting standards, including authors of previous SPIRIT and CONSORT papers³¹ and editors for leading journals (BMJ and JAMA). At the meeting, concepts which had reached consensus to include were discussed in terms of their utility in being included in the extension documents and in terms of their specific wording in the final documents. Concepts which did not reach consensus were also evaluated for their importance to include in the final documents.

Before the meeting, the concepts that met the criteria for inclusion in the extension checklist after the Delphi study, were reviewed by the study team. This resulted in the concepts being condensed into specific items. These items, along with draft E&E text, were circulated to the meeting members 1 week prior.

At the meeting, a brief introduction to the RAFT study was presented to the members of the consensus meeting, summarising the methodology process and outcomes to date. Session chairs were assigned to lead and focus discussions on the concepts and E&E text. Regular breaks were held to maintain the comfort and engagement of the members. Extension texts (ie, items and E&E) were discussed first in relation to CONSORT extensions and then in relation to SPIRIT extensions. This comprised the majority of the 2-day meeting. In the latter stages of day 2, the three items not at consensus were discussed, followed by approval of the dissemination strategy. Key discussions and outcomes were recorded.

Phase 4: checklist finalisation

Following the meeting, the core study team worked to finalise the wording of the checklist items and E&E documents. These were circulated back to the members for review through an iterative process. Three authors (AAM, BCK and EJ) pilot tested the extensions by each considering a different completed factorial randomised trial and using the checklists to briefly outline key points in the study protocol and main results publication manuscript. This process enabled the authors to reflect on the clarity of the item wording.

RESULTS

Phase 1: scoping review

The records identified, included and excluded are displayed in [figure 1](#).

A total of 3772 articles were identified from the MEDLINE search, with a further two articles included from our personal collections.^{12 32} After title/abstract screening, we were left with 112 possible articles. We

excluded a further 83 articles after full-text screening, leaving 29 included papers (see Appendix 1).

A total of 31 reporting recommendations were identified from the 29 articles, 19 recommendations were applicable to the SPIRIT extension and all 31 recommendations were applicable to the development of the CONSORT extension. These reporting recommendations were termed ‘concepts’ and formed the basis for the questions in the Delphi survey.

Phase 2: Delphi survey

Characteristics of the Delphi participants are described in [table 1](#). The samples were predominantly statisticians. The majority had over 15 years’ experience in clinical trials, had been involved in 1–2 factorial trials and were UK based.

After each survey round, a report was circulated and discussed by the RAFT study group via online videoconferencing.

The results of the Delphi survey are presented in online supplemental table S1 (CONSORT) and online supplemental table S2 (SPIRIT) (see online supplemental material).

Round 1 opened on 7 January 2022, and 104 participants took part (57% of those were sent personalised letters of invitation). Of the 50 concepts which were presented for rating across the proposed SPIRIT and CONSORT extensions, 38/50 concepts (76%) met the criteria for ‘consensus in’ (SPIRIT: 15/19; CONSORT: 23/31), 0/50 concepts (0%) met the criteria for ‘consensus out’ and 12/50 concepts (24%) did not reach consensus (SPIRIT: 4/19; CONSORT: 8/31). 18 new concepts were suggested in round 1. The core RAFT team reviewed the suggestions and concluded that many were already covered by concepts that had been included in the Delphi survey or were not uniquely specific to factorial trials. One new concept was subsequently added to round 2, which applied to both SPIRIT and CONSORT and therefore resulted in the addition of two new survey items (see online supplemental tables S1 and S2).

Round 2 opened on 21 March 2022, and 95 participants took part (91% of original sample). 52 concepts were rated in round 2 (SPIRIT: 20; CONSORT: 32). The cumulative totals over the two rounds were as follows: 46/52 concepts (88%) met the criteria for ‘consensus in’ (SPIRIT: 19/20; CONSORT: 28/32), 0/52 concepts (0%) met the criteria for ‘consensus out’ and 6/52 concepts (12%) did not reach consensus (including the two new concepts added; SPIRIT: 2/20; CONSORT: 4/32).

Round 3 opened on 30 May 2022, and 87 participants took part (84% of original sample; 92% of round 2). The six concepts which did not reach consensus in round 2 were rated in round 3. In summary, the cumulative total over the three rounds were as follows: 49/52 concepts (94%) met the criteria for ‘consensus in’ (SPIRIT: 19/20; CONSORT: 30/32), 0/52 concepts (0%) met the criteria for ‘consensus out’ and 3/52 concepts (6%) did not reach consensus (SPIRIT: 1; CONSORT: 2).

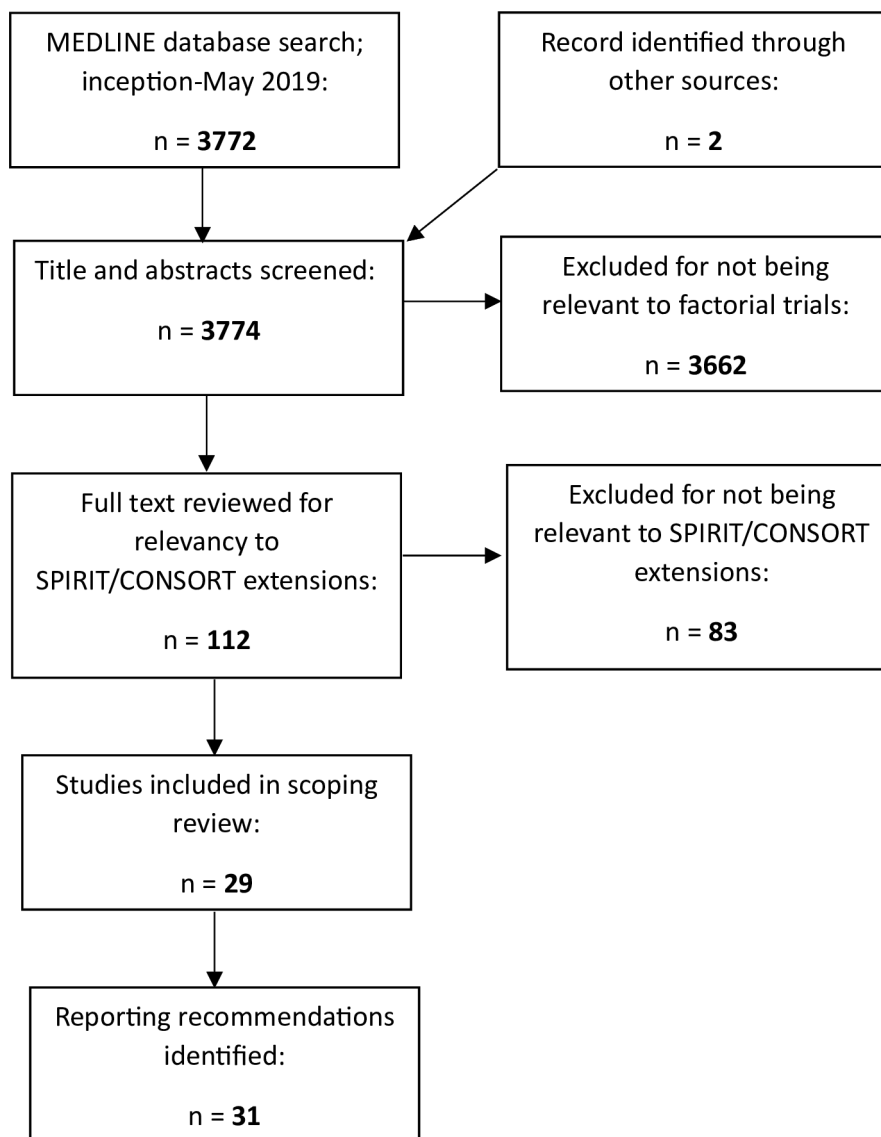


Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram for scoping review. CONSORT, Consolidated Standards of Reporting Trials; SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials.

The Delphi survey proved highly efficient in reaching consensus on the concepts to include in the development of SPIRIT and CONSORT extensions for factorial trials.

Phase 3: consensus meeting

The 30 CONSORT concepts which reached consensus in the Delphi survey were discussed at the consensus meeting. Five were removed from the final checklist, either due to overlap with other concepts or because it was agreed they were not essential for the checklist (see online supplemental tables S1 and S2 footnotes). In addition, the two CONSORT concepts which did not reach consensus in the Delphi survey were discussed and removed from the final checklist for similar reasons. The 25 concepts which remained were then combined into a smaller number of checklist items for the CONSORT extension. Overall, there were 17 factorial-specific items included in the extension checklist, 16 of which were modifications to the

CONSORT 2010 items and one which was a new item. The final CONSORT extension for factorial trials is available in a separate publication.¹⁷ Items relating to the CONSORT extension for Abstracts³¹ were not evaluated in the Delphi survey. Instead, these were chosen after the main checklist was completed in collaboration with the main author (Hopewell) of the original CONSORT for Abstracts.³¹ This approach enabled items that were agreed for inclusion in the main extension document to be considered and combined with expert topic knowledge for abstract writing, preventing the generation of an overly burdensome list given the word limits applied to abstracts.

The 19 SPIRIT concepts which reached consensus in the Delphi survey were discussed at the consensus meeting, and three were subsequently removed from the final checklist, either due to overlap with other concepts or because it was agreed they were

Table 1 Characteristics of the Delphi sample

Characteristic	% (n)		
	Round 1 (n=104)	Round 2 (n=95)	Round 3 (n=87)
Completion rate from round 1	100	91	84 (92 from R2)
Job role*			
Statistician	57.7 (60)	60 (57)	57.5 (50)
Clinical trialist	24.0 (25)	23.2 (22)	25.3 (22)
Chief investigator	18.3 (19)	16.8 (16)	18.4 (16)
Journal editor	16.3 (17)	14.7 (14)	16.1 (14)
Trial/study manager/coordinator	6.7 (7)	6.3 (6)	5.7 (5)
Health economist	2.0 (2)	2.1 (2)	2.3 (2)
Patient and public involvement representative	2.0 (2)	2.1 (2)	2.3 (2)
Funding committee chair/member	7.7 (8)	7.4 (7)	8.0 (7)
Ethics committee chair/member	3.8 (4)	4.2 (4)	4.6 (4)
Other†	10.6 (11)	9.5 (9)	10.3 (9)
Number of factorial trials involved with			
0	19.2 (20)	18.9 (18)	20.7 (18)
1–2	49.0 (51)	49.5 (47)	47.1 (41)
3–5	21.2 (22)	20.0 (19)	20.7 (18)
6 or more	10.6 (11)	11.5 (11)	11.5 (10)
Number of years of experience in clinical trials			
Less than 1 year	1.9 (2)	2.1 (2)	2.3 (2)
1–5 years	7.7 (8)	5.3 (5)	4.6 (4)
6–14 years	32.7 (34)	35.8 (34)	36.8 (32)
15 years or more	57.5 (60)	54 (56.8)	56.3 (49)
Country of residence			
UK	63.5 (66)	62.1 (59)	60.9 (53)
Canada	11.5 (12)	11.6 (11)	12.6 (11)
USA	9.6 (10)	10.5 (10)	10.3 (9)
Europe	6.7 (7)	7.4 (7)	6.9 (6)
Australia	3.8 (4)	4.2 (4)	4.6 (4)
China	1.9 (2)	1.1 (1)	1.1 (1)
Egypt	1.0 (1)	1.1 (1)	1.1 (1)
Other	1.9 (2)	2.1 (2)	2.3 (2)

*Respondents were allowed to select more than one job role if it described their role more accurately, so the table shows their different experiences. Respondents will fit into multiple categories, so the total for each column will not equal the total sample size.

†Respondents were encouraged to choose other if none of the options provided were appropriate.

not essential for the checklist. In addition, the one concept which did not reach consensus in the Delphi survey was discussed and removed from the final checklist for similar reasons. The 16 concepts which remained were then combined into a smaller number of checklist items for the SPIRIT extension. Overall, there were nine factorial-specific items included in the extension checklist, all of which were modifications to the SPIRIT 2013 items. The final SPIRIT extension for factorial trials is available in a separate publication.¹⁷ Plans for dissemination included identifying target

journals for the main checklist and accompanying E&E document, highlighting the importance of publishing the methods for developing the extension checklist to ensure transparency of the process. Plans to conduct workshops/talks, including at the 2024 International Conference for Clinical Trials Methodology Conference,³³ were also discussed with the aim of supporting researchers interested in factorial trials to conduct well-reported trials at all stages from design (ie, the protocol) to the results (ie, the manuscript).

Phase 4: checklist finalisation

Final wording for the extension items and E&E was developed through an iterative review process between the study team and consensus meeting members. Examples of changes to the item wording include using the term ‘main comparison’ instead of ‘primary comparison’ throughout. Examples of good reporting were sourced by the study team, through rapid reviews of the literature and reviewing existing databases of factorial trials held by team members. Preliminary versions of the checklists and E&E were piloted by members of the study team. This resulted in minor modifications to the item wordings to improve reader comprehension.

DISCUSSION

Through a four-stage approach, we have developed SPIRIT and CONSORT extensions for factorial trials. The checklists contain the minimum important information to accurately report on the design and results of a factorial trial. The shared E&E contains further elaboration on why and how these checklists should be applied in report writing, enabling flexibility to their application in practice where applicable. As part of our evaluation of the RAFT study, we aim to share our learning reflections and recommendations for future trials (box 1).

To our knowledge, there are no stand-alone reports of the methodological processes used in the simultaneous development of SPIRIT and CONSORT extensions. Indeed, to date, there has been considerably more investment in developing CONSORT extensions (n=21 published on the CONSORT website, excluding translations) compared with SPIRIT (n=7 published on the SPIRIT website, excluding translations) (data correct as of March 2023).

The RAFT study has highlighted the utility in developing SPIRIT and CONSORT extensions in combination. A well-designed and reported protocol is essential not only to assess scientific, ethical, safety and design issues before trial commencement but also to facilitate evaluation of the trial on completion.¹ Therefore, the concurrent development of both SPIRIT and CONSORT extensions promotes harmony between the two documents, encouraging consistent wording and structures to support the translation from a SPIRIT informed protocol to a CONSORT informed final report. Furthermore, developing these guidelines together proved an efficient use of resources, requiring just one panel of Delphi members and one consensus meeting.

There were several strengths to the RAFT methodology approach. First, we used a consensus-based approach, as recommended by EQUATOR¹⁹ for developing reporting guidelines. Our consensus threshold was clearly predefined to ensure that this was operationalised without uncertainty. Second, we involved key stakeholders from different job roles (eg, statisticians, trialists and editors), including potential end users who may have different perspectives throughout the process to help

Box 1 Reporting Factorial Trials reflections and recommendations

Study development

If extensions for Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) and Consolidated Standards of Reporting Trials (CONSORT) are required, it is resource efficient to develop them simultaneously, using one Delphi survey and one consensus meeting.

Delphi survey

Include key terminology and mapping documents in Delphi survey(s) to guide participants who may not be experienced in some aspects of the guideline development (eg, statistical terminology). A pilot process is useful for identifying potential areas for misunderstanding.

Consider usability and user-friendliness of Delphi software (including access to support) against the cost of licence use. There is a range of software available, which many academic institutions subscribe to, which have the functions required to run a Delphi survey.

To maximise Delphi participation, where possible, schedule the survey rounds avoiding holiday seasons and include prize draws as an incentive to complete all survey rounds.

Including larger (n>50) sample sizes in a Delphi survey has minimal logistical or workload impact.

Consensus meeting

Guideline development typically culminates in a consensus meeting to reach agreement on whether items which did not reach consensus in the Delphi survey should be included. However, meetings with an expert writing group, which focus on specific item wording, should also be considered to improve understanding of the final guidelines.

Hybrid meetings prevent restricting expertise to those members who cannot travel, as well as helping to reduce study costs.

Checklist finalisation

Identification of appropriate examples for including in the Explanation and Elaboration (E&E) represents a considerable aspect of the development of extension guidelines, but this process is often not reported. Appropriate resources should be allocated to this process to complete the extension guidelines on time.

A challenge of developing SPIRIT and CONSORT guidelines together was evident when considering publication strategies, specifically whether a shared E&E would be more appropriate than two separate ones. Key points to consider include the potential overlap between the E&E for the SPIRIT and CONSORT checklist, and conversely, the word limits imposed by many journals making a combined E&E potentially more challenging to synthesise.

ensure that the extensions were applicable to a wide audience. Our targeted recruitment strategy resulted in over 80% of our final sample having experience of at least one factorial trial, ensuring topic-specific expertise. However, it is also advantageous that 20% had not been involved in such trials to provide independent input on the face value of the proposed items. Additionally, both our core RAFT team and our wider expert group had extensive experience in writing and publishing SPIRIT and CONSORT guidelines. This proved valuable in refining item wording, and we hope for acceptability and uptake of the guidelines. Indeed, we spent time during the 2-day meeting considering the detailed wording of specific items derived from the concepts. This approach enabled



us to gain rich, nuanced information from international experts in writing (including editors of leading journals including JAMA and BMJ), which would not be possible in a Delphi survey.

Third, by conducting a scoping review of the relevant literature, we were able to consider issues which widely affect the reporting of factorial trials at an international level and then evaluate the potential importance of these issues through a Delphi survey. Indeed, this process proved highly effective, as evidenced by the majority of original items being rated as important to include and the high degree of consensus reached in the Delphi survey, with only three survey items not reaching consensus after round 3. Fourth, through our multi-strategy recruitment process, we successfully recruited experienced Delphi panel members, the majority of whom had 15 years or more experience in clinical trials. Furthermore, we were able to sustain high response rates over the three survey rounds. Reviews of the methodology of previous extensions to SPIRIT and CONSORT guidelines highlight that two-round Delphi surveys^{5 34–38} are more frequently used than three rounds,²³ sample sizes are typically smaller than achieved in RAFT^{4 5 25 36 38} and response rates in the final round commonly vary from as low as 55–66%,^{5 34 35} with 67% being achieved in reports of a three-round Delphi survey,²⁵ in contrast to the 84% achieved in round 3 of RAFT. It is possible that the inclusion of a prize draw for respondents who completed all three survey rounds contributed to these excellent response rates. By obtaining a larger sample of respondents, we were able to gain wider representation of expert opinions to reach consensus on the final extension documents. This helped to reduce content uncertainty and ensure that the guidance is broadly applicable to factorial trial designs.

Nonetheless, our study is not without limitations. It is possible that by conducting a scoping review as opposed to an extensive systematic review of the literature, some reporting recommendations were missed. This was mitigated in two ways. First, our experienced Delphi panel members were provided the opportunity to suggest additional items in round 1 of the Delphi survey. The review of these suggestions indicated that the majority were re-wordings of concepts already included. Second, members of the consensus meeting critically reflected on the clarity and comprehensiveness of the final guideline items. A second limitation of the RAFT study is that we only had two PPI participants. Participants were requested to select their main role in relation to factorial trials, and as such, it is possible that they also undertook other research-related roles. No additional feedback was requested from these PPI members regarding their inclusivity of the study design from a lay perspective. While PPI is not commonly considered in the development of reporting guidelines, patients often do their own information seeking with regards to healthcare options,^{39–41}

and therefore, developing guidelines which promote the readability of this information to patients could be important. How best to involve PPI members who may lack the necessary experience in key research terminology is an important consideration for future research, and since conducting this study, published guidance is now available.⁴² A third limitation is that all Delphi and consensus meeting members were from high-income nations. While reports indicate that over 80% of randomised controlled trials are conducted by researchers in the top five socioeconomic countries,⁴³ researchers from low- and middle-income countries are likely to have access to a different pool of resources and expertise. This may inform the judgements they make in deciding what represents critical concepts to include in reporting guidelines. Research should consider outreach activities to engage low- and middle-income countries in efforts to produce methodological guidelines.

Fourth, although a mixed-methods approach (including literature reviews and consensus-based methods), alongside substantial sample sizes, increases the generalisability and comprehensiveness of the proposed items, it remains possible that other important concepts relating to a factorial design should be reported in a study protocol and manuscript. As such, we advise that the SPIRIT and CONSORT extensions are viewed as a minimum set of items to include in the reporting process, as congruent with the original aim of the CONSORT and SPIRIT statements.¹³ Finally, the piloting process used represents the study's fifth limitation. The checklist items were internally piloted by members of the immediate research team. Research to support the validity of these extension items should include piloting with less experienced researchers, evaluating individual item comprehensibility.

CONCLUSION

Through a consensus-based approach, the RAFT study developed SPIRIT and CONSORT extensions to support the reporting of factorial trials. The impact of the extension documents for improving study design and reporting will be evaluated in due course, as will the applicability of the extension items in response to the proposed updates to the main SPIRIT 2013 and CONSORT 2010 guidelines (on-going at the time of writing). By reporting the methodological stages involved in the development of the RAFT guidelines, this article provides transparency for the process. Our reflections informed our recommendations for consideration in future efforts to promote best practice in the development reporting guidelines and improve the efficiency of future efforts in this area.

Author affiliations

¹Nottingham Clinical Trials Unit, School of Medicine, University of Nottingham, Nottingham, UK

²Faculty of Epidemiology and Population Health, The London School of Hygiene & Tropical Medicine, London, UK

³Centre for Research in Evidence Based Practice, Bond University, Gold Coast, Queensland, Australia

⁴Women's College Research Institute, University of Toronto, Toronto, Ontario, Canada

⁵Primary Care and Population Sciences, University of Southampton, Southampton, UK

⁶MRC Clinical Trials Unit at UCL, London, UK

X Sophie S Hall @Hall_S_Sophie

Acknowledgements We thank and acknowledge the contributions of all members of the Delphi study: Aaron Orkin, Aiping Lyu, Angela Fidler Pfammatter, Ben Cromarty, Catherine Hewitt, Christine Bond, Christopher Partlett, Christopher Schmid, Claire L Chan, David Moher, Derrick Bennett, Elizabeth George, Evan Mayo-Wilson, Giovannino Ciccone, Graeme S MacLennan, Halvor Sommerfelt, Hams Hamed, Helen Dakin, Himanshu Popat, Ian White, Jay Park, Jennifer Nicholas, Jonathan Emberson, Joseph C. Cappelleri, Julia Edwards, Julien Vos, Kath Starr, Kerry Dwan, Lee Middleton, Lehana Thabane, Lori Frank, Madelon van Wely, Marie-Joe Nemnom, Mark Hull, Martha Alejandra Morales-Sánchez, Martin Law, Martyn Lewis, Michael Forstner, Mike Bradburn, Monica Taljaard, Munya Dimairo, Nick Freemantle, Nuria Porta, Nurulamin Noor, Olalekan Lee Aiyegbusi, Patricia Logullo, Philip Pallmann, Ranjit Lall, Reuben Ogollah, Richard Haynes, Richard L. Kravitz, Robert Platt, Sarah Pirrie, Sharon Love, Shaun Treweek, Siobhan Creanor, Sunita Vohra, Susan Dutton, Suzie Cro, Tianjing Li, Tim Morris, Timothy Collier, Trish Hepburn, Vivian A Welch, William Tarnow-Mordi and Yolanda Barbachano.

Contributors Conception and study design: AM, BCK, EJ, DE, A-WC, EB and PL. Acquisition of data: SH, MB, BCK and EB. Data analysis and interpretation: AM, BCK, EJ, DE, A-WC, EB, PL, SH and MB. Drafting of manuscript: SH and BCK. Critical revision for intellectual content: BCK, SH, EJ, DE, AM, A-WC, EB, PL and MB. Approval of submitted manuscript: AM, BCK, EJ, DE, A-WC, EB, PL, SH and MB. Guarantor: SH.

Funding This work was supported by the Medical Research Council, Methodology Research Panel grant number [MR/V020803/1].

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by University of Nottingham Faculty of Medicine and Health Sciences Research Ethics Committee (FMHS 322-0821). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. Interested parties may contact the corresponding author to gain access to the dataset.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iDs

Sophie S Hall <http://orcid.org/0000-0002-9847-8269>

Edmund Juszczak <http://orcid.org/0000-0001-5500-2247>

REFERENCES

- Chan AW, Tetzlaff JM, Altman DG, *et al*. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 2013;158:200–7.
- Altman DG, Schulz KF, Moher D, *et al*. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663–94.
- Schulz KF, Altman DG, Moher D, *et al*. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Trials* 2010;11:1–8.
- Porcino AJ, Shamseer L, Chan A-W, *et al*. SPIRIT extension and elaboration for n-of-1 trials: SPENT 2019 checklist. *BMJ* 2020;368.
- Vohra S, Shamseer L, Sampson M, *et al*. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *BMJ* 2015;350.
- Piaggio G, Elbourne DR, Pocock SJ, *et al*. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 2012;308:2594–604.
- Eldridge SM, Chan CL, Campbell MJ, *et al*. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ* 2016;355.
- Juszczak E, Altman DG, Hopewell S, *et al*. Reporting of Multi-Arm Parallel-Group Randomized Trials: Extension of the CONSORT 2010 Statement. *JAMA* 2019;321:1610–20.
- Montgomery AA, Peters TJ, Little P. Design, analysis and presentation of factorial randomised controlled trials. *BMC Med Res Methodol* 2003;3:1–5.
- Park JJH, Grais RF, Taljaard M, *et al*. Urgently seeking efficiency and sustainability of clinical trials in global health. *Lancet Glob Health* 2021;9:e681–90.
- Park JJH, Mogg R, Smith GE, *et al*. How COVID-19 has fundamentally changed clinical research in global health. *Lancet Glob Health* 2021;9:e711–20.
- Kahan BC, Tsui M, Jairath V, *et al*. Reporting of randomized factorial trials was frequently inadequate. *J Clin Epidemiol* 2020;117:52–9.
- McAlister FA, Straus SE, Sackett DL, *et al*. Analysis and reporting of factorial trials: a systematic review. *JAMA* 2003;289:2545–53.
- Montgomery AA, Astin MP, Peters TJ. Reporting of factorial trials of complex interventions in community settings: a systematic review. *Trials* 2011;12:1–7.
- Freidlin B, Korn EL. Two-by-Two Factorial Cancer Treatment Trials: Is Sufficient Attention Being Paid to Possible Interactions? *JNCI* 2017;109.
- Cro S, Forbes G, Johnson NA, *et al*. Evidence of unexplained discrepancies between planned and conducted statistical analyses: a review of randomised trials. *BMC Med* 2020;18:1–8.
- Kahan BC, Hall SS, Beller EM, *et al*. Reporting of Factorial Randomized Trials: Extension of the CONSORT 2010 Statement. *JAMA* 2023;330:2106–14.
- Kahan BC, Hall SS, Beller EM, *et al*. Consensus Statement for Protocols of Factorial Randomized Trials: Extension of the SPIRIT 2013 Statement. *JAMA Netw Open* 2023;6:e2346121.
- Altman DG, Simera I, Hoey J, *et al*. EQUATOR: reporting guidelines for health research. *Open Med* 2008;2:e49–50.
- Moher D, Schulz KF, Simera I, *et al*. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010;7:e1000217.
- Montgomery AKB, Elbourne D, Beller E, *et al*. Reporting of randomised factorial trials: development of extensions to the CONSORT 2010 and SPIRIT 2013 guidance statements: open science framework. 2021. Available: <https://osf.io/e3n5r/files/osfstorage/61a8cd100db659030d835575>
- Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. *Inform Manag* 2004;42:15–29.
- Birko S, Dove ES, Özdemir V. Evaluation of Nine Consensus Indices in Delphi Foresight Research and Their Dependency on Delphi Survey Characteristics: A Simulation Study and Debate on Delphi Design and Interpretation. *PLoS One* 2015;10:e0135162.
- de Villiers MR, de Villiers PJT, Kent AP. The Delphi technique in health sciences education research. *Med Teach* 2005;27:639–43.
- Imran M, Kwakkenbos L, McCall SJ, *et al*. Methods and results used in the development of a consensus-driven extension to the Consolidated Standards of Reporting Trials (CONSORT) statement for trials conducted using cohorts and routinely collected data (CONSORT-ROUTINE). *BMJ Open* 2021;11:e049093.
- DelphiManager. n.d. Available: <https://www.comet-initiative.org/delphimanager>
- Harman NL, Bruce IA, Callery P, *et al*. MOMENT--Management of Otitis Media with Effusion in Cleft Palate: protocol for a systematic review of the literature and identification of a core outcome set using a Delphi survey. *Trials* 2013;14:1–8.



- 28 Montgomery AKB, Elbourne D, Beller E, *et al.* RAFT Delphi items. 2021. Available: <https://osf.io/e3n5r/files/osfstorage/61a8c9e0c7d9fb02fdeb562e>
- 29 Montgomery AKB, Elbourne D, Beller E, *et al.* RAFT study: keyword definitions: open science framework. 2021. Available: <https://osf.io/e3n5r/files/osfstorage/642aabdc8ef65b2d59429a93>
- 30 Montgomery AKB, Elbourne D, Beller E, *et al.* RAFT study: response to round 1 feedback. 2021. Available: <https://osf.io/e3n5r/files/osfstorage/642aad074ecbc423e026793d>
- 31 Hopewell S, Clarke M, Moher D, *et al.* CONSORT for reporting randomised trials in journal and conference abstracts. *The Lancet* 2008;371:281–3.
- 32 Kahan BC. Bias in randomised factorial trials. *Stat Med* 2013;32:4540–9.
- 33 ICTMC. 7th international clinical trials methodology conference. 2024. Available: <https://osf.io/ngb98>
- 34 Cruz Rivera S, Liu X, Chan AW, *et al.* Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Lancet Digit Health* 2020;2:e549–60.
- 35 Calvert M, Kyte D, Mercieca-Bebber R, *et al.* Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *JAMA* 2018;319:483–94.
- 36 Kendall TJ, Robinson M, Brierley DJ, *et al.* Guidelines for cellular and molecular pathology content in clinical trial protocols: the SPIRIT-Path extension. *Lancet Oncol* 2021;22:e435–45.
- 37 Dimairo M, Pallmann P, Wason J, *et al.* The Adaptive designs CONSORT Extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *BMJ* 2020;369.
- 38 Thabane L, Hopewell S, Lancaster GA, *et al.* Methods and processes for development of a CONSORT extension for reporting pilot randomized controlled trials. *Pilot Feasibility Stud* 2016;2:25.
- 39 Kubb C, Foran HM. Online Health Information Seeking by Parents for Their Children: Systematic Review and Agenda for Further Research. *J Med Internet Res* 2020;22:e19985.
- 40 Thapa DK, Visentin DC, Kornhaber R, *et al.* The influence of online health information on health decisions: A systematic review. *Patient Educ Couns* 2021;104:770–84.
- 41 Stevenson FA, Leydon-Hudson G, Murray E, *et al.* Patients' use of the internet to negotiate about treatment. *Soc Sci Med* 2021;290.
- 42 Elsman EBM, Smith M, Hofstetter C, *et al.* A blueprint for patient and public involvement in the development of a reporting guideline for systematic reviews of outcome measurement instruments: PRISMA-COSMIN for OMI 2024. *Res Involv Engagem* 2024;10:33.
- 43 Marshall IJ, L'Esperance V, Marshall R, *et al.* State of the evidence: a survey of global disparities in clinical trials. *BMJ Glob Health* 2021;6:e004145.