Response to Round 1 Feedback

In response to the comments left by some members of the Delphi survey, we have provided this table to clarify some items:

Item No.	Item Description	Additional Clarification
General and 13 & 14	Factor	In a 2x2 factorial trial, there are two factors, each with two levels: for example, factor 1 is active drug A vs placebo drug A, and factor 2 is high dose drug B vs low dose drug B. This can be generalised to allow simultaneous evaluation of three or more factors.
3 & 4	Scientific background and rationale for using a factorial design	Explain the rationale for evaluating more than one intervention in the same trial. For example, for efficiency, or to study interactions between intervention.
5 & 6	Justification for whether an interaction was expected or not	Explain whether an interaction between the different factors was expected and why this was/was not expected.
11 & 12	Type of factorial design (such as a full or partial factorial)	Different types of factorial designs are possible. In a full factorial design, participants may be randomised to all possible combinations of all factors. In a partial factorial design, some participants are ineligible for certain factors. A split-plot design allocates factors at different levels, for example cluster allocation to factor 1, and individual allocation to factor 2.
25 & 26	Time-point of randomisation for each factor	When does allocation to each factor occur, relative to other factors? For example, in a 2x2 factorial study, are participants allocated to both factors simultaneously, or are they first allocated to factor 1, and then at a later time-point allocated to factor 2.
27 & 28	Description of estimand(s) for each primary and secondary outcome (treatment comparison, population, outcome definition, population-level summary, handling of intercurrent events)	An estimand is the treatment effect we want to estimate in the trial. Factorial trials pose additional considerations for the estimand, particularly how other factors are incorporated.
29 & 30	Primary approach to statistical analysis (such as factorial; multi-arm) used to compare groups for primary and secondary outcomes; and details on why this approach was chosen	How the different combinations of factors are to be used in estimating intervention effects. This is expected to be consistent with the rationale for choosing a factorial design.
31 & 32	How the other factor(s) will be (or was/were) handled during analysis	How allocation/randomisation to the other factor(s) is accounted for when estimating treatment effect for each intervention (e.g., included as a baseline covariate in the regression model).

46	CONSORT only: Outcome data (including primary and secondary outcomes, harms, and adherence) presented by multi-arm group	For example, in a 2x2 factorial trial, the outcomes would be presented separately for each of the four possible combinations of the two factors.
48	CONSORT: Influence of potential interactions	What are the implications of any observed or potential interaction between interventions on the overall conclusions of the study?
49	Whether adherence to intervention might have been affected by inclusion of other factors	Receiving more than one intervention in a factorial trial may affect participant adherence to allocated intervention(s).

You may find it useful to read the comments left by other members of the Delphi survey, on an item-by-item basis. Please note, these comments have been condensed for ease of presentation.

Comments explaining why participants rated an item a certain way

Items [1] and [2]: Identification as a randomised factorial trial in the title

Item No and Title	Comment
[1] SPIRIT: Identification as a randomised factorial trial in the title	Fundamental for anybody attempting to do an initial sift by title alone; helpful for systematic reviews and literature reviews in general.
AND	There may be a more appropriate term than 'factorial' depending on the exact design. E.g.; 'split plot'.
[2] CONSORT: Identification as a randomised factorial trial in the title	I agree that it is critically important for indexing; however; I don't think it should be mandated in the title. This should be at least in the abstract - not necessarily the title.

Items [5] and [6]: Justification for whether an interaction is expected or not

Item No and Title	Comment
[5] SPIRIT: Justification for whether an interaction is expected or not	Shouldn't this be part of the rationale for using a factorial design?
AND [6] CONSORT: Justification for whether an interaction was expected or not	I think this is important in the protocol when planning the study. I don't see its value at the publication stage. I think readers will interpret the results presented and not whether interactions were expected on not.

Items [7] and [8]: Specification of the research question(s) relating to the factorial design

Item No and Title	Comment
[7] SPIRIT: Specification of the research question(s) relating to the factorial design AND [8] CONSORT: Specification of the research question(s) relating to the factorial design	I think it is more important the research question be specified; than it related to the factorial design.

Items [15] and [16]: Number of levels within each factor

Item No and Title	Comment
[15] SPIRIT: Number of levels within each factor AND [16] CONSORT: Number of levels within each factor	I think there is no great need to make these items overly specific; for example, CONSORT for parallel trials do not have an explicit "number of treatment arms"; as here; but of course it is part of the trial design description.

Items [17] and [18]: The eligibility criteria for each factor; with any differences between the factors if applicable

Item No and Title	Comment
[17] SPIRIT: The eligibility criteria for each factor; with any differences between the factors if applicable AND [18] CONSORT: The eligibility criteria for each factor; with any differences between the factors if applicable	This starts to become tricky as factorial trials can cut across other trials in a matrix fashion. So for example; a factorial trial design to look at interventions to improve retention may be run across a range of different unconnected clinical trials. So the eligibility criteria could be very different for different factors in the same trial.

Items [21] and [22]: Whether an interaction was assumed in the sample size calculation

Item No and Title	Comment
[21] SPIRIT: Whether an interaction was assumed in the sample size calculation AND [22] CONSORT: Whether an interaction was assumed in the sample size calculation	Shouldn't this be covered under details of how sample size was determined for each primary comparison?

Items [25] and [26]: Time-point of randomisation for each factor

Item No and Title	Comment
[25] SPIRIT: Time-point of randomisation for each factor AND [26] CONSORT: Time-point of randomisation for each factor	This would usually be "NA"- it's usual to randomise to all strata at the same time.
	A second factor trial may overlay a lot of separate clinical trials e.g., a methodological trial testing an intervention to improve retention may sit on top of a lot other RCTs which are all very different clinical trials which may not have referred to the 2nd factor in the original protocol; allowance has to be made for this. They are not always co-designed in parallel; a second factor may be added later.
	Taking this as when randomisation to this factor started, in the publication it is critical to know this.

Items [27] and [28]: Description of estimand(s) for each primary and secondary outcome

Item No and Title	Comment
[27] SPIRIT: Description of estimand(s) for each primary and secondary outcome (treatment comparison; population; outcome definition; population-level summary; handling of intercurrent events) AND [28] CONSORT: Description of estimand(s) for each primary and secondary outcome (treatment comparison; population; outcome definition; population-level summary; handling of intercurrent events)	This item is not specific to factorial trials.
	The word "estimand" might be not familiar to non-statisticians.
	A statement of the estimand is critical.
	Definitely for primary outcome; perhaps less so for secondary outcomes
	I've no doubt that this is essential; however; I feel this is a generic item that has nothing to do with factorial trials. I think it should be in the generic extension to the SPIRIT. Otherwise; implementation of these extensions will be confusing and challenging.

Items [29] and [30]: Primary approach to statistical analysis

Item No and Title	Comment
[29] SPIRIT: Primary approach to statistical analysis (such as factorial; multi-arm) used to compare groups for primary and secondary outcomes; and details on how this approach will be chosen AND [30] CONSORT: Primary approach to statistical analysis (such as factorial; multi-arm) used to compare groups for primary and secondary outcomes; and details on how this approach was chosen	How the approach was chosen is not so crucial.
	I don't see how this should be different from other trial designs as the analysis approach should be consistent with the study design to address research question.

Item [34]: Whether any adjustments for multiplicity were applied and method used

Item No and Title	Comment
[34] CONSORT: Whether any adjustments for multiplicity were applied and method used	Multiplicity is important to report; but the two main effects and one interaction in a 2x2 trial do not involve any issues with multiplicity.

Items [35] and [36]: Method(s) used to evaluate evidence of statistical interactions

Item No and Title	Comment
[35] SPIRIT: Method(s) used to evaluate evidence of statistical interactions	Important to choose between and state whether interaction will be assessed on an additive or on a multiplicative scale.
AND [36] CONSORT: Method(s) used to evaluate evidence of statistical interactions	It is really important that guidelines do not lead to an expectation of such analyses just because the treatments happen to be evaluated in the same patient material. Subgroup/interaction analyses should be planned for valid scientific reasons.

Items [37] and [38]: Likely impact of identified interactions on interpretation

Item No and Title	Comment
[37] SPIRIT: Likely impact of potential interactions on interpretation AND	A report should always describe the impact of interactions on interpretation of findings; but this would be interpreting the interactions found; not potential interactions; i.e., one should know the impact for the report, and it should not be potential.
[38] CONSORT: Likely impact of identified interactions on interpretation	I think this is a tricky issue; not sure everyone will be able to state this.

Item [39]: For each primary comparison; the numbers of participants who were randomly assigned; received intended treatment; and were analysed for the primary outcome

Item No and Title	Comment
[39] CONSORT: For each primary comparison; the numbers of participants who were randomly assigned; received intended treatment; and were analysed for the primary outcome	Is this meaningfully different to CONSORT 2010?
	Should this be by each factor; where possible? Of course there are feasibility issues if there are many factor levels

Item 41: Dates defining the periods of recruitment and follow-up

Item No and Title	Comment
[41] CONSORT: Dates defining the periods of recruitment and follow-up; if different across factors; describe reason(s) for the differences and any statistical implications	I don't think describing the reasons for differences in important at all here. At this stage; we just want to know what they did. I'm unsure of any statistical implications should be here.
	I put unimportant just to flag that although it is important for the individual factors; differences are probably usually unimportant.

Item [42]: A table showing baseline demographic and clinical characteristics for each primary comparison

Item No and Title	Comment
[42] CONSORT: A table showing baseline demographic and clinical characteristics for each primary comparison	There needs to be one table but difficult often to have a table that has each primary comparison represented. This is as participants would then be counted more than once. If need to show that arms are balanced then report by each group receiving same combination of factors; or other tables go in supplementary.
	Typically infeasible to show more than one baseline demographics table due to pushback by editors; however online supplements enable additional tables.

Item [44]: For each primary and secondary outcome; results for each primary comparison; the estimated effect size and its precision (such as 95% confidence interval)

Item No and Title	Comment
[44] CONSORT: For each primary and secondary outcome; results for each primary comparison; the estimated effect size and its precision (such as 95% confidence interval)	Sufficiently covered by original CONSORT.

Item [45]: For each primary and secondary outcome; the estimated interaction effect and its precision

Item No and Title	Comment
[45] CONSORT: For each primary and secondary outcome; the estimated interaction effect and its precision	This should be done for the primary outcome; however; I'd be concerned about multiplicity if this was done for all secondary outcomes.

Item [47]: All important harms or unintended effects in each primary comparison

Item No and Title	Comment
[47] CONSORT: All important harms or unintended effects in each primary comparison	I think this is critical; but I wonder if it is already covered by the language in CONSORT; since it refers to "each group".

Item [49]: Whether adherence to intervention might have been affected by inclusion of other factors

Item No and Title	Comment
[49] CONSORT: Whether adherence to intervention might have been affected by inclusion of other factors	I'm not sure how you would reliably tell.