SPIRIT|CONSORT-SURROGATE Delphi survey – summary of Round 1 & Round 2 now open for your completion

Dear colleague,

Many thanks for participating in the first round of Delphi survey for the development of the SPIRIT-SURROGATE and CONSORT-SURROGATE extensions. We had 195 participants from 29 countries who completed round 1. **Based on first round ratings, the second round is now open, and we would ask you to complete by 30th November 2022.**

In round 1, we received 179 free text comments on specific items, 28 general comments, and 26 additional suggested items. For the purposes of transparency and information sharing, these **are provided as appendices to this letter for your scrutiny**.

Nevertheless, we provide here **a summary of the project team’s consideration and actions around the feedback received after round 1.**

# Comments on specific items and modifications

* We used simple thematic analysis to analyse the comments and the common themes identified were support for the item; context of item importance or implementation; modification to the item; and limitations or challenges in implementation of the item – see “Free-text responses per item” section in Appendix 1.
* We carefully read and considered all the comments. Most of the comments especially on modification and context of item importance or implementation will be presented and further discussed during the consensus meeting and writing up of the extensions’ elaboration and explanation documents. We welcome any further comments in this respect.
* Comments on limitations or challenges in implementation of the item, particularly for items that were not highly rated (<70% rating the item 7-9), were used to modify the wording of the item. Modified items in the survey are highlighted with \*\* at the beginning and the modifications are highlighted below:
* **SPIRIT/CONSORT 6:** Clarify if the sample size calculation is/was explicitly informed by statistical metrics of surrogate validity ~~(such as the surrogate threshold effect (STE) or its equivalent)~~
* **SPIRIT/CONSORT 7:** State if trial participants will be/were informed before enrolment that trial was ~~powered~~ designed to evaluate an interventions effect using a surrogate endpoint ~~(rather than a patient relevant final outcomes)~~
* **CONSORT 11:** Provide an estimate (with a measure of uncertainty) of the predicted effect of patient relevant final outcome based on the observed effect on the surrogate endpoint; and if not possible then a qualitative assessment
* **CONSORT 13*:*** If surrogate endpoint and patient relevant final outcome data were collected in the trial; state the open access arrangements for the data for future secondary research including ~~the statistical evaluation of the~~ validation of the surrogate endpoint
* **Definition:** A biomarker or intermediate outcome used to substitute for a patient or participant relevant final outcome (i.e., severe morbidity; health related quality of life or mortality) and reliably predicts benefit or harm based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence

# General comments

* All the general comments are listed in “General comments” section of Appendix 1
* Most of these comments were complementary. Some of them raised important issues which we have taken note and will take forward in the next phases of the project.

# Suggested additional items

* We carefully considered all the 26 suggested items and have provided a point-by-point response to each – see Appendix 2.
* We felt that the majority of suggested items (n=18, 69%) were already covered in rated items or could be incorporated in those items during writing of the extensions’ elaboration and explanation documents. Furthermore, we thought that four items (15%) were beyond the scope of the project or were impractical to impose as reporting guidelines.
* One additional extension item was included for rating in this second round:
* SPIRIT/CONSORT item under Methods: State what other surrogate endpoints were considered and why the current one(s) were chosen over those
* One additional definition has also been presented for rating:
* An endpoint replacing a clinical endpoint that constitutes a basis for reliably predicting a treatment effect on the clinical endpoint in a defined context of use.

We hope you found this cover letter and high-level feedback summary useful. **We kindly ask you to complete this second and final round of the Delphi survey by 30th November 2022**. In case of questions or suggestions, do not hesitate to contact us via email or social media, see below for contact details.

With many thanks for your contribution towards the development of these important reporting checklists,

Dr Anthony Manyara, Professor Rod Taylor, and Professor Oriana Ciani on behalf of the SPIRIT|CONSORT-SURROGATE Project Team

Email : consort-surrogate@glasgow.ac.uk |Web page : <https://www.gla.ac.uk/spirit-consort-surrogate> | Twitter : [@Consort\_surr](https://twitter.com/Consort_surr)

# Appendix 1

Items that have been modified are highlighted, SPIRIT/CONSORT.

**Free-text responses per item**

**SPIRIT/CONSORT 1 & 2**

*Item 1: State that primary outcome(s) is considered a surrogate endpoint*

*Item 2: State the participant/patient relevant final outcome(s) that the surrogate endpoint is substituting and predicting for*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Support for the item | * Important for transparency
 |  |
| Support and challenges for having item in abstract | * Many researchers read abstracts only. If the surrogacy is not contained within the abstract it may be difficult to locate evidence
* If only the abstract will be read the evidence supporting surrogacy is overlooked.
* Usually limited word count in abstract; would need a couple of sentences to explain it
* Space in the abstract is always too much limited
 |  |
| Include item in headings | * It can also be mentioned in headings and subheadings
 |  |
| Modification to wording of item and definitional issues | * Not "is considered" but IS or is not a surrogate endpoint. Needs clear definition. Would define as "direct" patient outcomes of survival; patient symptoms or patient functions in their daily lives and everything else including Clinician Reported Outcomes are "indirect" measures. Too much vagueness on what is and is not a "surrogate". Could also divide into validated and candidate surrogates
* Considered by whom? There is no "official" list of surrogates... Journals should include a paragraph specifying their definition
* my preference would be to state that the primary outcome is a biomarker or intermediate outcome that is being considered (or has been considered) a surrogate endpoint.
* a clear distinction should be made between intermediate endpoints and surrogate endpoints. I think that many investigators will report intermediate endpoints as "surrogates" even if no formal surrogacy evaluation have ever been performed
* Again terminology important; not "final" outcome but direct patient outcome of survival patient symptoms or patient function; Too many studies use another surrogate as the "final outcome". Tuberculosis trials use a surrogate at an early time point with "final" outcome as same surrogate at later time point.
 |  |
| Need for validation evidence and challenges | * In early phase studies an outcome with strong correlation is probably sufficient. But in the case of a phase III study a surrogate should be a validated one. I feel that evidence on the validity of the surrogate will be difficult to find.
* This exercise is presumed to use a marker that has already been validated as a surrogate endpoint and therefore has evidence to adequately support the multidimensionality needed. For example; use of the BSES (ref) as the evaluation tool requires sufficient evidence in four domains study design; target outcome; statistical evaluation and generalisability. What follows assumes the same.
* Including discussion of evidence to support that claim and what clinical outcome the surrogate is intended to predict
 |  |

**SPIRIT/CONSORT 3**

*State the participant/patient relevant final outcome(s) that the surrogate endpoint is substituting and predicting for*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Support for item | * Absolutely crucial. Too many surrogates used in acute diseases where direct patient outcomes can be measured in short period of time. Used to "validate mechanism of action" which is not primary goal of confirmatory trials which should be to confirm patient benefits and harms e.g., urine culture in acute urinary tract infection - no reason to measure urine culture (not done in practice) in disease that last about a week
* Helpful to have this rationale
 |  |
| Move item to methods | * I would add this to the Methods section
* Could be combined with items in Methods below
 |  |
| Evidence for reasons and justification | * It will be better to provide evidence for their reasons
* justification for surrogate endpoint is critical
 |  |
| Other comments | * overall survival is not a surrogate outcome to me - cancer related death may be
 |  |

**SPIRIT/CONSORT 4**

*Justification for selected surrogate: Evidence of validation*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Challenges with validation evidence  | * Confusion regarding "validation" between patient level correlate independent of treatment effects and treatment effects at trial level on surrogate and direct patient outcome. The formed is only a "candidate" surrogate and not 'validated" but often presented as such. Also need to know WHEN the direct benefit is expected to occur. Too many studies state that direct patient benefit will be in "future" but WHEN? This is key to understanding relationship between surrogate and direct outcomes and also to ethics of study or even if and when surrogate should be used
* There are validated surrogate outcomes in major disease area like cancer; HIV infection; stroke... etc. But I suspect many other areas have putative surrogate with just associations considered sufficient.
 |  |
| Context of item importance | * Depends on how widely used the surrogate is
* might be the only available surrogate; even if not validated
 |  |
| Move to introduction | * Could be combined with Introduction above
 |  |

**SPIRIT/CONSORT 5**

*Justification for selected surrogate: Evidence of being specific to setting used e.g., intervention; disease; population*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Support of item | * Surrogate often used outside of setting in which they were developed of CHANGE definition or ANALYSIS METHOD of surrogate e.g., a valid surrogate using proportions as a specific timepoint is NOT necessarily valid when using AUC over time - see HIV viral load where there were attempts to change the definition of a previously valid surrogate
* Population and design features need reconciling with; e.g., BSES or near equivalent
 |  |
| Context of item importance | * The importance of the item is context dependent. Where the surrogate and its evidence are well-known this is less important; but for new surrogates it is crucial.
* some surrogates are very well established in the disease setting. This would be more important for new or rarely used surrogates
 |  |
| Unify with item 4 | * Item 5 is really included in item 4. The description in item 4 is incomplete if it does not include item 5.
* I think that this item can be unified with SPIRIT 4
 |  |
| Other comments | * It will be better to provide an evidence for their justifications
* not sure what difference between practical reasons above and justification here
 |  |

**SPIRIT/CONSORT 6**

*Clarify if the sample size calculation is/was explicitly informed by statistical metrics of surrogate validity ~~(such as the surrogate threshold effect (STE) or its equivalent)~~*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Context of item importance | * I think that more broadly the sample size calculations should be justified and explained; whether informed by STE or not
* only for new surrogates
* I think it would be more useful to require authors to report whether there is a STE (and if so, what it is) without linking it to sample size; as then readers can interpret both sample size and results based on this
* This depends on the disease; rare diseases or those with minimal research will not be able to provide this. It is also not a good way to develop guidelines when we are at a time of conducting simulation trials and digital clinical trials. AI use is becoming more common now and for those types of studies or trials or the use of these methods in other trials; this point is irrelevant.
* I consider it critical to establish in Methods that the study is powered/sample size is appropriate to support the surrogate outcome being met or not met. The details of what informed sample size calculation is important but less critical (for example; could be supplemental information provided to a stats reviewer; but of less interest to a clinician or patient)
 | Based on feedback received we have modified the item to strike out specification of surrogate threshold effect to reflect that the sample size calculation was informed by metrics of surrogate validity only  |
| Practicality in implementation and limitations of STE | * It will be so hard for some investigators to calculate STE for the surrogate
* For time-to-event endpoints; methodologies to derive STEs are mostly based on the hazard ratio as a measure of effect. Requiring the use of such metrics may perpetuate the predominant use of certain measures of effect, such as the hazard ratio.
* STE is a good criterion only if the trial-level surrogacy is very high; also critically the previous data used to estimate STE are with extremely high quality. In many cases, neither can be fulfilled. Furthermore, STE relies on strong linear assumption at the trial level; this measure takes different meanings for different types of endpoints; there is no clear idea how large STE is sufficient to show surrogacy evidence. Although it is important to specify sample size calculation regarding prediction of treatment effect on true endpoint; STE is not a good measure for this purpose.
* Very difficult on a practical basis... in cancer very few STE have been properly calculated.
* This will be probably difficult to find unless every study have a systematic review and meta-analysis as a precursor
* it seems unrealistic as the benefit of using a surrogate endpoint would be lost
 | See response above |

**SPIRIT/CONSORT 7**

*State if trial participants will be/were informed before enrolment that trial was ~~powered~~ designed to evaluate an interventions effect using a surrogate endpoint ~~(rather than a patient relevant final outcomes)~~*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Support for item | * Almost never occurs but this is critical to understanding benefit vs harm in the consent form as benefit is on surrogate and harms are often direct harms e.g. Alzheimer’s drug where 'benefit" was lowering amyloid and harms were brain bleeding and swelling
 | A note here is that in our subgroup analyses of this item, we found that it was favourably rated by PPI members (median of 8 against overall group median of 6) |
| Context of item importance or implementation | * would defer to the patients concerned in this and it might be subject specific
* would defer to patient and public views on this and it might be disease specific
* unless 'power' is explained before this question via; say; participant information sheet; this question may have no or limited relevance
* Powered to calculate is important. However; as someone who was an Ethics committee chair; I know the small sample sizes people have come up with and justified for a variety of practical reasons such as lack of funding for large sample sizes given limitations in resource. Especially in the U.K. where the NHS is struggling and limited academic institutions can independently recruit patients; this needs to be carefully thought after. Standardisation is great but it has a downside too unless funders and journals agree there needs to be a level of flexibility. Studies with 30 sample sizes are endless if you check either early feasibility or pilot trials. The reality is guidelines are to provide a guidance and its not a mandate. One cannot force this on research communities unless funders; regulators and all other parties have agreed to unified resourcing etc. There needs to be an acknowledgment here for example trials being conducted in low resource or income setting may not meet the standard written here. Having conducted studies in rural Africa I’m aware of the challenges associated with data collection. So, one needs to be pragmatic and be clear with these statements and its interpretation for example by journals and reviewers like.
 | See our response above |
| Challenges in implementation  | * clinicians are generally unaware of the nuances of surrogate outcomes. The lay public has no interest - nor should they.
* presumably which endpoint is used for a trial will be included in the consent. But whether it is articulated the endpoint is a surrogate endpoint could be questioned. In medical research; very few research really understand what surrogate endpoint is and what validations mean. It could be very challenging that the consent form is formulated accurately on this concept. A yes/no statement here may not be sufficient.
* Too difficult... acceptance would go down (and if the trial has been considered overall ethical this would be wrong) and the ones accepting would loose confidence in what they are going to do...
* Some reservations here because given lack of general understanding of clinical trials among patients; explaining endpoint surrogacy would be challenging to accomplish.
 | See our response above |
| Question | * What are precedents here?
 |  |

**CONSORT 8**

*If the primary outcome is a composite outcome that includes a surrogate endpoint; report the intervention effect on all components*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Support for item | * The importance of this cannot be overstated. Reporting on only the surrogate endpoint and ignoring other components of the primary outcome would count as cherry-picking data
* Critical because the conclusion would otherwise be biased; particularly when the overall effect size is being driven by a small number of components
 |  |
| Context of item importance or implementation | * At a minimum; I would want to see the outcomes that are making up the composite outcome and distributions of the individual outcomes; as well as the composite outcome
* But this should apply for all composite endpoints whether or not some components are surrogates
* if the choice to use a composite endpoint is justified I think that it is not mandatory to present results for each component
 |  |
| Challenges with implementation and support for item | * Study will likely be underpowered for components; however; even not significant trends are needed for understanding the meaning of the primary surrogate outcome
* This may be underpowered for individual components and hard to interpret; with multiplicity adjustments etc
 |  |
| Suggested addition | * Also need to be able to evaluate relationship between components e.g., how many have discordant outcomes like "success" on surrogate and failure in direct patient outcomes e.g., who has "negative" urine culture but still has symptoms and vice versa. Impossible to assess in most trials and also need to have how much missing data on each component - often see different denominators on different components. Finally, often see in infection trials the assumption that if surrogate not captures (e.g., patient can't make sputum in tuberculosis trial) that the surrogate is "negative" based on direct patient outcomes. This seems to obviate use of a surrogate if its outcome is assumed based on patient outcomes
 |  |
| Need for clarification | * Not sure that I fully understand this statement. A true endpoint is mostly likely a simple endpoint, such as overall survival. but it is very common that a surrogate endpoint is a composite endpoint; and most of time in oncology; the true endpoint is one of the components of the surrogate endpoint (e.g., DFS or PFS). Some clarifications of this statement would be great.
 |  |
| Other comments | * no spec
 |  |

**SPIRIT 8/CONSORT 9**

*Comment on whether the trial sample size and follow up period of the surrogate endpoint is sufficient to adequately capture potential harms of the intervention being tested*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Support for the item | * Critical as harms and benefit measured on different outcomes as noted above
* This is critical for any associated health economic analysis to establish whether and how to include the costs of adverse events.
 |  |
| Context of item importance or implementation | * It may be required to mention this as an inherent limitation when using surrogates as end points
* Depends on the objective of the study (either pivotal trial using a surrogate or an earlier proof-of-concept study)
* This would be very important for studies testing new medicines. I work with exercise/rehabilitation trials; so, the possible harms are minimum and acute.
* My response assumes that the primary aim of the trial is not to investigate possible harms.
* More interested in completeness of data for adverse events; such as the observed person-time
* It may be required to mention this as an inherent limitation when using surrogates as end points
 |  |
| Limitations of item | * Sample size has nothing to do with really in gathering SAEs or SUSARs and in general developing a safety profile. This is a separate matter entirely and this is being asked without considering what SAE/AEs mean to different populations. Before all of this there should be questions around age; gender and race/ethnicity. Its very unclear how one can talk about justifying sample size without first addressing key characteristics in a population and its relevance to the generalisability of its findings. So before talking about these sections this questionnaire should have been carefully thought about from a population demography perspective.
* Again how can one define answer to this question when follow up periods vary depending on the disease population features funding etc If you add a statement like this to a set of guidelines one needs to really think if we are introducing discrimination and prompting an elitist mentality towards conducting trials. The whole point should be to conduct flexible and high-quality trials that are relevant to population demand
* powering for safety and efficacy are unrelated.
 |  |
| Not specific to surrogate endpoints | * While obviously important; I'm not sure this is specific to surrogate outcomes? E.g., the same consideration could apply to a clinically relevant outcome; as often harms are rarer so would require longer follow-up/larger sample size to detect
* I'm not sure how to rate this since it is not specific to surrogate endpoints
 |  |
| Modification to item | * The additional wording doesn't seem to add value; because it is the sample size and follow up period of the trial that determines whether potential harms can be captured
* But this is better provided in the Discussion
 |  |

**SPIRIT 9/CONSORT 10**

*State if there are explicit plans to extend follow up and/or conduct subsequent analyses/studies to verify benefit of current findings on the patient relevant final outcome*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Modification to item | * And WHEN the confirmatory study will be done. Confirmatory studies still not done for many interventions a decade or more after initial study using surrogate
 |  |
| Limitation of item | * Not sure of the value of a non-binding intention
 |  |

**CONSORT 11**

*Provide an estimate (with a measure of uncertainty) of the predicted effect of patient relevant final outcome based on the observed effect on the surrogate endpoint; and if not possible then a qualitative assessment*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Support for the item | * The qualitative assessment is critical here; estimate of predicted effect is less so; as such estimates are often highly conditional
* This almost never presented - just assumed
 | Based on feedback of this item we have added “with a measure of uncertainty”  |
| Context of item importance or implementation | * Would be very useful to see this; but will require identification of a specific target population
* Depends on how accurate that estimate is likely to be - needs supporting evidence
* if possible, the true effect is a secondary outcome (e.g. OS)
* Potentially important; but raises other questions that would then need to be addressed in the reporting; particularly related to the validity of predictions
 | See above |
| Challenges in implementation | * I think this could be problematic; as estimation of the link between surrogate and clinically relevant outcome is quite challenging and will often be done poorly; so I'm concerned the providing a predicted effect on the clinically relevant outcome will provide a false sense of reassurance
* So far; based on what methodology in surrogacy exist; this is an impossible task. To force it; the consequence can be author-biased claims based on speculations.
* would be very subjective...
 | See above |
| Modification to item | * suggestion editing: “Provide an estimate of the predicted effect of patient relevant outcome based on the observed effect ..."
 | Done |

**CONSORT 12**

*Interpretation of findings of the trial in the context of using a surrogate primary endpoint including its known validity and the potential benefit-risk ratio of the tested intervention for participants*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
|  | * Explicitly state "benefit" is on surrogate, and harms are often direct outcomes
* Would rather see this as an independent exercise as this seems like a not very portable estimate. That is; it would depend on the rates in the target populations
 |  |

**CONSORT 13**

*If surrogate endpoint and patient relevant final outcome data were collected in the trial; state the open access arrangements for the data for future secondary research including ~~the statistical evaluation of the~~ validation of the surrogate endpoint*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Support for the item | * Important for transparency
* This is very important. There is a lot work that can be done using trial data in relation surrogate validation and evidence confirmation
* it would be good to encourage open access for future analysis of the data
 | Slight modification to read validation rather than evaluation of the surrogate endpoint |
| Context of item importance or implementation | * I think it would be more useful to require investigators to collect the patient relevant final outcome as part of the CONSORT statement; and for them to report both it; and the statistical evaluation of surrogacy
* Need patient level data to evaluate both patient level correlate and trial level surrogacy and missing data on each type of outcome
* Also as well as access to IPD we need structured machine readable outputs. This should include sufficient statistics to fully describe the models (e.g. vector of coefficients and the variance-covariance matrix of any models)
* NIH Data Sharing policy goes into effect in 2023 and this would need to be specified per that policy; at least in grant applications. This would need to be consistent with that policy/specific plans for the trial
 |  |
| Not specific to surrogates  | * for all studies; publicly funded or privately funded; a data access policy is preferred.
* Data sharing should be a must in any case
* Should be covered by general statements about open access
* This might be funding/resource dependent; not sure I would deviate from the standard data sharing statement
 |  |

**Definition 1**

*A biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence.*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Not inclusive  | * It's not only biomarkers that can act as surrogate endpoints. Suggest enlarging by "A biomarker; intermediate or any other endpoint"
* not exclusively a biomarker
* In oncology; does not address tumour response as a surrogate for survival; for example; since response is not necessarily a biomarker.
* Seems quite restrictive as not all surrogates are biomarkers
 |  |
| Other limitations | * Completely confusing based on the abused term "clinical" benefit. Clinician Reported Outcomes based on clinician observations are SURROGATES yet often presented as "clinical" endpoint - really really confusing for readers
 |  |
| Context of item importance  | * A fair description of an unvalidated surrogate outcome. Surrogates can never be endpoints!
* almost perfect definition in my view - without "based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence"
 |  |
| Suggested modification  | * the evidence should be experimental (statistical evidence of surrogacy) more than anything; event if the epidemiologic and pathophysiologic evidence is also important. I am not sur I understand correctly the statement
* I would add two mandatory statements like: 1) What is the mechanist evidence that surrogacy is acceptable? 2) what is and how was derived the relation that enables to compute the size of effect on the final outcome from the effect on the surrogate?
* It will be better to provide an evidence for their explanations
 |  |
| Comments on scale | * The headings on this page do not match the definitions of the nine categories given at the top of the page.
* scale above doesn’t match instructions
* the response should be continuum of complete and inclusive rather than importance continuum. Should complete and inclusive be the only response of interest? This is the traditional NIH definition and for this reason perhaps it should not be changed?
 |  |

**Definition 2**

*A laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful end point that is a direct measure of how a patient feels; functions; or survives and is expected to predict the effect of the therapy*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Not inclusive | * Clinical SIGNS often ignored as surrogates which they ARE and since they are measured by clinicians often presented as "clinical endpoints" but not direct patient endpoints
* a surrogate could be any information (e.g., speech or imaging); also and outcome in trials on prevention
* not exclusively in therapeutic trials
* should it just be therapeutic trials? e.g. What about diagnostic studies. I like that it includes a physical sign as. It depends on how one defines a biomarker. Can we suggest a definition?
* In oncology; does not address tumour response as a surrogate for survival; for example; since response is not a lab value..
* Again; not all surrogates are lab measures of physical signs
* This mixes the definition of a surrogate and the definition of a clinical endpoint. Effect of therapy is too unspecific
 |  |
| Modification to item | * predict the effect therapy "ON THE PATIENT-RELEVANT ENDPOINT"
* I would add two mandatory statements like: 1) What is the mechanist evidence that surrogacy is acceptable? 2) what is and how was derived the relation that enables to compute the size of effect on the final outcome from the effect on the surrogate?
 |  |
| Other comments | * Also relevant to vaccine trials and other prophylactic interventions
* This statement seems like a confusion of quantitative and qualitative measures - feels; functions and survives are three very different things.
 |  |

**Definition 3**

*A response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint.*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Lacks clarity | * Not written clearly at all
* too complicated
* requires a plain English explanation
* I cannot rate. It is a spurious statement. User of surrogate guess it is true but cannot support it; except from gut feeling
* Too statistical - it’s true; but clinicians won't understand this.
* Not sure what is meant by response variable - any kind of variable hat records a change in a patient’s condition or a questionnaire response. If the former this definition effectively includes all the others so is the most general
 |  |
| Other comments | * The problem with a null hypothesis reliance is that patients reading the study will mostly not understand the process here.
* This also refers to ANALYIS of treatment effects and not to WHAT Is measured. A surrogate can be used in an observational study as well which does not involve treatment effects. Separate issue of "what is a surrogate" vs how it is used
* Theoretically correct; but probably not of great help in practice.
* the concept of surrogacy is unrelated to statistical testing
* This could corresponds to a non eligible confounder
* statistical Prentice type definition
* the quantification of the expected benefit on the clinically relevant endpoint is also important
 |  |

**Definition 4**

*An endpoint that is used in clinical trials as a substitute for a direct measure of how a patient feels; functions; or survives. A surrogate endpoint does not measure the clinical benefit of primary interest in and of itself; but rather is expected to predict that clinical benefit or harm based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence.*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Modification to item | * My concern with this statement is the verb "predict". In surrogacy setting, predict means that there is a validated relation that enables to infer the size of the benefit on clinical outcome (the 'final' outcome) from the effect on the surrogate. Most of the tie such a function does not exist or is based on correlations (unreliable alternative; further often measured in different settings). Thus, instead of 'predict'; I suggest 'guess' in this statement.
* “is expected to predict” is not specific enough. Suggest adding in "has been shown to reliably predict".
 |  |
| Other comments | * This gets to HOW surrogate is used in a trial whereas second definition above refers to WHAT Is measured; this gets to HOW it is used to measure treatment effects
 |  |

**Definition 5**

*A biomarker or intermediate outcome used to substitute for a patient or participant relevant final outcome (i.e., severe morbidity; health related quality of life or mortality) and reliably predicts benefit or harm based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence*

|  |  |  |
| --- | --- | --- |
| **Theme** | **Free text comment** | **Response from project team if any** |
| Context of item importance | * Seems restrictive as not all trials seek to provide information about severe morbidity; HRQoL or death
* They may be all valid or not depending on whether or not there is a proper validation study...
* A fair description of a validated surrogate outcome
 |  |
| Modification to item | * ...predicts... probably not precise enough. How about " and reliably predicts"?
* Without the text "based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence" very useful
* predicts benefit or harm could be improved (i.e. more accurate) if changed to something like "is expected to predict" or "is likely to predict"
* Used IN CLINICAL TRIALS to substitute...? ...and IS EXPECTED TO PREDICT CLINICAL benefit or harm OF TREATMENT? Provide example of biomarker and example of intermediate outcome?
* My concern with this statement is the verb "predict". In surrogacy setting, predict means that there is a validated relation that enables to infer the size of the benefit on clinical outcome (the 'final' outcome) from the effect on the surrogate. Most of the tie such a function does not exist or is based on correlations (unreliable alternative; further often measured in different settings). Thus, instead of 'predict'; I suggest 'guess' in this statement.
 | “reliably” added to the definition  |
| Other comments | * The term "intermediate" is not well understood and would include patient symptoms and function not just distal measures of quality of life which is distinct from direct measures of symptoms
* there's so much variance for each of these points that I can't answer
* a surrogate is really designed for the context of comparative trials
 |  |

**General comments**

**Definitions**

* To repeat; surrogates can never be endpoints. Of clinical endpoints; I only know two: anus and death.
* You write overall survival as a surrogate or at least I read that. To me it matters whether or not I live or die so would use another example
* It is not meaningful to try to respond to the last 5 questions. The validity of surrogates needs to be decided upon in each individual case; e.g. sometimes a biomarker is just rubbish and sometimes it is a good predictor for a relevant outcome. Otherwise; very good work!
* Important to differentiate biomarkers (https://www.ncbi.nlm.nih.gov/books/NBK326791/; e.g., serum biomarkers; imaging biomarkers) using a formal FDA-defined Context of Use; as opposed to general endpoint surrogates (e.g., image response criteria; patient reported toxicity or progression-free- or recurrence-free-survival).
* Regarding definitions of surrogate endpoints there is the issue of whether intermediate outcomes should be included in the definition. There are pros and cons for this. If an intermediate outcome is how a someone feels; functions or survives; then this has face validity and may need no additional consideration. If the intermediate outcome is not how a someone feels; functions or survives; then it is a surrogate endpoint. Should we split; and call it an intermediate endpoint because it is not a biomarker or should all be collapsed into a generic term marker. it is possible; given the complexities of the issues; that no perfect definition can be crafted. I would be satisfied as long as there is transparency and consistency in reporting.

**Extension items**

* I have already made some comments in my earlier responses as I did not know the scope and specifics of the exercise. I scored all statements as critically important. ----I would like to raise one matter. In reading some of the statements I thought that it was possible that the statement may encourage publications of poorly validated surrogate endpoints. Maybe I am overthinking it but could some of the statements; as written; contribute to “fait accompli” of unvalidated surrogacy to readers that are unaware of the issues. For example, the first statement (SPIRIT/CONSORT 1): state the primary outcome (s) is considered a surrogate endpoint may imply to readers that it is a validated surrogate endpoint. I suggested alternative wording in my earlier comments.
* Furthermore, should there be a statement about whether it is a validated surrogate endpoint (but how to determine validation?) be in the introduction. Or would this be repetitious or redundant. I am sure that these issues will be addressed extensively in the publication.
* Statements regarding evidence (SPIRIT /CONSORT 4). I suggested explicitly reporting the different types of validation evidence a) Pathophysiologic and /or b) Epidemiologic and /or c) Therapeutic and /or d) Statistical and /or e) other scientific evidence of validation of the surrogate endpoint in the reported trial.
* Initially I thought there would be statements on “surrogacy statistics” in the Results. However, the aim of the exercise is to improve reporting of trials with surrogate endpoints and not the statistical validation of surrogacy so surrogacy statistics is not needed.
* Three major issues regarding the demonstration of relevance of proposed surrogacy seem far from being enough accounted for in the proposed reporting guide: 1) mechanist feature(s) of the link between surrogate and clinical outcomes (correlation or any other statistical link are too weak evidence); 2) quantitative link between the two efficacies which is needed to predict the size of benefit on the clinical outcome from the observed effect on the surrogate; 3) the domain of validity of this link both in terms of therapies and of patients ' profiles. These three issues support the strength of evidence of the guess on the efficacy that matters to patients.
* The clinical trial community needs to be better educated regarding the definition of a surrogate endpoint; and how it's distinguished from a clinical endpoint. However, it's not the job of a protocol to provide that education. It should be obvious to anyone reading a protocol whether or not the primary endpoint is a surrogate; and so it shouldn't be necessary for the protocol to mention that fact. It would be far more helpful to include in the protocol the evidence for why the study endpoint is a likely surrogate. However, I'm afraid that there's widespread lack of understanding on this point; too; and I suspect that many protocols will simply provide evidence for a correlation between the surrogate and the clinical endpoint; which in my opinion is insufficient.
* Ideal surrogate markers to be used within pivotal efficacy studies should be validated upstream by regulatory agencies, such as the decline in eGFR which was accepted by FDA in a joint effort with the National Kidney Foundation dedicated working group--Such a validation could remain a nice to have however not a must have when surrogates are used at a proof-of-concept development phase
* It will be better to add a question about adding "surrogate outcome" in the title of abstract or not.
* A surrogate biomarker can be measured repeatedly over time; allowing assessments of the within subject change from baseline; which can add clinical and statistical value over and above assessment at a single timepoint post randomisation

**Other comments**

* Bravo! Delighted to see this--Keen to link this to OMERACT Surrogate outcomes Working Group through Robin Christensen
* No comments. Tx!
* Thanks for providing me with the opportunity to participate in this survey.
* Thank you for the invitation to be part of this Delphi exercise. I am very impressed by the thought and effort that has been expended to develop the candidate items in this Delphi exercise.
* Congratulations to the project leaders and collaborators. I look forward to the next round.
* It took me a bit of time to figure out that the outcome definitions were available by "mousing over". You could indicate that. For the public participants, I think it's important to understand that these items will be on a reporting guideline as this could easily be confused with list of components of the clinical trial. The glossary was helpful. Thank you for the opportunity to participate.
* The questions were clear in this survey.
* You might want to spell-check the instructions on the last page; i.e. "relavant" ;)
* Interesting survey though I found the wording of some of the questions difficult to interpret.
* You seem to have selected things that we would all broadly support. Some are; of course; context specific. As presented, this feels more like giving a seal of approval to the selected items rather than really exploring what are the core mission critical pieces of information.
* Really important work you are doing and thanks for opening it up to participate as well as for making it easy to do so. This was a great system to do this in.
* This is a welcome update. Some of the definitions for surrogates were a bit difficult to parse at first go. They would benefit from rewording for clarity (not removing information; just improving the wording).
* Beware of the likely variation in responses to this survey that you may receive as the questions all assume a sound understanding of the terminology and concepts involved in surrogate endpoint studies. I would question the benefits of making the survey appearing to cover comprehensively all aspects of such study; but in reality might be better to focus on patients' area of interest; that is; what they perceive to be personally relevant and meaningful outcomes. In other words; patients know best what they want in order to benefit or protect themselves or wider public; but less motivated to assimilate the complexity and nuances of surrogate endpoint studies as an academic field.
* Very clear and straightforward.
* see ref BMC Medical Research Methodology March 2012 (BSES) see feedback 1.
* Thanks for reaching out! Happy to provide additional assistance if helpful to your goals. Best--Chris
* Very clear but I think some of the terms might have been explained more clearly for a lay audience (assuming they are participants in this process). Some scale anchors don't match the anchors in narrative instructions. Good job though all around
* The rating scale for the question about completeness and inclusivity still had the headings for the prior question (e.g., "critical"); rather than those associated with completeness and inclusivity;
* Please extend the time-out period for the survey from 30 mins to 45 mins.

# Appendix 2

**Suggested additional items and responses**

Below are additional items and point by point responses to each. Generally, we added one item which is highlighted in green.

|  |  |
| --- | --- |
| **Outcome** | **Response by project team** |
| 1. Race and ethnicity
 | We interpreted this comment in two possible ways. First, race and ethnicity as population descriptors of reporting an RCT, which is not specific to trials of surrogate endpoints and therefore should be addressed in core reporting guidance. No update required. Nevertheless, we acknowledge that validation of surrogates needs to be generalisable by various factors that could include race/ethnicity. We therefore added ‘population’ to the list of criteria in consideration if evidence of surrogate validation is applicable to a particular RCT under SPIRIT 5 [M] and CONSORT 5 [M] item: *[Justification for selected surrogate: Evidence of being specific to setting used e.g., intervention, disease, population]* |
| 1. Healthcare setting
 | As above |
| 1. Gender (biological and otherwise)
 | As above |
| 1. A surrogate endpoint is a sign or measurement used in place of another; to see whether a treatment works. Some studies use surrogates to allow study results to be measured sooner - for example, measuring how much a tumour has shrunk in place of patient survival. They are not as accurate as other measures; but can be useful when a treatment needs to be approved more quickly
 | It was considered that the content of this proposed text is already covered by the definitions we tested in our 1st e-Delphi round (see first and fourth definition in rated definitions). |
| 1. Demonstration that the link between the clinical outcome is mechanist and not solely a correlation
 | Whilst we agree with the participant comment, we decided to have biological plausibility and statistical validation as one item as both are important elements of surrogate validation. Therefore, should the item go forward and be included in the final extension we will explain requirement of both elements in the elaboration and explanation document |
| 1. Give the mathematical or statistical relation that links the size of a treatment effect on the two outcomes (clinical/final and surrogate)
 | See response to point #5.  |
| 1. Report on the robustness of the above-mentioned link
 | Whilst we agree with the sentiment of the participant comment, this level of detail (e.g., prescription of the statistical method of validation) is beyond the scope of this reporting extension. Furthermore, there is currently there is no agreed consensus on this issue.  |
| 1. Measure the effectiveness of the treatment by evaluating the strength of the association with the clinical end point and the value of the latter in a risk-benefit balance
 | This issue is already addressed by the item on justifying the surrogate by providing evidence of validation (SPIRIT/CONSORT 4 candidate items) and risk-benefit balance by CONSORT 12: both received high ratings in e-Delphi round 1 – median score of 8 |
| 1. I think it would be useful to require trialists to collect and report the clinically relevant outcome; obviously this would be underpowered; but as it's what we're actually interested in; it seems odd to me to not collect + report it
 | Whilst we agree with this point, it is outside the scope of this reporting guideline and may be impractical to impose.We note that a related item related to this was not highly rated in e-Delphi Round 1 (median score of 7): *State if there are explicit to plans to extend follow up or conduct subsequent analyses/studies to verify benefit of current findings on the patient relevant final outcome [SPIRIT 9 and CONSORT 10]* |
| 1. Given that the methodological quality of studies validating surrogates is often low; it seems quite relevant for investigators to assess this within their own trial and report results
 |  As response to point #9 it is not practical to ask trial investigators to do this over and above reporting the trial. Furthermore, few individual trials are unlikely to have sufficient data to perform a robust within trial validation. Nevertheless, we would note that the item on trial data availability access to facilitate future surrogate validation was not highly rated in e-Delphi Round 1: median score of 7 |
| 1. Background/Discussion to include regulatory actions (approvals/refusals) based on the surrogate endpoint
 | We initially included a similar item as part of the drafting of items to go forward for the 1st round Delphi survey. However, after discussion with our advisory Executive Committee it was excluded because (1) regulatory guidance usually applies to only drugs and medical devices and not all interventions. Also, regulatory approval of surrogates, such as the FDA listing, is not always evidence based. |
| 1. Regarding the level of surrogacy validation: report whether strong evidence for patient-level and/or trial-level surrogacy is available
 | If SPIRIT/CONSORT 4 item on validation should make it to the final extension checklist, we will include this comment in our explanation and elaboration document as the optimal practice in reporting |
| 1. Regarding the level of surrogacy validation: report whether surrogacy is consistent across various subgroups of interest
 | As response in point #12 above as the suggestion links to SPIRIT and CONSORT 5 item: *Evidence of being specific to setting used e.g., intervention, disease, population* |
| 1. Regarding the level of surrogacy validation: what is the relationship of the surrogate endpoint to the causal pathways between treatments and disease outcome?
 | As response in points #12 and #5  |
| 1. Would suggest including other forms of surrogates and HOW they are measured such as Clinician Reported Outcomes; Performance Tests; Observer Reported Outcomes etc. See Walton M et al PMID 26409600
 | Our operational definition of a surrogate endpoint includes both biomarkers and intermediate outcomes that incorporates clinician reported outcomes/performance tests.  |
| 1. Methods: State what other surrogate endpoints were considered and why the current one(s) were chosen over those.
 | We include this proposal in Round 2 Delphi exercise and present it for rating |
| 1. Definitions: perhaps "within a reasonable context of use" may be a helpful addition (FDA uses the "context of use" term but not sure how important it is to account for the regulatory aspects of the definition). In a workshop last month, the phrase "constitutes a basis" was suggested for effect of S to predict effect on T **[An endpoint replacing a clinical endpoint that constitutes a basis for reliably predicting a treatment effect on the clinical endpoint in a defined context of use]**
 | Included for rating in Round 2 |
| 1. Justification of the selected surrogate ENDPOINT. Evidence for validation.
 | Already included as SPIRIT 4 and CONSORT 4 candidate items |
| 1. Justification of the selected surrogate ENDPOINT. Pathophysiologic and /or Epidemiologic and /or Therapeutic and /or Statistical and /or other scientific evidence of validation.
 | Already included in SPIRIT/CONSORT 4 candidate item and see our response to combining biological plausibility and statistical validation in point #5  |
| 1. ABSTRACT/INTRODUCTION The primary outcome is a marker endpoint(s) that is/are being considered (or has/have been considered) a surrogate endpoint for how a patient feels functions or survives (here put the specific feels/functions/survives endpoint)
 | We will consider this suggestion in writing the explanation and elaboration document |
| 1. A surrogate endpoint is a marker (laboratory; imaging; physiological or other biomarker; a physical or clinical sign or a composite of any of these) that is intended to substitute for how a patient feels; functions or survives based on epidemiologic; therapeutic; pathophysiologic; or other scientific evidence.
 | Already covered by definitions 1 and 2 that were assessed in round 1 e-Delphi. We note that neither of these definitions have not received a high rating – median score of 7  |
| 1. Provide a complete context for which the surrogate has been used (e.g., patient population; level of disease activity; when and how it is assessed [change from baseline; absolute value]. etc.}
 | The context of use is already covered by items SPIRIT/CONSORT 5: *Justification for selected surrogate: Evidence of being specific to setting used e.g., intervention, disease, population*The second point on assessment is not be specific to surrogates. We will seek to refer to this issue during writing of the explanation and elaboration document |
| 1. Provide details of the methodology for measurement of the surrogate including what value or threshold is clinically meaningful and how the surrogate information is interpreted.
 | See response to point #22We note that the item on predicted effect for rating did not get a high rating in round 1 (median score of 7): *CONSORT 11: Provide an estimate of the predicted effect based on the observed effect on the surrogate endpoint, and if not possible then a qualitative assessment*Furthermore, we also included the following item that covers this and has been rated highly in Round 1 (median score of 8)*: CONSORT 12: Interpretation of findings of the trial in the context of using a surrogate primary endpoint including its known validity and the potential benefit-risk ratio of the tested intervention for participants* |
| 1. Differentiation of FDA BEST-defined biomarkers and surrogate endpoints (imperative for US community)
 | Given the international reach of this project, we would not be seeking to make country specific recommendations.  |
| 1. I don't think I saw it captured that in certain situations; clinical outcomes can't be used because the trial would take too long for that event to happen clinically. So some type of surrogate measurement is needed to serve as a proxy. eGFR is a good example of a surrogate endpoint in nephrology clinical trials; as kidney function decline can be progressive and slow.
 | This could be included under the item on practical reasons of using a surrogate which was rated highly. When writing the elaboration and explanation document, we will seek to provide such examples. |
| 1. CONSORT [N]: State if there are known registered trials that will (further) examine the validity of the selected surrogate
 | This is beyond the scope of this reporting guideline. Furthermore, we note that the item on stating explicitly whether there are plans of follow up or subsequent analyses/studies to verify benefit on final outcome, but this was not highly rated in Round 1 (median score of 7): SPIRIT 9/CONSORT 10  |