Appendix 1

Items that have been modified are highlighted.

# 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
 |  |