# Free text feedback

Text highlighted in blue is from Delphi round 2.

**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** |
| Support for the item | * Important for transparency * Initially I did not consider that a preliminary statement would be necessary to introduce the subsequent item * Patient needs to be aware of this. * Should be clear that a surrogate outcome is being used * Rated higher in the benefit of transparency. The protocol should state that the primary outcome is a biomarker or intermediate outcome that is being considered a surrogate endpoint. * On reflection; patient relevant final outcomes that use surrogate endpoints should clearly be stated up from in the protocol |
| 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 * I am still a little uncertain about this one; as I feel there can be overlap in the interpretation of what a surrogate marker is; I changed my original rating due to my uncertainty so that my response was more in line with the average; For example using potentially inappropriate prescribing measures- some may argue these are process measures/markers of quality but others that they are surrogate measures as PIP is associated with drug related morbidity and is often not feasible to power studies to look for this. |
| 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. * I feel that S2/C2 incorporates/implies S1/C1 because if you report the PRFO that the surrogate endpoint is substituting and predicting for; well then you are also stating that the primary outcome is a surrogate endpoint. Perhaps S2/C2 could include a little extra (explicit) wording to better combine S1 and S2 (C1 and C2)? (which would mean that one of them does not need to be added) |
| 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** |
| 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 * Acknowledging the practical challenges of looking at a true endpoint and the way a surrogate endpoint helps is of value. * Without this, the reader may question the intentions of the researcher. * Information on the practical reasons seems to me rather interesting for the planning of subsequent research projects - I think I gave more importance to this aspect in the second round. * Critical as justifications should be made and informed in advance especially if a certain surrogate is not particularly well-accepted as validated in the literature (and for searchability purposes in research) * Seems a reasonable requirement for justification * This may encourage greater thought on whether it is necessary or appropriate to use a surrogate endpoint in the trial |
| Context of importance of item | * When the surrogate is used first time; yes; then likely we need to mention. However, when the surrogate is used in many other studies; it may not be that important. The first ranking was not accounting for the number of times the surrogate was already used for the indication. |
| 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** |
| Support for item | * Recent work on FDA Biomarker Qualification applications and Quantitative Imaging Biomarker Alliance have highlighted need for reporting rigor * It will be important for readers to understand why these outcomes were chosen * Whenever there is evidence of validation in a proposed surrogate outcome that should be provided; or otherwise stated that it is a putative surrogate in need of validation. * The endpoints selected should be fully justified as a validated outcome |
| 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** |
| 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 * This is vital for a study; since without this evidence; the study should not be performed. * Recent work on FDA Biomarker Qualification applications and Quantitative Imaging Biomarker Alliance have highlighted need for reporting rigor * It will be important for readers to understand why these outcomes were chosen * Readers would expect relevant surrogate endpoints for the setting when reading trial results |
| 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 * Rated higher as it is important to ensure that selected surrogate is being used within the setting in which it was developed. HOWEVER, I agree with other comments in that it is context dependent, and a more effective solution is to unify this item with item 4 as they are intimately linked. |
| 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 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** |
| Support for the item | * Important for understanding whether the trial is adequately powered * It is essential to know if the sample size is enough to find confident results on the surrogate endpoint * Recent work on FDA Biomarker Qualification applications and Quantitative Imaging Biomarker Alliance have highlighted need for reporting rigor * A degree of indication of the influence of surrogate validity on this calculation (especially very small sample sizes) would help inform reporting of study design and methods before the trial takes place * Sample size calculations should be fully transparent for CONSORT; and a detailed indication of the influence of surrogate validity on this calculation (especially very small sample sizes) would help; thus I upgraded to critical |
| 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) * although this may not be that important in general; there are really countries like Germany that need this. I had my general hat on last time. * I was basing this on the PFS surrogate for OS; but for new surrogates this would be more important. * Sample size calculations more relevant for protocol than final report * We should always report which outcome variable was part of the sample size and power considerations. (Independent of surrogacy) |
| 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 * Reconsidering... why not? But unfortunately I guess that there will be few cases where this happens |

**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** |
| 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 * Lastly reading comments the ethical issues here need to be addressed; Participants need to know what they are signing up for. The basic fact that the trial measures a lab test and WILL NOT be able to tell if the intervention helps them or anyone else is a major piece of information that could influence whether people choose to participate. Merely stating "people don't understand" is insufficient and calls into question whether consent was informed at all. * To reflect better patient engagement * Importance of letting patients know what is being evaluated slightly more important than previously accounted for * Would encourage greater ethics and transparency; allow patients to make an informed choice about participating in a given trial * ethical consideration. Patients need to know the purpose of the study. |
| 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. * It may be important; but still feel that it should not be mandatory. * Seems important but not critical in a report of a clinical trial because unlikely to affect interpretation of the findings * While it is important to share a clinical trial's goals with participants; the knowledge of where or not a surrogate is a goal of the trial does not impact potential risks to patients who choose to participate. |
| 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. * On reflection; I think most trial participants would be confused with the concept of surrogate endpoints |
| 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** |
| 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 * changed my views - while not yet possible in complex interventions of lifestyle change; this is definitely an aspiration * On rethinking I feel that only reporting the composite outcome can hide how the importance of the various components. * Realised that it would be important to know which of the components was responsible for the change as this could affect interpretation of results. * Greater transparency. In some cases; for some components; a relatively fewer number of measures may have been possible, but it is still important for this information to be appropriately presented. |
| 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 * Feels like further work but for not much gain |
| 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** |
| 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. * Originally rated 6 because most trials are not powered to detect all potential harms; but it is important to know whether this is the case * It does seem important to know whether or not the trial was designed to detect harms * To integrate few safety trials that are designed using a surrogate * The sample size and follow-up period must be sufficient to adequate report patient safety |
| 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 * although this may not be that important in general; there are really countries like Germany that need this. I had my general hat on last time. * Again I based original on PFS; this would be important for shorter term surrogates. |
| 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 * Important to include in discussion that this as an inherent limitation when using surrogates as end points. |

**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** |
| Support for the item | * follow-up plans is important * Including this in final report will help inform interpretation of findings * For low mortality risk studies; the relation between surrogate and late outcome seems to only concern the limited follow-up of RCTs; while for those studies the death percentage can be as low as 10%. Seeing such a study changed my mind. * Greater emphasis on 'patient relevant' aspects |
| Context of importance | * I think it can be important for some of chronic diseases (e.g. cardiovascular diseases or rheumatologic diseases) to have a plan for probable extension of follow-up time. |
| 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 * It is worth while to get it into the DISCUSSION section of the paper |
| Limitation of item | * Not sure of the value of a non-binding intention * we usually have plans - but they often go awry * It's a best practise to have a clinical assessment plan we all laid out a priori; but many a times it is not feasible; as we would want to see the preliminary/interim efficacy and safety first and then plan out for future trials/ extend the follow up * I think this scenario is unlikely to occur; and it would not be an issue as this would only further benefit the 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*

|  |  |
| --- | --- |
| **Theme** | **Free text comment** |
| 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 * qualitative research is important * Thought to give a little more weight; also considering the potential relevance of a qualitative assessment * Increases the reproducibility * From stats perspective this may not be as important. However; for the patient it is as they want to know what the treatment does mean for them. I learned that during the past months because of preparation of a presentation. * Important to be able to predict how the surrogate outcome will perform to facilitate future research into validity of surrogate. * It can be important to provide more proper methods and interpretation for our trial to be more generalizable trial. * Greater emphasis on 'patient relevant' aspects * I have come to believe this is a more critical element of the process to enhance transparency * Nonetheless; if data is reliable and available; then it will be important to increase the effect size of the inferences of the study. * I think this is crucial for future research (simulation studies, consistency in replication of studies etc). Having a collection/repository similar to ICECaP, for instance, across different therapies and modalities would be very useful going forward |
| 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 |
| 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... * C11: Potentially problematic; may introduce substantial bias; may result in undue importance being placed on unfounded statements. * Although this is important; data may not always be readily available. Also; the association between the surrogate endpoint and the clinical endpoint may be moderated by specific variables. |
| Modification to item | * suggestion editing: “Provide an estimate of the predicted effect of patient relevant outcome based on the observed effect ..." |

**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** |
| Support for the item | * Increases the reproducibility * Since the ultimate goal is to assess the benefits and harms; it did not seem right to me to rate the item as "not important" * I have come to believe this is a more critical element of the process to enhance transparency |
| Other comments | * 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** |
| 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 * open access is key for improving the quality in research; I have thus revised my previous judgement and proposed a higher rate for this specific point * fair data principles are important * Transparency in clinical trials is important; and in order to replicate studies I have reconsidered my previous rating * this may lead to more open science activities that could help revisit surrogate endpoints debate. * I have come to believe this is a more critical element of the process to enhance transparency |
| 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 * will take time to implement - in CVD only now seeing such analyses on major studies * same thought; it would be good to have this stated in the analysis plan but might not be feasible/evident unless we know the preliminary performance of the drug in the patients. |
| 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** |
| 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 * Surrogates are not just 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 * I like a more general definition that does not go into details such as biomarker or a laboratory finding; as it may include other endpoints; like rating scale. I suggest to leave a very general definition but provide examples such as these. * I object to the premise and the options given making it hard to score anything: Surrogate endpoints can only measure benefit = not harms and half of identified failures of surrogates are of “validated” endpoints such as blood sugar and cholesterol level * I think it is the word "predict" that I am uncomfortable with; ie; part of our validation definition is based on meta-regression 1-to-1: Effect Size1 vs Effect Size2 * logged me out - does not have feels functions or survives in definition * because "biomarker" does not represent studies that use other kind of response measures as a subjective questionnaire. |
| 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** |
| 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 * What about surrogates that are not lab measures or physical signs? What about mental illnesses or neurological conditions? There may be MRI scans etc used for such conditions but presumably non-physical signs are also used? * Went through comments from others. What about surrogates that are not lab measures or physical signs? What about mental illnesses or neurological conditions? There may be MRI scans etc but presumably non-physical signs are also used? * I do not feel that any definitions could be described as inclusive |
| 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** |
| 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 * If the whole purpose of a definition is to increase understanding of a term using terms that are commonly understood and whose meanings are clear; this definition completely fails to achieve that. It is not likely to be commonly or easily understood. All medics won't even understand it. * I don’t feel strongly about it (Poor reading – too long; no need for the “null hypothesis” statement in it) * A clinical definition is more appropriate * Reading the comments made me to understand better the statement. I believe that linking the hypothesis testing in the setting is a spurious exercise. Much useful to pride with the equation (stat or algebraic) that enables to predict the size of the benefit on the clinical outcome from the observed difference on the surrogate * Too convoluted * Too much statistical... * too many technical words * I agree that the wording is unclear; in particular; the terminology of "response variable" may be too vague and there is no explicit mention of the predictive ability of a candidate surrogate endpoint * This is too statistical and might not be relevant to wider population (patients; healthcare practitioners). |
| 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** |
| Support for item | * A comprehensive definition * endpoint is generic and it is different in kind form feels functions and survives; I did not rate it highly in first round as it stated clinical' trials' and I would have preferred clinical 'studies' so as to not exclude observation studies. so our these only to be used in intervention studies; if so then this definition is my preferred definition of all. It also includes types of evidence |
| 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". * The definition becomes more relevant if "is expected to predict" is modified to "has been shown to reliably predict" as suggested by a Delphi participant; and I would then give it a higher rating.. |
| 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** |
| 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. * Great that you have referred to both biomarkers and intermediate outcomes in the definition. What about referring to context/setting as well? Concerns about "reliably predicts" - what about surrogates (that are used or considered surrogates) but have low validity? I know I am wrongly mixing reliability and validity BUT my point is: are all surrogates only considered to be surrogates if they are 100% known to reliably predict PRFO? The definition seems incomplete - do you want to mention "clinical trials" and "treatment" (or similar?). i.e.; A biomarker or intermediate outcome used to substitute for a patient or participant relevant final outcome (i.e.;...) in clinical trials and reliably predicts benefit or harm of treatment based on ... |
| Limitations | * As stated previously; the readers of medical journals are often clinicians with little to no familiarity with these terms. “Intermediate endpoint” is meaningless to practicing clinicians. And “predict” confuses the major issue in surrogacy which is confusing patient level correlation and trial level surrogacy. So this definition is not helpful to many if not most readers * participant relevant final outcome (i.e.; severe morbidity; health related quality of life or mortality) is a rating of 'degree' and not change 'kind' * Other definitions are better in defining and clarifying the essential features of a surrogate endpoint. * A doubt I had was on the use of the word “final” in patient-relevant final outcome – there may be non-final clinical outcomes like functional related measurements that may be sufficient, and HRQoL isn’t necessarily a final outcome (e.g. EUHTAnet document on clinical endpoints) |
| 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 * I agree with the addition of intermediate outcomes/endpoints since biomarkers do not necessarily capture subjective elements of health status. |

**General comments**

**Definitions**

* To repeat; surrogates can never be endpoints. Of clinical endpoints; I only know two: anus and death.
* As an aside, it would be informative to clarify the (consistency of) usage of outcome vs endpoint, these may be synonymous to some but not to others
* Transparency of choices; definitions and reporting of surrogate endpoint is critical
* 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.
* For the definitions, I think there is still scope to improve. There are elements of some ranked poorly; that are good or could be improved on. I think the key message to get across is that they are a measure that accurately predicts a gold standard clinical measure. But clinical measure also then needs to be defined.
* My preferred definition is “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.” It is parsimonious (an endpoint) If you remove in clinical trials then it is general enough for all clinical research; but that is not a deal-breaker if this statement applies only to clinical trials---- it has face validity (non-medical people know 'feels; function; survives')----it has content validity in specifying the type of evidence for validation (epidemiologic; therapeutic; pathophysiologic; or other scientific evidence)----finally it is almost identical to the original NIH definition/s from 20 + years ago except 'endpoint' is used instead of 'biomarker' and combines the 'biomarker'; 'surrogate endpoint' and 'clinical endpoint' definitions. Therefore, it also has historical / contextual traction

**Extension items**

* For systematic review/literature searchability purposes, it would be good to include the word “surrogate” in the title/abstract – many articles don’t mention this/use “marker””
* Many items are aspirational in my area of work (lifestyle; obesity; diabetes) and have feasibility issues for implementation. Could the use of these be piloted in different areas (mental health; counselling vs. drugs; etc.) before being adopted? Room in papers to discuss the underlying issues would be most welcome.
* I still fear that a more formal definition of a surrogate is lacking. The missing arm is the few critical details on what should be drawn from the biological and epidemiological evidence to support surrogacy: 2 dimensions; one is statistical; the correlation between the potential surrogate and the clinical outcome; the other one point that should not be missed is the equation (or any algorithm) that enables to predict the quantity of benefit. Missing the latter bullet point leaves surrogacy demonstration half the way.--Another issue that is not addressed at all is the consistency of surrogacy through various classes of treatment targeting the same disease/outcome. This is a difficult issue. A surrogate which the surrogacy of which is demonstrated for a single class of treatment is of little use for R&D.
* 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**

* The instructions for the use of these checklists need to again clearly state; that the use of patient-relevant outcomes must always be given priority. That even a neat processing of the checklists to justify the use of surrogate parameters does not release from the ultimate necessity of validation (surrogate and corresponding patient relevant outcome). The scientific community must continue to adhere to this and generate instructions and examples in order to make progress in this regard - to support validation activities, thank you for your effort!
* 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
* Many thanks for inviting me and including in your survey; and I'd like to congratulate you for conducting such an important survey to improve RCT reports.
* These are important items and additions!
* Helpful to see how I scored things as well as others.
* A bit tricky asking for detailed reasons for changes at end of document; better placed proximate to the where the changes occurred
* I object to the options given - patients and doctors should be informed that surrogate endpoints can ONLY measure benefit - not harms and that half of identified surrogate failures --(that do not translate to clinical benefit) are of so-called "validated" surrogates; such as blood sugar and cholesterol levels. There needs to be much greater emphasis on the need to --seek genuine patient-oriented clinical benefit. Cancer drugs lead the way for bad outcomes based on various surrogate endpoints such as progression free survival or tumour size measures.
* One comment in general as a practicing clinician AND a researcher who deals with my practicing colleagues. Many of the CONSORT documents are written in a way that sounds like what my colleagues call "statistical-eze". Since the readers of journals are often practicing clinicians (and increasingly the lay press with even less familiarity with these issues and terms) would be key to write these documents in a way that makes them accessible to both statisticians; investigator/authors of manuscripts as well as practicing clinicians. --Since these issues have been discussed for over 30 years one cannot help but hypothesize that some of this is dissembling on the part of those writing manuscripts to make it unclear what was measured; WHY it was measured (eg don't need a surrogate in an acute disease where patient outcomes are measured in the same time frame); how it is measured and analyzed; and the relationship between surrogate and patient outcomes (eg often stated in HIV trials that "viral load is primary outcome' but this is incorrect as any patient outcomes that happen like death or AIDS defining events during the course of the trial should override outcome on the surrogate - therefore outcome is and should be a composite of patient outcomes and viral load). --This is a wonderful effort as the definitions in papers are often unclear and often obscure where a surrogate is used at all; how it is defined; why it was used etc. And "regulatory precedent" is not necessarily based on data -just someone at regulatory agency agreed with or without evidence.
* Excellent work aimed at providing quali-quantitative justification for the choice of surrogate outcomes
* This is the first time (the survey) I viewed these items from both the protocol and scientific report perspectives. I realized that much of what is in the protocol should also be in the scientific report.
* Thank you for conducting this valuable work in benefit of the research community!
* I would like to see the outcome/ know the next step for this. All the best!
* 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
* Thank you again for your interest in this important topic! In case you were not aware; here is a link for FDA resources regarding surrogate endpoints: https://www.fda.gov/drugs/development-resources/surrogate-endpoint-resources-drug-and-biologic-development. Included are links to additional information including: Surrogate Endpoint Table that is updated twice a year to help sponsors know which surrogates are supported by FDA and for which type of approval. A content outline for Type C Novel Surrogate IND meetings that describes the key topics that should be addressed when proposing a novel surrogate endpoint----Thanks!--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.
* The items and the survey in general are very well planned.
* Thank you for driving this important piece of work.
* Good work overall!
* results so far looking good