BMJ-2021-064952

Dear Dr. Doshi,

Thank you for the opportunity to submit a revised version of our manuscript (BMJ-2021-064952). We have carefully reviewed the comments from the Reviewers and Editors. As we describe below in our point-by-point response, we have addressed their comments in full. We believe that doing so has further strengthened our paper.

Thank you for considering our manuscript for publication in *The BMJ*. We look forward to your decision.

Kind regards,

Huseyin Naci
Associate Professor of Health Policy
Department of Health Policy
London School of Economics and Political Science
Email: H.Naci@lse.ac.uk

Cc: **Dalia Dawoud, Oriana Ciani, Sylwia Bujkiewicz**

Editors' Comments:

1-Our overall feeling was that we'd like to offer a chance for a major revision, but we had a lot of trouble with the bit about NICE as it threw us while reading and felt like an abrupt ending. Essentially, it seems like either the NICE guidance becomes the front and center focus of the article, in which case we need to know a lot more about it (what it says and why it matters), or alternatively, it would better to take it out. We suspect that not much can be said about NICE at the moment, so the latter approach of taking it out makes more sense. In that case, you probably need to find a better hook to make the piece feel topical.
This was our only major comment - there are a few more minor comments - and a lot of helpful feedback from the reviewers that you will find below.

Many thanks for this suggestion. To fully address this comment, we have removed the section focusing on NICE, and instead listed these proposed changes to NICE's Methods Guide in Box 1, as also recommended by the second reviewer.

In our revised paper, we have used the recent FDA approval of aducanumab based on a surrogate endpoint as the 'hook'. The introduction paragraph of the revised paper now reads as follows (page 2, line 39):

> *"On 7 June 2021, the US Food and Drug Administration (FDA) granted accelerated approval to aducanumab for the treatment of Alzheimer's Disease. The FDA based its decision on the drug's amyloid-reducing effects despite evidence from several earlier studies that shrinkage of beta-amyloid protein plaques does not predictably delay cognitive impairment in patients. [1]The decision has drawn significant attention to the use of surrogate endpoints —laboratory values, radiographic images, or other physical measures*

that may serve as indicators of clinical outcomes such as symptom control or mortality— in clinical trials of new drugs.[2] In fact, the approval of aducanumab is only the latest example of growing regulatory reliance on surrogate endpoints. Using surrogate endpoints in drug trials has increased substantially in recent years.[3]"

2. The details on validations methods will likely not be terribly engaging for the majority of readers. Instead, it might be more useful to spend more time explaining an aspect of the argument that is not likely obvious at first blush like why a surrogate endpoint can be considered valid for some drugs but not for other drugs.

Thank you for this suggestion. We have substantially shortened this section of the article and instead added further detail explaining why a surrogate endpoint can be considered valid for some drugs but not for other drugs. This section of the revised paper now reads as follows (page 5, line 164):

> *"Evidence at the individual patient level alone is insufficient to evaluate surrogate endpoints especially when such evidence is obtained from a single trial.[38] This is because the observed surrogate-to-clinical outcome relationship for one drug may not hold for another. Fleming and DeMets describe a number of scenarios where, depending on the mechanism of action of a treatment, a surrogate may not fully mediate the treatment effect on the clinical outcome, as different interventions may affect disease pathways in different ways, involving sometimes one endpoint but not the other. [38] Meta-analysis, which combines data from a number of randomised trials, is therefore more appropriate for evaluating the association between the treatment effects on the candidate surrogate endpoint and on the final clinical outcome.[39]"*

3. Would be nice to have a little more scaffolding for readers not fully engaged in this area: what led to surrogates becoming acceptable and more widely used in the approval process? Is it driven by the pharmaceutical companies seeking faster routes to market? Equally if the regulatory bodies said surrogate endpoints weren't acceptable, then the nature of trials would surely have to change? Authors highlight that surrogate endpoints are used in the majority of expedited programs - is there an alternative to using surrogate endpoints in this context that wouldn't increase the time to market?

We have now added further details about the regulatory acceptance of surrogate endpoints. For example, on page 4, line 5, we write the following:

> *"The recent proliferation of surrogate endpoints is partly due to the increase in the use of 'expedited' regulatory programs that are aimed at speeding up the development, review, and approval of drugs.[22] Over the past quarter century, lobbying by pharmaceutical companies has put pressure on policymakers to establish several expedited programs in Europe and the United States.[23] These programs also meet perceived patient demand for faster access to potentially effective therapies."*

To address the 2nd part of this comment, we have added a new section to outline the potential implications of raising the bar for using surrogates by regulators (page 6, line 196):

> *"Raising the bar for the regulatory use of surrogate endpoints may increase the cost and duration of drug development. In cancer, trials collecting data on clinical outcomes lasted on average 11 months longer than those measuring surrogate endpoints.[5] However, this need not hamper pharmaceutical innovation. In the past, regulatory guidance encouraging manufacturers to evaluate the cardiovascular outcomes of anti-diabetic medications*

incentivised the generation of patient-centred evidence without adversely affecting research and development.[49,50]"

4. "…sound methods for surrogate validation" - sound methods according to who? What standard are the authors trying to hold the regulators to?

We no longer use the term "sound methods" in our revised paper. On page 5, line 149, we have amended the sentence as:

> *"The level of depth and scrutiny applied by different agencies varied, with a relatively infrequent consideration of the level of available evidence and application of statistical methods for surrogate validation."*

5. is a surrogate endpoint the same as a validated predictor?

Surrogate endpoints, if valid, are predictors of clinical benefit – this is the definition we have used consistently in our revised paper, including in the Standfirst. In more technical terms, this refers to the treatment effect on the surrogate endpoint as predictor of treatment effect on the final outcome. We have used this wording in the revised 'Validating surrogate endpoints' section (page 6, line 171):

> *"Meta-analysis, which combines data from a number of randomised trials, is therefore more appropriate for evaluating the association between the treatment effects on the candidate surrogate endpoint and on the final clinical outcome."*

6. Your article seems largely accepting of surrogate measures, with the tack that we just need to improve statistical methods. But aren't there clinical scenarios where surrogate endpoints would not be acceptable e.g. acute, short-lived, self-limiting infections? Why would we need surrogates for drugs to treat uncomplicated skin infections? Perhaps you can acknowledge this in the part where you set the scene, so the scope is clearer.

Thanks for the suggestion. We have now clarified that surrogate endpoints are neither necessary nor appropriate to use when evaluating drug effects in acute conditions. We have emphasised this in the section outlining the benefits and harms of using surrogates (page 3, lines 69-71):

> *"By contrast, surrogate endpoints are not useful when a drug's effect on the final clinical outcome can be observed within a relatively short time frame, e.g., in acute conditions.[8]"*

We also repeated this in a newly included section entitled "Way Forward" where we state on page 6, lines 190-193:

> *"Using surrogate endpoints should only be considered in chronic disease settings with substantial unmet need, especially when collecting data on clinical outcomes requires clinical trials with unfeasibly long follow up durations."*

7. When revising, please ensure the article is within 1800-2000 words.

The word count now is around 2100. We hope this is still acceptable.

Reviewer(s)' Comments to Author:

Reviewer: 1

Generally, this article is timely and of great importance. The article's title makes the case that the science of surrogate endpoint markers has advanced and seeks to showcase NICE's efforts to address that emerging science. But after some appropriate review and setup, the reader arrives at p. 9 lines 4-10 (for the reviews) and p. 9, lines 27-33 (for the proposals) expecting some details on precisely what NICE has found and is proposing but finds instead some generalities implying that the authors support a more restrictive approach to surrogate markers, but little confirmation that this is so. Clearly, the authors are treading a difficult line between too much and too little information, but the current approach leaves the reader unsatisfied. Perhaps tables providing more detail would help.

Thanks for this suggestion. As this comment closely parallels the Editors' comments, we have now removed this section of the paper and instead included the details of the proposed changes (as well as their rationale) in Box 1. We hope this strikes the optimal balance.

There is substantial relevant evidence missing from the article in its current draft. There is a large body of papers addressing this subject, but the authors need to do more to convince the reader that their review was systematic.

We can confirm that we have cited some of the key systematic reviews that summarised the individual studies evaluating the validity of surrogate endpoints (e.g., those in oncology)., However, our article is not a systematic review of the literature on this topic. Our aim is to provide a balanced and evidence-based overview of the key issues on the use of surrogate endpoints by regulators and health technology assessment bodies, and highlighting some key methodological developments in this area.

Pg. Line Comment

4 5/6 The authors' statement is true but doesn't fully capture regulatory agencies' ability to require Phase 4 post marketing studies.

This sentence has been revised to further clarify this point and now reads (page 4, line 116):

> *"Not all expedited programs explicitly facilitate the use of surrogate endpoints. However, reviews of recent approvals have shown that the use of surrogate endpoints is common across all expedited programs."*

Later on in the paper, we further clarify this point (page 4, line 122):

> *"The use of surrogate endpoints in certain expedited regulatory programs like the FDA's accelerated approval pathway is linked to "conditional" approvals where drug manufacturers are legally mandated to conduct additional trials to demonstrate the clinical benefit of their products."*

4 15/16 Can the authors point to systematic efforts to assess the frequency of unconfirmed surrogates? See, e.g., https://pubmed.ncbi.nlm.nih.gov/31135808/

We can confirm that this is already cited in our paper (reference 28) to support this statement (page 4, line 124):

"Even when post-approval studies are required, clinical efficacy of drugs initially approved on the basis of surrogate endpoints is often subsequently "confirmed" on the basis of other surrogate endpoints."

We have also added references to other studies that systematically assessed the frequency of unconfirmed surrogates in regulatory submissions (page 4, line 128):

"Unfortunately, surrogate endpoints accepted by regulators are not routinely validated. In a recent study, researchers found only weak correlations between surrogate endpoints and survival in breast cancer.[28] In another analysis, researchers found that none of the surrogate endpoints used in EMA expedited approvals were validated.[29] "

4 31 Do you mean that that surrogate markers were the sole basis for approval?

Yes, we have clarified this by stating (page 3, line 99):

"Pivotal studies that form the basis of regulatory approvals for new drugs frequently use surrogate endpoints alone."

4 51-52 This description does not accurately convey the nature of expedited programs, at least in the US. "Expedited programs" is an umbrella term that includes priority review, fast track, breakthrough therapy and accelerated approval. These are described in detail here: https://www.fda.gov/media/86377/download. Only the last of these inherently involves surrogate markers and then sponsors are supposed to conduct Phase IV studies that document the impact of the drug upon non-surrogate outcomes.

We have further clarified this statement, which now reads as follows (page 4, line 116):

"Not all expedited programs explicitly facilitate the use of surrogate endpoints. However, reviews of recent approvals have shown that the use of surrogate endpoints is common across all expedited programs."

5 7/8 Reference needed for "In 2018, ¾ of new drugs approved by FDA benefitted from at least one expedited program…."

The reference has been added.

5 37/38 Please provide reference for 2018 FDA table

We no longer refer to this table.

5 35 Recommend deleting sentence as topic sentence does not align will with rest of paragraph

This has been deleted.

5 42-7 Can the authors comment on what to them constitutes satisfactory validation?

The revised section titled "Validating surrogate endpoints" provides a balanced discussion of the history of surrogate validation efforts, different frameworks, and emerging consensus

for the use of meta-analysis. We no longer refer to "satisfactory validation" but instead state (page 6, line 174):

> *"There is growing methodological consensus for using bivariate meta-analysis methods to evaluate the surrogate-to-final relationships. [40–44]"*

6 20-26 Are the authors aware of any attempts to determine what percentage of surrogates used to approve drugs actually meet the IQWiG threshold?

In an earlier study (which is cited in our revised paper), some of the authors of this Analysis article compared the application of three surrogate evaluation frameworks (i.e, Elston&Taylor, IQWIG, Biomarker-Surrogacy Evaluation Schema (BSES3)) to 31 different surrogate validation meta-analyses in advanced cancer settings. As we mention in our revised paper, the study concluded that the frameworks lead to potentially heterogenous conclusions about the validity of the same putative surrogate endpoints across different jurisdictions. According to IQWiG's framework, progression-free survival achieved acceptable evidence of surrogacy only in metastatic colorectal and ovarian cancer treated with cytotoxic agents.

A more recent comparison of HTA assessments using surrogate endpoints (which is also cited in our revised paper) revealed that IQWIG rejected the validity of the candidate surrogate in 9 out of 13 reports, and in the remainder never declared acceptability.

6 35/36 Expand on what the variations are

We have now clarified what the "variations" referred to (page 5, line 149):

> *"The level of depth and scrutiny applied by different organisations varied, with an infrequent consideration of the level of available evidence and application of statistical methods for surrogate validation.[34] Such variation across HTA bodies yields potentially heterogenous conclusions about the validity of the same putative surrogate endpoints across different settings.[35]"*

6 51/52 Expand on greater transparency for NICE; what is the transparency for others?

After a revision in the related section, we no longer refer to this aspect. This section of the paper now reads as follows (page 5, line 149):

> *"The level of depth and scrutiny applied by different organisations varied, with an infrequent consideration of the level of available evidence and application of statistical methods for surrogate validation".*

8 10-17 Brought up and included in previous page (lines 42-7)

We have considerably revised the "Validating surrogate endpoints" section of the paper, as recommended by the Editors. This suggestion is therefore no longer applicable.

8 10-27 Authors could consider a table describing all the methodologies for establishing the predictive value of a surrogate marker

All the methods we mention (bivariate meta-analytic methods) use a similar mechanism for establishing the predictive value of a surrogate marker. Following the advice from the Editors, we decided not to include more technical details, but we do highlight this in our revised paper (page 6, line 174):

> *"There is growing methodological consensus for using bivariate meta-analysis methods to evaluate the surrogate-to-final relationships. [40–44] These methods take into account not only the correlation between the treatment effects, but also uncertainty. The correlation between treatment effects on the surrogate endpoint and the final outcome, or a related measure of association between the correlated effects is needed to quantify the surrogate relationship.[43,44]"*

8 30/31 "Proposed Changes for surrogate of health technology evaluation". This leads to a summary of the reviews in lines 3-10 on p. 9, but this is at a very high level. Can the authors produce a table that would provide more detail?

In response to Editors' comments, we have now removed this section and instead included a detailed list of the proposed changes by NICE and their rationale in Box 1.

9 33/34 Summarize the proposals here in another table would give more granularity to the article; we do not know from this description what these proposals are

A list of the proposed changes are now included in Box 1.

10 12-36 Key Messages are too generalized

We have revised these to make them more specific.

16 Box 1 Could be deleted; not well conceptualized. Critical points can be incorporated into the text.

This has been done. We have also moved some of the text from the box to the section on "Validating surrogate endpoints" (page 6, line 174).

Reviewer: 2

This is a well written manuscript on an important topic that is applicable across diseases and conditions. The authors lay out their arguments for the necessity of stricter requirements for the use of surrogate endpoints in regulatory trials of drugs or technologies. The discussion is balanced and clear.

Thanks for these positive remarks.

What I am struggling with is that it is still unclear what the authors' specific asks are – they mention that "these proposals were included in the consultation on the changes to NICE technology evaluation methods that concluded in December 2020" but the decisions won't be made until September 2021 when the program manual is published. I think the box is a way to clarify this, but as currently written, the box has a lot of text without listing specific criteria. Could the authors make this clearer? How about a box that has current criteria on the

left and proposed criteria (specifically delineated) on the right?

We have undertaken a major revision of our paper to clarify the main crux of our article, which argues for raising the evidence bar for accepting surrogate endpoints for regulatory and health technology assessment agency decisions. This is now clearer in the title of the paper, standfirst, key messages, and throughout the main body of the text.

We have also fully taken on board the Reviewer's suggestion to add a box on NICE's current criteria and the proposed criteria.

Specific Comments:
Event-free survival may not be the best example of a surrogate endpoint. Event-free survival is typically defined as the time the patient remains free of certain complications or events that the treatment was intended to prevent or delay. These events may include the return of the cancer or the onset of certain symptoms, such as bone pain from cancer that has spread to the bone. Although event-free survival may be based on a surrogate, such as radiographic appearance of the cancer, it doesn't have to be. It may be based on patient-important outcomes of pain or symptom control. Better examples of surrogate endpoints could be the laboratory markers in diabetes trials, such as urine microalbumin or glycemic control levels as measured by HbA1c. A table with examples of surrogate and patient important outcomes may be helpful to the readers. Overall survival is important but not always the most important for all patients in all stages of life.

Thanks for this very helpful suggestion. We have now added a new table which lists examples of surrogates validated using meta-analysis of RCTs. Our reference to event-free survival was not accurate, and we have therefore corrected this by referring to disease-free survival instead.

Authors argue that the main benefit of surrogate markers is the speed of the regulatory approval process but – to be fair – it should be mentioned that the use of patient-important outcomes may simply not be feasible in some situations. These situations may include the long time to the actual outcome (e.g., in diabetes, blindness or renal failure) or studies or rare or ultra-rare disorders (in which power cannot be achieved yet progress needs to be made).

We thank the Reviewer for these thoughtful suggestions. We have amended the paper to fully address these comments. We now write on page 2, line 63:

> *"Using surrogate endpoints can reduce the duration, cost, and complexity of clinical trials prior to regulatory assessment, and facilitate faster patient access to new therapies.[3] Surrogate endpoints are useful in chronic disease settings when they can provide early and accurate measurement of a drug's effect, especially when long follow-up is required before the final clinical outcome can be assessed.[4]"*

We revisit this issue later on in the paper, in the newly-added "Way forward" section (page 6, line 190):

> *"Using surrogate endpoints should only be considered in chronic disease settings with substantial unmet need, especially when collecting data on clinical outcomes requires clinical trials with unfeasibly long follow up durations."*

One issue with surrogate endpoints that the authors point out is the lack of validation and lack of agreement on the frameworks for validation of surrogate endpoints. Although the authors describe some of these frameworks and methods for assessing validity of surrogate endpoints, there is an inherent problem with this. Even when surrogate endpoints are highly predictive of the outcomes, they may not be predictive of treatment effects. Although authors mention ways to determine the predictive ability of surrogates for treatment effects, applying this across treatments is inherently problematic. Different treatments (including different agents within a drug class) may have unintended consequences or may work through pathways unrelated to the measured surrogates. Although validating surrogates is important, I don't think they can 100% ensure that surrogate endpoints and outcomes will align in any given trial.

We fully agree with the Reviewer about the importance of this issue. To address this comment, we have amended the language in the "Validating surrogate endpoints" section in two ways.

First, we have clarified why the surrogate-to-clinical outcome relationship for one drug may not hold for another (page 5, line 164):

> *"Fleming and DeMets describe a number of scenarios where, depending on the mechanism of action of a treatment, a surrogate may not fully mediate the treatment effect on the clinical outcome, as different interventions may affect disease pathways in different ways, involving sometimes one endpoint but not the other. [38]"*

Second, we have clarified the importance of estimating the uncertainty of the surrogate-to-clinical outcome relationship in meta-analysis (page 7, line 216):

> *"If recommended meta-analysis methods are used for validation, the strength (or weakness) of the surrogate will be reflected in the uncertainty around the predicted treatment effect on the final outcome. A weaker surrogate will yield a larger interval and hence greater uncertainty."*

Authors mention many European decision-making bodies but do not mention the U.S. FDA or other non-European drug regulatory bodies. This may be included for completeness.

Many thanks for this suggestion. We have further emphasized several examples from the US FDA context throughout the paper. In particular:

- *Page 2, line 39 (introduction paragraph)*
- *Page 4, paragraphs 2, 3 and 4 (section on "Use of surrogate endpoints by regulators"*

Reviewer: 3

The title of the paper is somewhat disingenuous. Its focus is on HTA methodology. The authors take a quite narrow view of the role of surrogate endpoints and describe a 'case for change' which they appear to be already actively embarked upon with NICE. There is no up-to-date reporting of what steps regulators such as EMA and FDA are taking with regard to surrogate endpoints. The 'case for change' needs to be made against the background of a

wider understanding of clinical research, both as it currently is and the directions in which it is heading.

We are grateful to the Reviewer for their thoughtful and detailed assessment of our paper.

Starting with the revised title and standfirst, we hope that our revised article clarifies that we are no longer focusing on HTA methodology.

We have also removed the section focusing on NICE to address the Editors' and Reviewers' comments. This further ensures that the article strikes a balance between regulatory and HTA use of surrogate endpoints.

Although the decision is the Editors' I feel that for a journal like BMJ this paper needs to reflect a wider background and identify the place this work has in that environment. However a paper like this is good for the BMJ's audience. I would like to see a Plain Language Summary added to it.

Although Plain Language summaries do not appear to routinely accompany Analysis articles, we would value the Editors' guidance on how to proceed in addressing this suggestion.

Having made that criticism I enjoyed reading this paper, and I am not a methodologist. It opened up a rare insight into the world of methodology and statistical analysis.

Thanks for this positive remark.

It has a view of a real-world issue but the lack of context means that the paper is unbalanced, lacking a rounded summary of the views of (and work being undertaken by) other stakeholders.

To fully address this comment, we have provided further context (or "scaffolding" as suggested by the Editors) throughout the article.

The authors do indicate that acceptance of uncertainty (properly contextualised) is appropriate and mention a couple of times that on occasions 'flexibility' is required in decision making. As a patient involved in NICE Technology Appraisals I have seen first-hand such flexibility at work (NICE TAs 179 and 185) and I enjoyed the convoluted language employed to justify it, albeit that was some time ago.

We thank the Reviewer for this comment and for reference to real world examples of appraisals based on surrogate endpoints. Although we endorse uptake of robust methods for surrogate validation and a more structured and more standardized consideration of this issue in healthcare decision-making, we are aware there might be situations where these methods are not feasible, e.g. rare/ultra-rare conditions, high unmet need. In those instances, flexibility around the criteria for the acceptability of the putative surrogate endpoint is required to proceed with a decision in the best interest of patients.

We have further clarified this point in our revised paper (page 6, line 190):

> *"Using surrogate endpoints should only be considered in chronic disease settings with substantial unmet need, especially when collecting data on clinical outcomes requires clinical*

trials with unfeasibly long follow up durations. When generating evidence on clinical outcomes is not feasible, regulators should routinely evaluate the validity of the proposed surrogate endpoints."

The paper focusses on HTA methodology and change rather than that of the drug regulators, principally FDA and EMA. I would agree that the two main marketing authorisation regulators have weak processes regarding surrogate endpoints. They consult with pharma companies developing a registrational study and accept ahead of the trial which surrogates are used. My understanding is that HTA considerations are not sought at this stage, creating challenges for HTA agencies. This context is not mentioned. This will change in the UK with the ILAP programme managed by MHRA/NICE/SMC. FDA/EMA they do have a systematic approach, the question is how consistently it is applied. (I have been a member of scientific advisory panels for EMA).

We fully agree with the Reviewer that regulatory agencies could improve their processes regarding the use of surrogate endpoints. We now argue for raising the bar for the evidence required by regulators.

To address the Reviewer's comment, we have now added further details about the context in which surrogate endpoints are accepted. For example, we write on page 6, line 187):

"During clinical development, manufacturers engage with regulators to agree on the designs of studies (including endpoints) that will be used to support regulatory assessment. Regulators therefore have an opportunity to adopt rigorous criteria for accepting surrogate endpoints."

We have also referred to the importance of early dialogue between regulators and HTA bodies, and referenced ILAP as recommended by the Reviewer (page 7, line 206):

"Involving HTA bodies in early regulatory interactions with manufacturers may help align evidence requirements on surrogate endpoints. The UK Innovative Licensing and Access Pathway managed by the Medicines and Healthcare Products Regulatory Agency, NICE and the Scottish Medicines Consortium is aimed at facilitating such alignment.[53]"

The practical problems with surrogate endpoints are legion. The paper, understandably touches on the statistical uncertainties, rather than the clinical uncertainties which underlie them. As patient advocates we like certainty and surrogates do not help. We are aware of more than 20 different descriptions of the PFS endpoint in Patient Consent Information leaflets and are (anecdotally) aware of research clinicians using a different definition from the one in the Consent Information for the study they are working on. The challenge presented by surrogates is well summarised in this paper by the examples of RCT data contradicting data from the Phase II registrational study using a surrogate endpoint. Oloratumab in soft tissue sarcoma is engraved on my advocacy experience.

We fully agree with this and we hope our article further helps inform others.

The issue of regulators requiring confirmation studies is touched upon in the paper but the wider context of 'real-world' studies to provide confirmation of benefit is mentioned almost in passing, nor are the methodological challenges such studies present. The detail is not for this paper but the challenge deserves mention. As the paper points out too often the surrogate is replicated. High quality real-word data is, it must be said, something patient groups are

keen to see and more robust methodology would be valued.

Phase II studies with surrogates are now dominating cancer research and are increasing elsewhere. The description in the paper of methods for validating surrogate endpoints are a mathematical presentation which goes (at least) one step beyond my ability to follow but I can appreciate the rigour and the value it offers to the development of robust future procedures. I leave other reviewers to consider the maths. My concerns are about the long-term implementation, and thus the value of this work. The shape of future research will be governed by genetically defined disease rarity, the resulting small size of patient cohorts and an urgency to get drugs approved, this is where patient advocacy has impact. This reality is the key context for the future, it is an unstoppable movement. The solutions in this paper are seen from a statistical/methodological viewpoint, although the importance of 'flexibility' in decision-making is appropriately mentioned, but the implementation awaits NICE guidance.

Patient involvement in clinical trial design and development is a growing field. We know from evaluations in cancer, mental health, arthritis and other diseases that patients have important contributions to add in clinical research. Patients involved in this work usually learn about such issues as endpoints 'on the job'. It is often easy to accept a proposed surrogate without being aware of, let alone considering, the complexities but it is important they get to grips with them and are able to challenge the research teams they work with. There is an opportunity for all regulatory agencies to target these patient groups and support their training.

This brings me to PROs – Patient Reported Outcomes. The paper dismisses HRQoL without discussion. Lets consider HRQoL and PROs together. HRQoL uses questionnaire tools to calculate a score representing a patient's quality-of-life, possibly in a number of domains. PROs are data gathered from patients reporting specific effects of treatment, developing disease, function etc. In isolation these tools offer clinicians and patients information which can support decisions about treatment. There is however lack of comparability, no consistency, frequent missing data, and scarcely concealed bias which means that such data are in the 'flexibility' category for HTA appraisals. The issues are being addressed (SISAQOL, COMET etc). FDA and EMA see a future for these tools in providing co-primary endpoints (alongside a surrogate) in a drug trial, thus offering subjective information to regulators. HTA cannot ignore this, even though I accept that these are still developments they present another context specific to the future of surrogate endpoints for regulatory use which this paper should reflect.

We thank the Reviewer for these important reflections and suggestions. To address these comments, we have now added a new paragraph on the importance of patient involvement in clinical trial design and development in the "Way forward" section of our paper (page 7, line 220):

> *"Greater involvement of patients (and organisations representing patients) in regulatory and HTA processes is also essential to ensure that uncertainty related to using surrogate endpoints is explicitly presented and taken into account. Earlier studies showed that patients often overestimate the benefits of treatments on the basis of surrogate endpoints.[54] Patient input can help guide regulatory and HTA decisions regarding the appropriate use of surrogate endpoints."*

Summary.
A revised title making it clear the work is from an HTA standpoint.

Thanks for the suggestion. We have now revised the article (title, standfirst, key messages, and the main body of the text) and clarified that we cover this issue from both a regulatory and HTA perspective.

A Plain Language Summary

As mentioned above, we would value the Editors' guidance on how to proceed in this suggestion.

A significant re-write to:
• Add descriptive information so that the background to the use of surrogate endpoints is more completely contextualised and stakeholder views recorded

We have expanded the background to describe the issues in more detail and cover the benefits and limitations of using surrogates to the general reader. The section titled: "Benefits and harms of using surrogate endpoints in drug trials" sets the scene for this.

• Consider how the 'market' is moving and the role that surrogates will play in future. Discussion of the relevance of the ILAP project may be valuable here.

Thanks. This has been done and a new section titled "Way Forward" now covers the proposed approach.

• Consider how patients could 'toughen up' the challenge to trial designers employing surrogate endpoints at an early stage in development

Thanks for this important addition. As mentioned avove, we have added the following paragraph in the section titled "Way forward" (page 7, line 220):

> *"Greater involvement of patients (and organisations representing patients) in regulatory and HTA processes is also essential to ensure that uncertainty related to using surrogate endpoints is explicitly presented and taken into account. Earlier studies showed that patients often overestimate the benefits of treatments on the basis of surrogate endpoints.[54] Patient input can help guide regulatory and HTA decisions regarding the appropriate use of surrogate endpoints."*

• Discuss PROs used as co-primary endpoints with surrogates, if only to confirm the methodological challenges which PRO developers must address if this is to happen.

Thanks for the suggestion. We have now included HRQoL as an important clinically-meaningful and patient-relevant outcome to capture in clinical trials.

Comment to Editors

This paper has a very narrow view. I have no doubt that the rigour they propose for validating surrogate endpoints will add statistical value to regulatory review when surrogates are used but as described in its present form it would limit the paper's reach. We have to recognise that the real-world has contexts which impinge on just about every element of their narrative. Thus I am concerned that the paper will be seen as self-serving, rather than adding value to the continuing discussion about surrogates without revision. As a result I have focused my comments on 'contexts'. My mention of the ILAP Project, which brings together MHRA, NICE and SMC for a co-ordinated approach to drug regulation, is relevant to the eventual use of the authors' work.

Thank you once again for flagging this. In the newly-added "Way forward" section of our revised paper, we have outlined the changes required from both regulators and HTA agencies in response to increasing use of surrogate endpoints in clinical trials of new drugs. We have also mentioned ILAP, which is indeed very relevant (page 7, line 206):

> *"Involving HTA bodies in early regulatory interactions with manufacturers may help align evidence requirements on surrogate endpoints. The UK Innovative Licensing and Access Pathway managed by the Medicines and Healthcare Products Regulatory Agency, NICE and the Scottish Medicines Consortium is aimed at facilitating such alignment.[53]"*

When an established market is disrupted by an incomer the instinct of the traditionally grounded market occupant is to focus efforts and tighten its control. Meanwhile the disruptor innovates, reaches customers in new ways, and develops new products. An outsider might take a similar view on drug regulation although it is an established 'market' with occupants who are impossible to displace. There has been a continuing erosion by a combination of patient demand for access to new drugs, changing study methodology to meet evolving science, shortening timescales, smaller trial cohorts etc and now COVID has been an additional disrupting force. Innovation takes on many guises, rarely building on the familiar, more usually coming from a new direction.

These questions need addressing. We can recognise that rigorous standards should apply to surrogate endpoints but what is the place of that rigour in the value chain from trial design to real world use in a rapidly evolving (disrupted) environment? Is the acknowledged need for 'flexibility' more valuable than statistical rigour? Should methodologists be considering how they can add value to the formal inclusion of 'flexibility' in HTA appraisals without constraining the innovation or originality of such contributions? One such area is PROs. Rather than ignoring them and hoping they will go away (market occupant approach) regulatory bodies and their methodologists should get involved with making them better (innovator approach).

Thanks for this thoughtful input. We do agree with the need to engage with innovation in methods as well as in product development, this is why we stressed the need for pragmatism and flexibility when providing methods guidance to accommodate new and disruptive developments.