PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Real-world data from expanded access programmes in health technology assessments: a review of NICE technology appraisals
AUTHORS	Polak, Tobias; Cucchi, David; van Rosmalen, Joost; Uyl-de Groot, Carin

VERSION 1 – REVIEW

REVIEWER	Rob Hodgson University of York, Centre for Reviews and Dissemination
REVIEW RETURNED	22-Jun-2021

GENERAL COMMENTS	Review of: Real-world data from expanded access programs in health technology assessments: a review of NICE technology appraisals The authors present the results of a review of NICE technology appraisals considering whether they use data from expanded access programmes (EA). My main criticism of this study is that it does not go far enough in examining how EA data is used in NICE appraisals. The exemplars provided are most useful, but without a systematic analysis, the conclusions that can be drawn are limited. In particular, it is not clear whether EA data is of particular importance to decision-makers. I am also concerned that the discussion and conclusions do not fully match with the results obtained from the study. For example, the authors refer to the increasing use of EA data. This does not appear to be justified given the results of their analysis. Overall, I found this an interesting and mostly well-written manuscript but feel that the authors should revisit the paper before this study is accepted for publication. Abstract Methods: There appears to be some inconsistencies in the formatting. Some of the text appears to be in a smaller font. Main text Page 5 line 20 delete the word “on” prior to cost-effectiveness. Page 5 line 26. While ERG analyses is typically relates to the economic analyses it may also include reanalysis of clinical data. Page 6 first paragraph: For clarity it may be worth editing to make it clear that you are referring to pre-approval by regulators rather than HTA agencies. This may not be clear to a reader not familiar with the regulatory and reimbursement process. Page 6 second paragraph: There appears to be some inconsistencies in the formatting. Some of the text appears to be in a smaller font. Page 6 line 26. You make reference to Polak et al’s review of EA data by reimbursement agencies. It would be helpful (for context) to add some details of what this study found. This could be done in either the introduction or as part of the discussion. Page 8 line 36: This sentence is not very clear I think what your

	saying is that NICE committee's principally evaluate technologies value using a cost -per QALY approach. I would also add that NICE reference case requires that technologies evaluate benefits using QALY. The only exceptions to this is in rare cases where a cost minimisation approach is permitted (Fast track appraisals). The ambiguity in this sentence regarding the use of a cost-per-QALY approach is there not warranted. Page 8 line 36: You make generic reference to payer rather specifically to NICE – is this intentional? I'm not sure statement is true for all payers as many payers do not use QALYs to evaluate benefits. Page 12 line 3: Your results indicated that there was no trend in the proportion of appraisals making use of EA. This conclusion is therefore not appropriate. Page 12 paragraph 2 last sentence: This point seems highly speculative. There are range of reasons that this may be the case it may simply be that the need for the use of EA is less warranted in these disease areas. Page 12 line 26/27: While this sentence may be accurate, it would be more informative to refer to the proportions of decisions rather the absolute number of decisions. Page 12 line 30/31. This second point is somewhat redundant. Why would a regulatory agency make use of resource data given the remit of its decisions? Page 12 line 36: What does add "degrees of freedom" mean in this context? I do not follow the authors meaning. Page 12 paragraph 3 last four lines. I don't follow the point being raised. Please clarify. I would also add that NICE committees tend to value consistency in data sources wherever possible. Use of multiple sources for different parameter values would therefore be a cause for concern. Page 12 last paragraph. I found the points being raised in this paragraph difficult to follow and not entirely justified by the data. The authors can say little about the value of EA given the data they have collected as they have not assessed the relevance of EA to decision making. I would also add that a large proportion of technologies receive regulatory approval and NICE recommendation without an RCT. The authors may therefore wish to rephrase to refer to regulatory studies rather RCTs per se. Page 13 line 23: There appears to be some inconsistencies in the formatting. Some of the text appears to be in a smaller font. Page 13 line 23: This sentence is not well phrased. Consider revising. Reference to transitions in this context is also not appropriate. Transition probabilities are derived from effectiveness estimates and are not about resource use. Page 13 line 25: Trial values may also not be that informative as they typically multinational and do not contained data relevant to a particular national health system. Page 14 line 17/18: While this is true, conditional access via the cancer drugs fund will often require further data collection and therefore this divide is not as clear cut as indicated by the authors.

REVIEWER	Alison Smith University of Leeds Faculty of Medicine and Health
REVIEW RETURNED	12-Aug-2021

GENERAL COMMENTS	The authors present an analysis of NICE technology appraisals (TA), exploring how often and in what capacity TAs have used early access (EA) data. Whilst the research question answered here may not be considered a high priority topic, it is interesting from the

	perspective of understanding NICE decision making and research methodology. In general the paper is well written and provides some interesting results. There are some minor aspects I think the authors should address before publication, and further insights that could be provided (see my suggested topics provided at the end). Specific comments provided below. Abstract Please provide the definition for acronym TA. In the abstract you state that appraisals were searched from 2010-2020, but then in the summary bullet points you state 2021. Since you only search until the 1st Jan 2021, I would stick with describing your search as covering the period 2010-2020. Methods 'independently, manually, reviewed' should be 'independently and manually reviewed'. I do not believe enough detail on the computer script and text extraction process are reported to enable reproduction of this study – could the authors perhaps provide full details in supplementary material? Discussion The authors conclusion at the beginning of the discussion section states that NICE frequently includes EA data. However the core documentation for many of the TAs will be the manufacturer submission. As it is, the authors do not address the question of who it is that is using EA data - i.e. the manufacturer, the ERG, NICE, or the committee/other stakeholders? Some more information here breaking down where exactly the EA data is introduced would be of interest. The authors state a couple of times in the manuscript that the number of appraisals using EA data appears to increase over time. But this is only true in absolute and not relative terms, as the authors themselves point out in the results section. I do not think it is therefore appropriate to include these statements. Something that I think is missing from the Discussion, and perhaps the manuscript more broadly, is more information for readers to understand when and how EA data can and should be included within health technology assessments. Can the authors provide more information for readers interested in including EA data in their analyses as to a) how to assess if it is appropriate to include EA data, and b) how/where to find such data? Conclusion Given that you have stated that you could not quantify the added value of EA data on TAs, the final sentence of the conclusion section is too strong. The inclusion of EA data may help to inform NICE decision making. Based on my understanding of the results presented, further research is required to understand when EA data can and should be included in TAs. Overall comments: There are several interesting questions that I think the authors could address in this analysis but currently do not. Please can the authors review whether or not it would be possible to address these questions: 1. Does the inclusion of EA data affect NICE committee decisions? The authors mention that they could not quantify the added value of EA data. But they could have explored the impact of it's inclusion on

	NICE committee decisions, by comparing outcomes for those that did vs. did not include EA data. 2. Did the NICE committee ever comment on the use of EA specifically? If so were the comments favourable or negative? 3. Were any advantages and limitations with the use of EA data ever discussion by the manufacturer/ERG/committee? What were the most common advantages and limitations highlighted by each group? 4. Was the use of EA data mostly in manufacturer submissions, ERG reports, NICE documentation or other stakeholders?

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Reviewer: 1

Dr. Rob Hodgson, University of York

Comments to the Author:

See attached comments

Review of: Real-world data from expanded access programs in health technology assessments: a review of NICE technology appraisals

Comment 1: The authors present the results of a review of NICE technology appraisals considering whether they use data from expanded access programmes (EA). My main criticism of this study is that it does not go far enough in examining how EA data is used in NICE appraisals. The exemplars provided are most useful, but without a systematic analysis, the conclusions that can be drawn are limited. In particular, it is not clear whether EA data is of particular importance to decision-makers. I am also concerned that the discussion and conclusions do not fully match with the results obtained from the study. For example, the authors refer to the increasing use of EA data. This does not appear to be justified given the results of their analysis. Overall, I found this an interesting and mostly well-written manuscript but feel that the authors should revisit the paper before this study is accepted for publication.

Response 1*:* We thank the Reviewer for their time, critique, and suggestions to improve our work. We agree with the Reviewer that the theoretical use of EA (or any type of RWD for that matter) may be of less interest to HTA or government officials, yet there is substantial uncertainty for medical companies whether EA programs can (i) generate data and (ii) whether these data can be submitted for regulatory or reimbursement purposes. We have provided references to these discussions in the Introduction.

Secondary, such programmes may potentially collect real world data in a regulatory pre-approval setting, but the generation and useability of evidence derived from these programmes remains a topic of debate.(ref10–16)

Therefore, our primary goal was to show **that** these data can or are being used in the first place. We have added references to demonstrate the debate on whether these data can be used in the first place to highlight the relevance of this topic. Providing evidence of the use of these data by regulators, payors, or medical professionals, can help incentivize companies to run these programs and may speeden access to investigational treatments.

Furthermore, we have also attempted to provide more insights into **how** these data can be used, which may be of interest to a more HTA-focused audience with additional analyses. Although we agree that a systematic analysis of all occurrences of EA terms in appraisals may lead to more robust insights on the usage of these data, the sheer size (>4500 terms) and complexity of this problem hampers this analysis and renders it beyond the scope of this manuscript. As both reviewers bring up this point, we have performed several additional analyses, added supplementary material, and provided the entire dataset and code for further research. We have described in detail our additional work in the 'Exploratory Analysis' section in our response to Reviewer 2.

Methods:

Comment 2: There appears to be some inconsistencies in the formatting. Some of the text appears to be in a smaller font.

Response 2: We have now reformatted the complete main text to "Times New Roman" font 10.

Main text

Comment 3: Page 5 line 20 delete the word "on" prior to cost-effectiveness.

Response 3: We thank the reviewer for noticing this error and have now rewritten this sentence slightly to make it grammatically correct.

In the United Kingdom (UK), the National Institute for Health and Care Excellence (NICE) conducts technology appraisals (TAs) to evaluate cost-effectiveness of technologies (e.g. drugs, medical devices) and to determine their impact on health care budgets[ref1].

Comment 4: Page 5 line 26. While ERG analyses is typically related to the economic analyses it may also include reanalysis of clinical data.

Response 4: The reviewer is right that the ERG re-analyses clinical data in some cases. We have now adjusted our statement accordingly.

The submission is scrutinized by an independent Evidence Review Group (ERG), which critically reviews the manufacturer's submission and performs additional exploratory analyses of cost-effectiveness; in some cases the ERG even re-analyses clinical data[ref1,4–6].

Comment 5: Page 6 first paragraph: For clarity it may be worth editing to make it clear that you are referring to pre-approval by regulators rather than HTA agencies. This may not be clear to a reader not familiar with the regulatory and reimbursement process.

Response 5: We thank the reviewer for this suggestion and agree that this clarification is important for readers that are not familiar with regulatory and reimbursement processes. We have now incorporated the requested clarification.

*It offers a potential opportunity to collect real world data in a **regulatory** pre-approval setting*

Comment 6: Page 6 second paragraph: There appears to be some inconsistencies in the formatting. Some of the text appears to be in a smaller font.

Response 6: We thank the reviewer for noticing this and have now corrected the formatting inconsistencies.

Comment 7: Page 6 line 26. You make reference to Polak et al's review of EA data by reimbursement agencies. It would be helpful (for context) to add some details of what this study found. This could be done in either the introduction or as part of the discussion.

Response 7: We agree with the reviewer that including details of this study is needed to provide more context for the current study. As a clarification, our previous study focused on the use of EA data by regulators, not by reimbursement agencies. Our aim was to identify and characterize the instances when regulators used EA data to inform the clinical efficacy profile of submissions. We now adopted the most important conclusion of the study by Polak et al in the introduction.

These data are increasingly accepted to support evidence of efficacy by regulators, especially when collecting data in controlled settings is infeasible, such as in ultra-rare diseases, or is deemed unethical, in the case of extremely large treatment effects.[ref10]. However, the use of EA data by payers or HTA bodies remains unquantified.

Comment 8: Page 8 line 36: This sentence is not very clear. I think what your saying is that NICE committee's principally evaluate technologies' value using a cost -per QALY approach. I would also add that the NICE reference case requires that technologies evaluate benefits using QALY. The only exception to this is in rare cases where a cost minimisation approach is permitted (Fast track appraisals). The ambiguity in this sentence regarding the use of a cost-per-QALY approach is not warranted.

Response 8: We agree with the reviewer that there are exceptions to the NICE reference case requiring benefits evaluated by using QALYs. Although cost-minimisation approaches implicitly assume a zero difference in benefit in QALYs, we follow the reviewer that our current phrasing is ambiguous. We therefore changed the wording and added reference to NICE documentation.

NICE requires that benefits of technologies are evaluated using quality-adjusted life years (QALYs), as NICE's decision to recommend or not recommend a product for reimbursement depends (among other things) on the willingness-to-pay for an incremental year in perfect health – the so-called cost-per-QALY approach.

Comment 9: Page 8 line 36: You make generic reference to payer rather specifically to NICE – is this intentional? I'm not sure if this statement is true for all payers as many payers do not use QALYs to evaluate benefits.

Response 9: We agree with the reviewer and we have integrated Comment 8 and Comment 9 as follows.

NICE requires that benefits of technologies are evaluated using quality-adjusted life years (QALYs), as NICE's decision to recommend or not recommend a product for reimbursement depends (among other things) on the willingness-to-pay for an incremental year in perfect health – the so-called cost-per-QALY approach.

Comment 10: Page 12 line 3: Your results indicated that there was no trend in the proportion of appraisals making use of EA. This conclusion is therefore not appropriate.

Response 10: We agree that only in absolute terms, the number of TAs that make use of EA increases. We have now changed the conclusion accordingly .

NICE uses EA data in over one in five (21.1%) of their appraisals, and this number appears to remain stable over time.

Comment 11: Page 12 paragraph 2 last sentence: This point seems highly speculative. There are range of reasons that this may be the case; it may simply be that the need for the use of EA is less warranted in these disease areas.

Response 11: We agree that this point is speculative and that there are a range of other reasons explaining this phenomenon. By no means did we mean to imply that cardiologist or ophthalmologist would be neglective - only because the use of EA may be less warranted in these disease areas they naturally may be less acquainted with EA. We have now rewritten this part, making it clearer that our statements are speculative and also including other potential explanations.

There is a range of possible explanations for this discrepancy. Perhaps, drug developers in these areas may be less familiar with collecting and using EA data, or cardiologists and ophthalmologists may be less acquainted with EA than haemato-oncologists – simply because EA may be less warranted in these disease areas.

Comment 12: Page 12 line 26/27: While this sentence may be accurate, it would be more informative to refer to the proportions of decisions rather than the absolute number of decisions.

Response 12: We agree with the reviewer that the proportion of decisions may provide additional information on top of the absolute number. Unfortunately, the denominator (in case of FDA/EMA approvals) is not easy to compute as the instances described in Polak et al are a mix of first approvals, indication extensions, and population extensions. Therefore, this has been omitted from Polak et al. in the first place and we are unable to incorporate that into this submission.

Comment 13: Page 12 line 30/31. This second point is somewhat redundant. Why would a regulatory agency make use of resource data given the remit of its decisions?

Response 13: We agree with the reviewer that this argument comes across as superfluous. Therefore, we have removed the second reason. This section now reads:

One reason for this may be that payors have a higher uptake of RWD in their decision making. Furthermore, they also assess comparative effectiveness rather than efficacy .

Comment 14: Page 12 line 36: What does add “degrees of freedom” mean in this context? I do not follow the authors meaning.

Response 14: We agree that this wording is vague. Because NICE investigates resource use and comparative effectiveness, more data are required than for the analysis of efficacy and safety that is performed by regulators. We have now removed this sentence to avoid ambiguity.

Modelling cost and comparative effectiveness by definition necessitates a variety of input parameters, every one of them potentially coming from different sources, such as EA.

Comment 15: Page 12 paragraph 3 last four lines. I don’t follow the point being raised. Please clarify. I would also add that NICE committees tend to value consistency in data sources wherever possible. Use of multiple sources for different parameter values would therefore be a cause for concern.

Response 15: We agree with the reviewer that consistency in data source. We have removed the ambiguous sentence as stated above in Response 14. We also note that NICE seems to be ambiguous in valuing this consistency, as we highlight in the Supplementary Files:

‘*The company stated that, based on its observations from the compassionate use scheme for caplacizumab, in NHS clinical practice, people would have it for a shorter duration than in the trials. The committee in general prefers not to disassociate estimates of cost and effectiveness from a trial.*

However, it appreciated that many assumptions about caplacizumab's effectiveness in this model were not taken from the main trial.' (From, Supplementary Files, TA667)

Furthermore, there are also arguments to be made for combining various data sources to improve generalizability. The exact parameter of interest influences this choice. For example, when the 'average weight of the patient population' needs to be estimated in our example of ipilimumab (TA268, Results section, Resouce use paragraph), combining estimates from trial and compassionate use seems obvious. When it comes to other parameters, such as efficacy, we agree with the reviewer that direct evidence is usually preferred above indirect analyses.

Comment 16: Page 12 last paragraph. I found the points being raised in this paragraph difficult to follow and not entirely justified by the data. The authors can say little about the value of EA given the data they have collected as they have not assessed the relevance of EA to decision making. I would also add that a large proportion of technologies receive regulatory approval and NICE recommendation without an RCT. The authors may therefore wish to rephrase to refer to regulatory studies rather RCTs per se.

Response 16: Indeed the points we wanted to make were stated somewhat unwieldy. We aimed to clarify the points being raised in this paragraph as follows:

For safety, the use of registries, post-approval safety studies, or pharmacovigilance during EA, is useful to detect infrequently occurring adverse events. Indeed, we identified such an example in TA533, where the compassionate use programme led to the identification of a rare but serious adverse event. Overall, the evidence for assessing safety and efficacy should primarily come from regulatory studies and can be synthesized with RWD or other non-randomized sources, such as EA programmes.

We have performed additional analyses and provided supplementary files to assess the relevance of EA to decision making, see: Exploratory Analysis section (last points of Reviewer 2)

Comment 17: Page 13 line 23: There appears to be some inconsistencies in the formatting. Some of the text appears to be in a smaller font.

Response 17: We thank the reviewer for noticing this and have now corrected the formatting inconsistencies.

Comment 18: Page 13 line 23: This sentence is not well phrased. Consider revising. Reference to transitions in this context is also not appropriate. Transition probabilities are derived from effectiveness estimates and are not about resource use.

Response 18: We have improved our sentence based upon the Reviewers suggestion. This sentence now reads as follows:

'Furthermore, modelling resource use requires estimates of a large number of input parameters, such as costs, incidence and also transition parameters that determine the amount of time spent in a disease state. Some of these parameters can only be estimated from studies with lengthy follow-up periods, so that patient or population registries or EA programmes would be best suited to inform decision making on these model inputs.'

We acknowledge that in most of the times, transition probabilities are derived from effectiveness estimates. However, this depends on the nature of the disease. For chronic illnesses, transition rates determine the duration of time spent in a certain state, and this duration is highly influential to the use

of resources. Another counter example is transition to a state of 'other cause mortality', which is usually not (or should not be) based on effectiveness estimates.

Comment 19: Page 13 line 25: Trial values may also not be that informative as they are typically multinational and do not contain data relevant to a particular national health system.

Response 19: We thank the reviewer for this important comment and have incorporated this comment in the text.

Finally, trial values may not be sufficiently informative, as they are typically multinational and do not contain data relevant to a particular national health system.

Comment 20: Page 14 line 17/18: While this is true, conditional access via the cancer drugs fund will often require further data collection and therefore this divide is not as clear cut as indicated by the authors.

Response 20: We agree with the author that in specific national instances in certain medical disciplines, such as the Cancer Drugs Fund, or the 'conditional approval' in the Netherlands, drugs can generate real-world data prior to receiving reimbursement/approval. We suggest adding the word 'typically' to cover these specific situations.

– this makes EA data different from general RWD sources (e.g. electronic health records or claims and billing data), as the latter will typically only start generating evidence once the drug has been approved.

Reviewer 2

Reviewer: 2

Dr. Alison Smith, University of Leeds Faculty of Medicine and Health

Comments to the Author:

Comment 1: The authors present an analysis of NICE technology appraisals (TA), exploring how often and in what capacity TAs have used early access (EA) data. Whilst the research question answered here may not be considered a high priority topic, it is interesting from the perspective of understanding NICE decision making and research methodology. In general the paper is well written and provides some interesting results. There are some minor aspects I think the authors should address before publication, and further insights that could be provided (see my suggested topics provided at the end). Specific comments provided below.

Response 1: We thank the reviewer for the careful consideration of our paper. We have incorporated the suggestions provided to improve our paper.

Abstract

Comment 2: Please provide the definition for acronym TA.

Response 2: We have now provided the definition of TA in the Objectives section of the abstract.

Comment 3: In the abstract you state that appraisals were searched from 2010-2020, but then in the summary bullet points you state 2021. Since you only search until the 1st Jan 2021, I would stick with describing your search as covering the period 2010-2020.

Response 3: Indeed, 2010-2020 is the correct period. We have now changed this in the Strengths and Limitations section of the manuscript.

We evaluated all NICE appraisal documentation from 2010 to 2020.

Methods

Comment 4: 'independently, manually, reviewed' should be 'independently and manually reviewed'.

Response 4: We thank the reviewer for this correction and have now adjusted this in the Methods section of the manuscript.

When at least one of these 'expanded access terms' were present, two authors (T.B.P. and D.G.J.C.) independently and manually, reviewed the context of the term.

Comment 5: I do not believe enough detail on the computer script and text extraction process are reported to enable reproduction of this study – could the authors perhaps provide full details in supplementary material?

Response 5: We have now provided the detailed protocol in the Supplementary Files, as well as a link to the Github page from the main author where the individual scripts and original data are now accessible.

https://github.com/TobiasPolak/RWD-from-EAP-in-HTA-a-review-of-NICE-technology-appraisals

Discussion

Comment 6: The author's conclusion at the beginning of the discussion section states that NICE frequently includes EA data. However the core documentation for many of the TAs will be the manufacturer submission. As it is, the authors do not address the question of who it is that is using EA data - i.e. the manufacturer, the ERG, NICE, or the committee/other stakeholders? Some more information here breaking down where exactly the EA data is introduced would be of interest.

***Response 6**: T*o improve reproducibility, we have added the entire list of terms that occurred in documentation available on NICE' website, including the document title, page number, direct link to the page, and the two sentences before and after the 'expanded access' term. In total, these terms occur in over 4500 pages.

It is difficult to address who it is that is using EA data. First, these data often get copied over in several documents (Manufacturer Submissions, ERG review, etc). Second, 'using' data is often not a binary decision, but involves a complex discussion on how to weigh data or in what way and to what extent data may be biased. To illustrate the complexity of this issue, we have in detail discussed the 5 appraisals in which EA terms most frequently occurred. This discussion can be found in **Supplementary Files 2.**

We have now rephrased this sentence to *'We have found that data from EA programmes are frequently included'*

Comment 7: The authors state a couple of times in the manuscript that the number of appraisals using EA data appears to increase over time. But this is only true in absolute and not relative terms,

as the authors themselves point out in the results section. I do not think it is therefore appropriate to include these statements.

Response 7*:* We agree that only in absolute terms, the number of TAs that make use of EA increases. We have now changed the conclusion accordingly in the abstract and conclusion.

Abstract: The number of TAs that utilize EA data remained stable over time, and EA data utilization was disproportionately distributed across disease areas (p=0.001).

Discussion: The use of data from EA programs appears to be stable over the years.

Conclusion: NICE uses EA data in over one in five (21.1%) of their appraisals, and this number appears to be stable over time.

Comment 8: Something that I think is missing from the Discussion, and perhaps the manuscript more broadly, is more information for readers to understand when and how EA data can and should be included within health technology assessments. Can the authors provide more information for readers interested in including EA data in their analyses as to a) how to assess if it is appropriate to include EA data, and b) how/where to find such data?

Response 8*:* We thank the reviewer for this comment. We now refer to the evidence submission guidelines from NICE and have rephrased the following sentence to align with these guidelines.

Whether using EA data (or other non-randomized data) for payer decision making is wise, depends in part on the robust design and execution of the EA program, and the relevance to the decision problem.

Furthermore, we have suggested several sources of data at the end of the discussion:

Results from EA programs can be obtained via peer-reviewed publications, if published. Alternatively, data can be requested via the medical company using data sharing platforms, such as Vivli.[ref 26] Finally, data may be available through local investigators (See HST7, ***Supplementary Files 2****).*

Conclusion

Comment 9: Given that you have stated that you could not quantify the added value of EA data on TAs, the final sentence of the conclusion section is too strong. The inclusion of EA data may help to inform NICE decision making. Based on my understanding of the results presented, further research is required to understand when EA data can and should be included in TAs.

Response 9: We thank the reviewer with this suggestion and agree that our current wording is too strong. Therefore, we have changed the phrasing into:

In synthesis with evidence from well-controlled trials, data collected from EA programmes may meaningfully inform NICE decision making - but further research is required to understand when EA data can and should be included in health technology assessments.

Overall comments:

There are several interesting questions that I think the authors could address in this analysis but currently do not. Please can the authors review whether or not it would be possible to address these questions:

Exploratory Analyses:

1. Does the inclusion of EA data affect NICE committee decisions? The authors mention that they could not quantify the added value of EA data. But they could have explored the impact of it's inclusion on NICE committee decisions, by comparing outcomes for those that did vs. did not include EA data.

Response: We thank the reviewer for this comment. To investigate this issue, we could indeed associate the recommendation decision (yes/no) for single technology appraisals with the usage of EA data (yes/no). Out of the 36 non-recommended appraisals, EA was used in 9 (25%), whereas of the recommended appraisals, EA was used in 58 (19%). (p=0.4, Pearson's Chi-squared test). We have some reservations that, in our view, preclude the inclusion of this analysis in the main paper.

First, appraisals that probably would not become recommended can be withdrawn by the manufacturer or terminated. For these appraisals there is no detailed documentation available. Using only the appraisals with documentation therefore biases our dataset for this analysis. Second, the recommendation is heavily influenced by the cost of treatment (products are approved after confidential discounts) and expanded access is of no influence to this decision. Therefore, we fear that presenting this association could suggest a causal relationship that most probably is absent. Third, we would hardly expect a substantial effect of a binary predictor of EA-usage, as there may be a plethora of other factors that in part determine the decision.

2. Did the NICE committee ever comment on the use of EA specifically? If so, were the comments favourable or negative?

Response: Please see 'Exploratory Analyses' under 3.

3. Were any advantages and limitations with the use of EA data ever discussed by the manufacturer/ERG/committee? What were the most common advantages and limitations highlighted by each group?

Response: Please see 'Exploratory Analyses'.

Exploratory Analyses

We thank the reviewer for the suggested exploratory analyses to improve our research. Based on these suggestions, we have performed the following analyses and supplemented the following files:

1. We reran our analysis to screen for use of EA terms in Technology Assessments and synthesised a complete list of all TAs, URLs linking to these TAs, occurrence of an EA term and the page on which the term occurred. We reclassified one TA (TA487) to have used EA data, identifying 80 instead of 79 instances.
2. As the reviewer requested more context, we added this context to the list, by automatically including the two lines before and after the EA term.
3. We provide this list with this rebuttal to the reviewers, facilitating insights in the context and scope of our project. Furthermore, it is freely accessible on Github. Each term provides the URL to the page on which the EA term was identified. This includes over 4500 EA terms used in TAs for the reviewer to peruse.

With these above analyses, we aimed to answer the additional questions (Nos 2. and 3.) of the reviewer. We found that the NICE committee indeed comments on the use of EA data specifically. However, the sheer dimension (>4500 terms) in combination with the heterogeneity of comments and discussions regarding data from expanded access hampers a systematic analysis of <u>all</u> occurrences and renders it beyond the scope of this manuscript. To illustrate this and to provide an idea of the comments, limitations and advantages discussed, we analysed in detail the five TAs comprising the

most individual occurrences of search terms, and provide a summary of these analyses in the **Supplementary Files 2**. This shows that an unambiguous quantitative conclusion is hard to provide. However, the detailed and nuanced discussion of EA-data in these examples provides informative qualitative insights. In TA391 (see Supplementary Files 2), the EA data are even topic of debate in a formal Appeal procedure, highlighting the importance of these data.

Furthermore, we sought to determine the exact occurrence/use of EA data. We found that this is a very difficult problem to address. Second, it is difficult to determine who is the 'first': the company may mention an EA program but did not include data, or did not include the data to the same extent as the ERG and/or NICE decide to. Therefore, we are unfortunately unable to systematically quantify the most common advantages and limitations by each group. But, we provide several qualitative examples of discussions of the use of EA data in **Supplementary Files 2.** We hope that, although not a systematic analysis, this provides more insight.

In the main text, we have added this to the Methods section:

Additionally, we provide a narrative summary of the 5 appraisals that contain the most occurrences of the search terms to illustrate the use of EA data qualitatively.

In the main text, we now refer to these examples as follows in the Results section.

'To further illustrate how EA data are appraised by the manufacturer, ERG and NICE committee, and what the advantages and limitations of its use may be, a detailed discussion of the top-5 appraisals in which the search terms most frequently occurred can be found in the Supplementary Files 2.' This includes representative examples in the areas of haemato-oncology (e.g. prostate cancer, follicular lymphoma) and rare diseases (e.g. spinal muscular atrophy).

We have described the above limitations and need for future research as follows:

Additionally, it is difficult to measure the impact of EA data, as it is not always clear how these data have exactly been used: the use of EA data – and the appraisal thereof - in HTA by the manufacturer, ERG or NICE committee are difficult to quantify due the complexity and extent of the discussions described in the documentation. Although we have provided the reader with both high-level quantitative statistics and with illustrative qualitative examples from our data set, future research could attempt to systematically analyse these topics.

VERSION 2 – REVIEW

REVIEWER	Rob Hodgson University of York, Centre for Reviews and Dissemination
REVIEW RETURNED	19-Nov-2021

GENERAL COMMENTS	I have no further comments.