

PEER REVIEW HISTORY

BMJ Medicine publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Clarifying what it means to be "representative"
AUTHORS	Rudolph, Jacqueline; Zhong, Yongqi; Duggal, Priya; Mehta, Shruti; Lau, Bryan

VERSION 1 - REVIEW

REVIEWER 1	Riley, Richard; University of Birmingham, Institute of Applied Health Research. Competing Interest: None
REVIEW RETURNED	04-Nov-2022

GENERAL COMMENTS	<p>This is a very important topic, and highly suitable for a methods primer in BMJ Medicine. Representative is rarely defined, yet is inherently a key part of undertaking, designing and appraising most medical research articles. Hence, it is important that the issue of what we mean by representative is discussed, and so this article is timely and can make a big difference. It is also well-written. I have the following comments for the authors to consider in their revision.</p> <ol style="list-style-type: none">1) "We presume perfect internal validity of study results" – I think the words 'internal validity' are also ambiguous and need to be defined. Or could remove this and rather say that it is assumed the study have no risk of bias concerns (e.g. as defined by a risk of bias tool).2) Some risk of bias tools also have an aspect called 'applicability' (e.g. see PROBAST and QUADAS tools). I think this should also be mentioned, and the idea of representativeness linked to this concept too (if the authors feel it does link of course).3) "We suggest that a study sample is representative of a well-defined target population when ..." – I think the authors need to be stronger here and say that they define (and not suggest) a study sample to be representative of a well-defined target population when ...4) It may help to give the paragraphs on 'generalizable in estimate' and 'generalizable in interpretation' their own sub-headings (perhaps in bold italics, within the section on defining representativeness), to make them stand out clearer, as they are key.5) I think the paper would benefit from an example woven in early on – essentially a running example. I appreciate that the four examples at the end are already there (and are excellent), but there are lots of sentences (which I agree are needed) to explain the nuances of the concepts early on, so having a running example would complement this. For example, as most readers know about RCTs, I think it makes sense to draw the reader in with an RCT running example. The other examples can still occur at the end (or perhaps in Boxes).
-------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<p>For example, when talking about the idea of estimands and estimates, it would help to focus on the treatment effect from an RCT. This would hold the readers hand more, and bring the more general (non-statistically trained) reader in more easily</p> <p>6) “stratum-specific estimates” – define stratum and give example</p> <p>7) Effect modifiers – again, this needs explaining for the general reader.</p> <p>8) In the ‘generalizable in interpretation’ paragraph, an example would really help too.</p> <p>9) “size of the covariate strata need not exactly match the target population” – I think relative size is what is intended here? No sample matches the target population in terms of actual size.</p> <p>10) “If we wished to generalize to target populations beyond New York (e.g., the entire US), we would need to make assumptions about whether there are effect measure modifiers that differ between New York and the entire US and whether we have them measured.” – and also assumptions about the magnitude of the effect modification?</p> <p>11) I think it may help to move the examples to 3 boxes to reduce main text length and to make them stand out even more (with the RCT as a running example in the actual text, as mentioned), with then just a brief section signposting the examples and what they broadly show.</p> <p>12) The conclusions section is all very interesting, although usually it would be better to end a research methods paper with a briefer section – like just the first paragraph of the conclusion section. The other parts I still found interesting but I wondered if they should be raised earlier in the paper a new section called “Which is more important: generalizing the interpretation or the estimate?”</p> <p>In summary, a very important paper, that is well-written and I enjoyed reading it. I hope these comments are helpful to the authors moving forwards, and I look forward to seeing their revision.</p>
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

REVIEWER 2	Lu, Haidong; Yale University School of Public Health. Competing Interest: None
REVIEW RETURNED	26-Nov-2022

GENERAL COMMENTS	<p>In this paper, the authors attempted to provide a comprehensive definition of “representativeness” in medical research. Such bold endeavors are appreciated, especially considering the ambiguousness of the term “representative”. However, there are some areas/concerns that need to be clarified or addressed, in order to make the paper more useful. Below are my comments.</p> <p>1. The authors may need to give more compelling explanations/rationales about why their new definition of “representativeness” is useful and important. Many people’s perception (including mine) of the term “representative” is probably in line with the (1st) definition (page 4 lines 37-39) of the 2013 IJE series papers that representativeness occurs when “study sample is</p>
-------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

a simple random sample of the target population” through representative sampling. In the current work, the authors gave a new broad definition of “representative” that the study sample is representative when the results are generalizable in estimate or in interpretation. Especially, the authors mentioned, if the sample can be quantitatively generalized to the target population through weighting or standardizing, the study sample can still be considered representative even when the distribution of the key covariates differ between the sample and the target population. However, such definition is a bit counterintuitive and awkward. Under such definition, most study samples will be representative once the effect modifiers are measured and accounted for, and the positivity assumption for generalizing is satisfied. It deviates from people’s classic perception of “representative” and its relation to “representative sampling”, and may be difficult for readers to digest. Therefore, to make this new definition of representativeness approachable and useful, more rationales should be provided.

Importantly, I am thinking about descriptive studies estimating the prevalence of disease: if we wanted to estimate the depression rate in the overall population in the US, and selected non-random sample that is majorly male, is it reasonable to claim this sample is representative even if we are able to account for the difference in sex distribution?

2. The illustration of “generalizable in interpretation” is not very clear (page 6 lines 1-29). Especially, it lacks some references to support this concept. From my interpretation, this type of “generalization” is more like the generalization in “effect sign”. That is, if the effect in the target population (e.g., humans) and the effect in the study sample (e.g., lab mice) are in the same direction (e.g., positive effect), one can say the results are generalizable in “effect sign”. In my view, “generalizable in effect sign” is more intuitive than “generalizable in interpretation”. A useful reference that could be used to illustrate such sign generalization:

EGAMI, N., & HARTMAN, E. (2022). Elements of External Validity: Framework, Design, and Analysis. *American Political Science Review*, 1-19. doi:10.1017/S0003055422000880

3. Continuing the Comment 2, as in authors definition, if the results from lab mice are generalizable in interpretation to humans, the study sample will be representative. However, it is a little bit awkward to say, the study sample of lab mice is representative of the human beings. Further, the lab mice in the study sample is probably unrepresentative in its species, as they are homogeneous with identical genes (also mentioned in the 2nd paragraph of Greenland’s IJE paper in 2013). So representativeness may not be necessarily linked to generalizability. The knowledge can be generalizable when the sample is not representative.

The authors devoted majority of the text to generalizability, but what is lacking is a strong case why the definition of representativeness should be tied to generalizability, and why such definition is more intuitive and useful (similar to my comment 1).

4. As in the text, generalizing in estimate involves both descriptive studies and causal studies. It may not be appropriate to say “effect

	<p>measure modifiers” for descriptive studies (on page 7 lines 2-13).</p> <p>5. Page 11 lines 47-48: “the 36.3% prevalence of SARS-CoV-2 could potentially be generalized to the full sample of people who inject drugs in this region, at least during the time frame examined.” The authors may want to briefly describe the sampling method used in this descriptive study and give a reasonable speculation that it is a random sample.</p>
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

VERSION 1 – AUTHOR RESPONSE

We thank the editor and reviewers for considering our paper. Below, see our responses to the comments. All page numbers refer to the clean document without track changes.

Reviewer 1

This is a very important topic, and highly suitable for a methods primer in BMJ Medicine. Representative is rarely defined, yet is inherently a key part of undertaking, designing and appraising most medical research articles. Hence, it is important that the issue of what we mean by representative is discussed, and so this article is timely and can make a big difference. It is also well-written. I have the following comments for the authors to consider in their revision.

RESPONSE: We are pleased that the reviewer sees the importance of the paper’s message. We respond to each of the reviewer’s comments below.

1. “We presume perfect internal validity of study results” – I think the words ‘internal validity’ are also ambiguous and need to be defined. Or could remove this and rather say that it is assumed the study have no risk of bias concerns (e.g. as defined by a risk of bias tool).

RESPONSE: We have removed the reference to internal validity and have revised the sentence to say (p. 4), “We presume no bias in the study’s results; in any real-world study, bias will need to be weighed alongside whether the sample is representative or its results are generalizable or applicable to a target population.”

2. Some risk of bias tools also have an aspect called ‘applicability’ (e.g. see PROBAST and QUADAS tools). I think this should also be mentioned, and the idea of representativeness linked to this concept too (if the authors feel it does link of course).

RESPONSE: Thank you for pointing us toward these risk of bias tools, as we had not previously considered them in the context of representativeness. One definition we found for a lack of applicability was “when the population, predictors, or outcomes of the study differ from those specified in the review question,” such as when the sample the prediction model was built in differs from the target population. There is certainly some overlap with what we discuss in this manuscript. Thus, we have added the following text on p. 10:

“Fourth, the concepts of representativeness and generalizability discussed above are also related to the term “applicability” used in certain risk-of-bias tools, such as the PROBAST and QUADAS (15,16). All concepts center on the idea that it is important to assess a study and its results in terms of how well they can be related to some target population. While here we discussed causal and descriptive studies, the two tools mentioned apply this concept to predictive and diagnostic studies.”

3. “We suggest that a study sample is representative of a well-defined target population when ...” – I think the authors need to be stronger here and say that they define (and not suggest) a study sample to be representative of a well-defined target population when ...

RESPONSE: As suggested, we have strengthened our language and now state (pp. 4-5), “We define a study sample to be representative of a well-defined target population if the results estimated in that sample are generalizable to the target population.”

4. It may help to give the paragraphs on ‘generalizable in estimate’ and ‘generalizable in interpretation’ their own sub-headings (perhaps in bold italics, within the section on defining representativeness), to make them stand out clearer, as they are key.

RESPONSE: We added sub-headings for the paragraph(s) explaining generalizability in estimate and in interpretation.

5. I think the paper would benefit from an example woven in early on – essentially a running example. I appreciate that the four examples at the end are already there (and are excellent), but there are lots of sentences (which I agree are needed) to explain the nuances of the concepts early on, so having a running example would complement this. For example, as most readers know about RCTs, I think it makes sense to draw the reader in with an RCT running example. The other examples can still occur at the end (or perhaps in Boxes). For example, when talking about the idea of estimands and estimates, it would help to focus on the treatment effect from an RCT. This would hold the readers hand more, and bring the more general (non-statistically trained) reader in more easily.

RESPONSE: We have taken the reviewer’s suggestion and fully integrated the randomized control trial example into the explanation of concepts. All other study design examples have been moved to Boxes.

6. “stratum-specific estimates” – define stratum and give example

RESPONSE: We define “stratum-specific estimate” by way of example (p. 6): “More generally, even if the distribution of the key covariates differs between the sample and target population, the sample may still be representative within strata of the key covariates, such that the stratum-specific estimates (e.g., risk difference within age categories) can be generalized from the sample to the target population.”

7. Effect modifiers – again, this needs explaining for the general reader.

RESPONSE: After the first mention of “effect measure modifiers,” we define this term as follows (pp. 5-6): “These key covariates are those that affect the variable under study (e.g., hospitalization or death) and thus are potential effect measure modifiers of the effect of an exposure on that variable. By effect measure modifiers, we mean variables where the effect of the exposure differs by levels of that variable on some scale (risk difference, risk ratio, odds ratio, etc.). In our example, age might be an effect measure modifier because the effect of molnupiravir on hospitalization or death (as quantified by the risk difference) might differ across ages.”

8. In the ‘generalizable in interpretation’ paragraph, an example would really help too.

RESPONSE: See our response to Comment 5 above.

9. “size of the covariate strata need not exactly match the target population” – I think relative size is what is intended here? No sample matches the target population in terms of actual size.

RESPONSE: Yes, this was the meaning intended – that the prevalence proportion (not the absolute size of the group) does not have to be the same for a subgroup in both the sample and the target

population. This has now been clarified (p. 6): “While this requires that all the key covariates be measured, the proportion of the sample in the covariate strata need not exactly match the proportion who fall into that subgroup in the target population.”

10. “If we wished to generalize to target populations beyond New York (e.g., the entire US), we would need to make assumptions about whether there are effect measure modifiers that differ between New York and the entire US and whether we have them measured.” – and also assumptions about the magnitude of the effect modification?

RESPONSE: We are interpreting the reviewer’s comment to mean the case where there is heterogeneity in the effect of the modifier on the outcome (and thus the magnitude of its modification of the exposure-outcome relationship). In our view, this additional heterogeneity would be caused by the existence of additional effect measure modifiers, which also affect (or are associated with) the outcome. Provided the modification all occurs on the same scale, these variables could potentially also be modifiers of the exposure-outcome relationship. If we measured and accounted for all such variables (via stratification or methods to generalize), then there should be no concerns related to heterogeneity of the effect of the modifier.

If we missed the reviewer’s point, we would be happy to revise the text, after additional clarification.

11. I think it may help to move the examples to 3 boxes to reduce main text length and to make them stand out even more (with the RCT as a running example in the actual text, as mentioned), with then just a brief section signposting the examples and what they broadly show.

RESPONSE: As mentioned in our response to comment 5, we have integrated the RCT example into our explanation of concepts and have moved the other examples to Boxes.

12. The conclusions section is all very interesting, although usually it would be better to end a research methods paper with a briefer section – like just the first paragraph of the conclusion section. The other parts I still found interesting but I wondered if they should be raised earlier in the paper a new section called “Which is more important: generalizing the interpretation or the estimate?”

RESPONSE: When integrating the RCT example, we reorganized the paper. We now include a Discussion section that contains the paragraphs that previously began with “There are two points related to defining representativeness that are worth highlighting” and the paragraphs on the relative importance of generalizing the interpretation or the estimate. We then have a shorter Conclusion section that summarizes the main points of the paper.

Reviewer: 2

In this paper, the authors attempted to provide a comprehensive definition of “representativeness” in medical research. Such bold endeavors are appreciated, especially considering the ambiguousness of the term “representative”. However, there are some areas/concerns that need to be clarified or addressed, in order to make the paper more useful. Below are my comments.

RESPONSE: We are pleased to see that the reviewer appreciates our attempt to better define representativeness. We respond to each of the reviewer’s comments below.

1. The authors may need to give more compelling explanations/rationales about why their new definition of “representativeness” is useful and important. Many people’s perception (including mine) of the term “representative” is probably in line with the (1st) definition (page 4 lines 37-39) of the 2013 IJE series papers that representativeness occurs when “study sample is a simple random

sample of the target population” through representative sampling. In the current work, the authors gave a new broad definition of “representative” that the study sample is representative when the results are generalizable in estimate or in interpretation. Especially, the authors mentioned, if the sample can be quantitatively generalized to the target population through weighting or standardizing, the study sample can still be considered representative even when the distribution of the key covariates differ between the sample and the target population. However, such definition is a bit counterintuitive and awkward. Under such definition, most study samples will be representative once the effect modifiers are measured and accounted for, and the positivity assumption for generalizing is satisfied. It deviates from people’s classic perception of “representative” and its relation to “representative sampling”, and may be difficult for readers to digest. Therefore, to make this new definition of representativeness approachable and useful, more rationales should be provided.

Importantly, I am thinking about descriptive studies estimating the prevalence of disease: if we wanted to estimate the depression rate in the overall population in the US, and selected non-random sample that is majorly male, is it reasonable to claim this sample is representative even if we are able to account for the difference in sex distribution?

RESPONSE: This is an excellent point, and we think it directly relates to the sentence now on p. 9: “While the study sample may not be representative of the target population as observed, it could be made representative using methods for generalizability or transportability.” We believe part of the confusion was the result of including the paragraph on methods to make one’s estimate generalizable in the section defining representativeness. We now include this paragraph in the new Discussion section, to highlight that we are connecting these concepts but not necessarily including them in our definition. We also edited the text in other places, such that mentions of methods to generalize are confined to this Discussion paragraph.

To answer the particular question about the descriptive study, we would say that the sample is not representative as observed (save perhaps within strata). If we applied generalizability weights, though, the resulting weighted sample would be representative, given the necessary assumptions were met.

2. The illustration of “generalizable in interpretation” is not very clear (page 6 lines 1-29). Especially, it lacks some references to support this concept. From my interpretation, this type of “generalization” is more like the generalization in “effect sign”. That is, if the effect in the target population (e.g., humans) and the effect in the study sample (e.g., lab mice) are in the same direction (e.g., positive effect), one can say the results are generalizable in “effect sign”. In my view, “generalizable in effect sign” is more intuitive than “generalizable in interpretation”. A useful reference that could be used to illustrate such sign generalization:

EGAMI, N., & HARTMAN, E. (2022). Elements of External Validity: Framework, Design, and Analysis. *American Political Science Review*, 1-19. doi:10.1017/S0003055422000880

RESPONSE: Thank you for pointing us toward the paper by Egami and Hartman. We agree that there is great overlap between their term “generalization in effect sign” and our term “generalization in interpretation.” Given many of the comments from Reviewer 1 about limiting the amount of jargon for the journal’s audience, we have decided to leave the phrase as “generalizable in interpretation,” as we believe it might be more broadly intuitive for this audience and reflective of the idea that we are interested in applying the general inference from one study to another. However, we now include a reference to the idea of generalization of effect sign when defining this phrase and cite the paper by Egami and Hartman on p. 7:

“While the estimates obtained in the sample are not quantitatively the same within a margin of error as those that would be estimated in the target population, one can hypothesize based on background knowledge that the interpretation (which could be the direction of effect, general inference from the results, or knowledge gained from an experiment) would remain the same.”

A similar reference is given on p. 11.

3. Continuing the Comment 2, as in authors definition, if the results from lab mice are generalizable in interpretation to humans, the study sample will be representative. However, it is a little bit awkward to say, the study sample of lab mice is representative of the human beings. Further, the lab mice in the study sample is probably unrepresentative in its species, as they are homogeneous with identical genes (also mentioned in the 2nd paragraph of Greenland's IJE paper in 2013). So representativeness may not be necessarily linked to generalizability. The knowledge can be generalizable when the sample is not representative.

The authors devoted majority of the text to generalizability, but what is lacking is a strong case why the definition of representativeness should be tied to generalizability, and why such definition is more intuitive and useful (similar to my comment 1).

RESPONSE: We recognize the reviewer's concerns. This paper was motivated by the experience of our author team in seeing the word "representative" being used in a number of different ways across the literature, not always with clear meaning. Furthermore, in our view, representativeness and generalizability are fundamentally connected because generalizability is the goal of representativeness, even when representativeness is defined solely as a simple random sample. When researchers make statements about whether or not they believe their sample is representative, what they are fundamentally thinking about is whether their results can be applied to other samples or populations. Results being applied to other samples, either quantitatively or qualitatively, is what we mean by generalizability. We also believe that a sample should not only be deemed representative if it is a simple random sample of the target population. A sample is also representative when the information gained from the study can be applied to the target population because the underlying scientific principles or biological processes are applicable. We also believe the distinction being made in our paper and the accompanying examples are useful for the target audience of this journal, who rely on the results of a broad spectrum of science.

We hope that many of our edits throughout the paper, as well as the particular response to Comment 1 above help allay some of the reviewer's concerns.

4. As in the text, generalizing in estimate involves both descriptive studies and causal studies. It may not be appropriate to say "effect measure modifiers" for descriptive studies (on page 7 lines 2-13).

RESPONSE: In the sections defining representativeness, we had been careful to note that we were had to account for key covariates that affect the outcome (which could be effect measure modifiers when estimating a causal effect). However, in other places we only mentioned effect measure modifiers. We have updated the text throughout to be more general.

5. Page 11 lines 47-48: "the 36.3% prevalence of SARS-CoV-2 could potentially be generalized to the full sample of people who inject drugs in this region, at least during the time frame examined." The authors may want to briefly describe the sampling method used in this descriptive study and give a reasonable speculation that it is a random sample.

RESPONSE: In that paper, the study recruited people who inject drugs through street outreach and mobile vans. The authors do not comment on whether they believe the study sample approximates a random sample of the target population. To our paper, we have added text to allow for some flexibility in the use of perhaps imperfect examples. For example, we state on p. 5:

"In Boxes 1-3, we explore these definitions in the context of other study designs. As with the RCT, we use the research question, study design, and sample description of these published works to construct a theoretical example. We do not delve into the specific details of the study and whether or not the researchers achieved what we discuss."

In Box 3, we have expanded the bullet point on generalizability in estimate to state:

“Under an appropriate sampling and recruitment strategy, the 36.3% prevalence of SARS-CoV-2 could potentially be generalized to the full sample of people who inject drugs in this region, at least during the time frame examined.”

VERSION 2 – REVIEW

REVIEWER 1	Riley, Richard; University of Birmingham, Institute of Applied Health Research. Competing Interest: None
REVIEW RETURNED	15-Feb-2023

GENERAL COMMENTS	The response and revision are excellent. The new running example, and the moving of the other examples to boxes, make a big difference. I have no further comments
-------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------