Response to Reviewers

PCI-RR submission entitled "Mapping Cross-Disciplinary Perspectives on Responsible Conduct of Research: A Delphi Study"

24 May 2024

Editor's Summary of reviewers' points:

- Clearly state the Research Questions in the Introduction (to satisfy review criterion 1A]
- Consider a minimum number for to represent the four UoA panels
- Consider clarifying the first paragraph to improve readability
- Include a protocol for 'pseudonymising' data to ensure participant's data remain anonymous
- Link the initial reference document in the main manuscript (and not just within the response to reviewers)
- Include the rules/processes applied to the inclusion of new dimensions
- A personal reflexivity (as recommended by Thomas Evans) [although I note that these are present in your supplementary materials on your OSF Project Page, so you may wish to simply refer the reviewer to this or make this point more salient in your main manuscript).
- Consider and mitigate risks associated with breaking down results by discipline

Editor's personal reflections:

I.   The Abstract states "Responsible conduct of research (RCR) is generally agreed to be a laudable goal", but I would argue it is a necessary and imperative goal given that research is a public good. I recommend reconsidering the use of the word 'laudable' within this opening sentence.

Thanks for pointing this out. The word "laudable" indeed is a severe understatement in this context. We have rephased the opening sentences as follows:

"*Responsible conduct of research (RCR) is imperative for the quality and trustworthiness of research, and the health of the whole research system. However, it is as yet unclear to what extent RCR differs across disciplines*" (p. 2).

II.   The following part of this sentence within the Abstract is also a little cumbersome: "then will proceed with several rounds of rating the importance of each dimension to particular discipline".

We have redrafted the relevant section to give a clearer and more concise summary of our research:

"*We will conduct a modified reactive Delphi, in which panellists progressively refine their judgement of the importance of individual dimensions of RCR to their respective discipline, starting from a provisional list of dimensions derived from previous literature and interviews*" (p. 2)

III.   You state within the eligibility criteria that participants must have published at least one article including the following key words: "RCR", "RRI", "responsible research and innovation" "research integrity" or "responsible research". I understand this is based upon your scoping review, but do you not think that the key words of 'open research', 'open science', and/or 'open scholarship' could also be applicable here?

While we recognise that there is overlap between terms like open science and RRI, there are important differences. For one thing, the RRI and research integrity concepts predate the somewhat recent 'open science' movement (we do not refer to the earlier open access push in the 1990s by people like Peter Suber, but the 'crisis of confidence' related open science movement that is currently receiving much research funding and attention), and they originate in slightly different (though again, overlapping) communities and disciplines. They also tap into slightly different concepts and dimensions. Open science and RRI/RCR are at different stages of academic embedment - while RRI/RCR are well established, open science and its movement are still emerging to some degree. While openness and related concepts are highly relevant to RRI and our RCR framing, they're only one of many aspects. We want to collect a sample of participants that are across the RRI framework more broadly so they have an overview of the whole RRI landscape. Not only that, but open science and the related movement is relatively new (since somewhere around 2010/2011 most literature estimates) and quite popular currently. Many people are familiar with and advocate open science now, as a result of the momentum of the movement and the associated community. Not all of those, who include open science and related keywords in their online profiles, will be experts in RRI/RCR. Many won't even be experts in open science. The power and informativeness of a Delphi lies in the panel and how it is chosen, and we do not wish to inadvertently dilute its quality by including people who might be familiar with one single aspect of RRI/RCR but are not generally thought of (nor think of themselves) as experts in the overall existing scholarship on RRI/RCR. This certainly keeps our pool narrow, and it will be difficult to ensure we have the kind of panel we would like and which would be methodologically optimal, but in our opinion that is preferable to having a diluted and 'skewed-to-open research' panel.

However, we recognise that the inclusion criteria, as currently written, are perhaps too stringent, potentially leaving out people in fields that don't use terms such as 'responsible' or 'research integrity'. For that reason we have included a 'secondary' tier of inclusion, where people involved in some particular aspect of RCR (open science, ethics, public engagement) might be recruited IF they agree, when consulted, that their expertise in RCR extends beyond just that one aspect. In that regard, these 'secondary' keywords for inclusion will be a starting point for us to identify potential participants who may have relevant knowledge but not use the main keyword terms.

Therefore the amended text about keywords now reads:

"*Relevant keywords include "RCR/responsible conduct of research", "RRI/responsible research and innovation", "research integrity", "responsible research". We will also consider using keywords pertaining to particular areas of RCR (such as "open research", "open science", "research ethics", "public engagement", or any of the other aspects identified in our initial reference document) as ways to identify potential participants, but for these more particular terms we would require that, upon consultation, the potential participant provides additional justification that they have expertise to answer the Delphi on RCR more generally*" (p 12).

IV.   There seems to be an inconsistency in the section on panel size. In the first paragraph you state "Ideally, we would recruit two to three members of each of the 34 UoA's, in order to maximise the chance that the outcomes of the study are in fact disciplinary differences instead of personal differences", which would mean approx' 68 participants invited (2 from each of the 34 UoA's). However, in a later paragraph you state: "As such, we have set the minimum panel size at the start of the process to be 30 panellists." If, as you suggest, you successfully recruit 30 panellists, do you think you can rule out that the results will represent disciplinary differences and not personal differences? Can this section be clarified?

Our apologies for the confusion. We have shortened and clarified this section in the current revision. Please find the added relevant paragraph pasted below:

"The ideal panel size in a Delphi study is largely dependent on the purpose of the particular study. We are using this Delphi study to map the disciplinary differences regarding RCR. As such, it is important to reasonably minimise the chance that differences in judgements between panellists stem from their personal convictions rather than their disciplinary backgrounds. One way to do so is to ensure we have a sufficient panel size, to balance out differing opinions. However, we will also mitigate this risk of noisy data by the fact that we recruit experts, whose views should already take into account their familiarity with the range of differing opinions in their field. Additionally, we explicitly ask participants to answer based on their knowledge of the field, rather than their personal beliefs. Finally, since the primary purpose of our study is to investigate diversity between disciplines in general, rather than making inferences about specific disciplines, we do not consider a minimum number of panellists per discipline to be necessary. Of course, the expected increase in reliability and representativeness of an increased panel size must be weighed against the feasibility of the study, both in terms of difficulty of recruitment and retention, and in terms of time and resources required to analyse (qualitative) data and draft the feedback reports. As such, we have set the minimum panel size at the start of the process to be 30 panellists" **(p. 13).**

V.   Recall that at Stage 2, the main aspects of the Stage 1 manuscript cannot be revised/changed (other than, for example, minor tense changes from future to past). For this reason, you may wish to reconsider how the section titled "expected results" is written.

Thank you. We have made some minor changes to this section (e.g., removing speculations about what the data might look like). Now, the only changes we will anticipate making are tense changes or updating the reference number for tables and figures, etc. It now reads:

"Our results will be structured as follows: first, as an overview of the main findings, we will provide a table containing all final dimensions from Phase 2 of the Delphi process, noting which dimensions reached stability, and outlining the final variances and interquartile ranges, as a proxy for consensus, and the median rated importance, as an indication of importance (see Table X, below). This table contains data from the entire panel, across all represented disciplines. On our 1-7 importance rating scale, we will interpret median ratings corresponding to numerical values

lower than 3 as low importance, between 3 and 5 as moderate importance, and greater than 5 as high importance. These straightforwardly map onto the verbal labels of the rating scale.

However, we will not declare any a priori thresholds for interpreting levels of consensus in the data. While it is imperative to state specific thresholds for interpreting data in a Delphi study, if that study aims to attain consensus (Grant et al., 2018; Williams & Webb, 1994), our goal is different. We aim to refine and map the existing perspectives on dimensions of RCR, rather than determine a consensus of which dimensions are most important. Indeed, multiple authors have argued that the emphasis on a binary consensus-nonconsensus divide is a crude way of employing the Delphi technique. For example, Scheibe and colleagues note that "considering that there is a strong natural tendency in the Delphi for opinion to centralize, resistance in the form of unconsensual distributions should be viewed with special interest" (2002, p. 271), see also (Linstone & Turoff, 2011).

As such, we do not propose any confirmatory analyses for declaring 'consensus' for any given dimension, but will instead present the quantitative data descriptively. We are interested in at least three categories of response distributions: either a strong peak around a single point (i.e., universal agreement), a relatively flat distribution across all points, or a multi-modal distribution with two or more distinct peaks. We do not propose to differentiate these statistically, given the relatively small sample sizes and exploratory nature of this work, but instead will present our descriptive interpretations along with visualisations of the complete importance rating data for each dimension, so readers can visually assess the shape of distributions to infer the 'universality' of responses across the panel (see Figure(s) X, Y, Z, below).

In addition to the primary results above, we will also provide, for context, a table providing a complete view of the Delphi results broken down by discipline, reporting the median rated importance of each discipline category per dimension and the number of respondents per discipline category. Since our sample will include only a small number of panellists per discipline category, we must be cautious in making any inferences on the basis of these analyses, and advise readers from making any strong inferences either.

To contextualise all the abovementioned quantitative findings, we will also provide an analysis of the qualitative data from throughout the Delphi process for all Phase 2 dimensions" **(pp. 24-25)**

Reviewer 1

The authors have responded to my previous comments adequately, and I was pleased to be able to review the additional materials that were omitted from the last round of review. I wish them good luck with the study!

Thanks again for the helpful feedback!

Reviewer 2

This manuscript is a revision of a Stage 1 proposal for which I previously served as a reviewer. The authors have done an outstanding job with the revision, clearly addressing the broad concerns I had

with the previous version. The revision, to me, is clear and compelling. I had just two small suggestions:

2.1   Registered Reports (well, all types of reports) should have clearly stated research questions. In the current version, I sort of had to piece together and infer the RQs, and then it was not until reading the expected results section that I got a firm sense for what the RQs are, namely to a) map the existing dimensions with respect to importance, and b) describe any variations in this mapping by field. Stating these RQs in the Introduction would help clarify the purpose and contribution of the project. Alternatively, if I have the RQs wrong, then the analysis plan need to be modified—this is why clear RQs are needed, as without them the project is difficult to evaluate properly.

Thanks for pointing this out. Indeed, any study's research question should be clearly apparent from reading the introduction. We have condensed the introduction and have clearly stated our research goal. Please find the relevant paragraph pasted below:

"*In this study, we aim to build upon previous literature and guidelines to construct an updated collection of dimensions of RCR and a mapping of how important they are to different research disciplines, including those traditionally underserved by existing frameworks. Using the Delphi technique of structuring communication, we will enlist a multidisciplinary expert panel to help us map the manifestations of RCR in different disciplines*" (p. 5).

2.2   The sampling process is now much clearer. However, it is still possible that with a 15-discipline minimum only two of the four panels could be represented. Having no representation from some of the panels would severely compromise the impact of the project. I suggest adding a layer, in which there should be some minimum number (3?) from each of the four panels.

Thanks for this suggestion, which we kindly followed. Please find the relevant revised paragraph outlining this addition pasted below:

"*To avoid disciplinary bias in our sample, i.e., where a disproportionate amount of experts would have a background in a specific discipline, we also decided to include two measures of diversity. First, we will require a minimum amount of disciplines represented in our sample before we start the Delphi study. As such, the starting panel must represent a minimum of 15 disciplines, with no more than three participants from one single discipline. Second, we will require a minimum amount of three participants from each of the four "panels" of the REF UoA structure that we used as our guide for disciplinary diversity (the four panels roughly correspond to: health and life sciences, natural sciences, social sciences, and arts and humanities)*" (pp. 13-14).

Congrats on a fine project, and I look forward to seeing the results!

Reviewer 3:

I have read the revised version and the authors' responses. The authors have carefully responded to feedback and in my view the study is now ready to be carried out.

3.1   If other reviewers or recommenders end up asking for more revisions, I would like to leave a short cosmetic note: while I can appreciate the poetic tone of the very first paragraph, it is also a bit difficult to follow (at least for this non-native English reader) and might be a good to streamline -- or

if the word count matters -- even remove. This minor comment on style can naturally be ignored and it's certainly not worth considering unless another review round is commenced for other reasons.

We followed your suggestion, and for the sake of concision removed the first paragraph.

All best wishes for the work and looking forward to read the Stage 2,

Reviewer 4:

Thanks for the opportunity to re-review this exciting Registered Report. It was great to see the amount of support and constructive feedback invested in your work on the previous submission! The revisions have contributed to a much more refined proposition and I am grateful for the positive approach to the feedback the authors have taken. You will be pleased to see that my response to the manuscript is significantly reduced in volume since last time, and this is much a reflection of the work conducted to address the comprehensive range of ideas and concerns raised by the previous round of reviews. I feel like many of the larger concerns have been directly addressed and that the work as a whole represents a valuable contribution to our understanding.

A few further ideas to consider:

4.1   A protocol for 'pseudonymising' data that might identify participants is absent. I appreciate it's unlikely based upon the style of questioning but being a little clearer about what you will and won't edit to ensure it's easily sharable might cover off concerns later.

We agree this would be a valuable addition to our manuscript. Both for transparency towards the reader, and guidance for ourselves. We have added the following paragraph explaining exactly what actions we will take to anonymise the data if necessary:

***"Specifically, this means that the feedback reports will not contain names of persons, institutions, or geographical locations that may be present in the qualitative data. In addition, if we determine that a particular section of text in the qualitative data – such as a description of an institution or location – allows the panellist to be identified, we will delete this as well. All deletions will be clearly marked"* (p. 21)**.

4.2   I can't see the initial reference document signposted in the main manuscript (apologies if I've missed this!)– I can see it in the response to reviewers document so it should be in the main document. This looks to be a helpful resource in of itself and I can see lots of different applications of this so I'm hopeful that the final work will produce an even more insightful mapping of ideas. I think personally I still have concerns about some of the blurred lines of scope of constructs/dimensions included but I'm not recommending any action based on this as I think it's probably an inherent part of the landscape.

Thanks for pointing this out. We have added a link to the initial reference document in the main manuscript, see the following addition:

***"We plan to use a so-called 'reactive' Delphi method, in which panellists respond to an pre-constructed reference document (available on https://osf.io/jrf47) instead of creating one themselves (Salkind, 2007, p. 243)"* (p. 7)**.

We acknowledge the reviewer's concerns over the blurred lines between the dimensions of RCR, and agree that this is likely an inevitable aspect of engaging in any type of discussion around the topic.

4.3   At stage 2 – inclusion of new dimensions will be at the authors' discretion – will there be any rules/processes applied here?

Again, we appreciate the reviewer pointing out potentially unclear aspects of the protocol. This will likely make the process of executing the study much easier. We have (re)drafted and included the following paragraph outlining our protocol of accepting proposed new dimensions in Phase 2:

"Note that in Phase 2, participants will not be asked to add new dimensions. Should participants still wish to add dimensions, despite this, they may indicate so in the response boxes provided. These suggestions will be screened by the research team on the basis of the same two points as after Phase 1, as outlined above. However, the barrier for including new dimensions during Phase 2 will be much higher, as addition of new items at this point in the process substantially complicates the process, risking higher attrition and discontinuity. As such, new proposed dimensions must contain a highly compelling argument that it is sensible and non-redundant. What is and is not required of participants for each phase will be clearly communicated to them" **(p. 20).**

4.4   I really like the idea of participants at each stage being able to look at the 'analysis' or interpretation whilst also being able to cross-reference this with the raw data – this is a really transparent and effective way of engaging with this and I really appreciate the attempts to increase openness in this type of design.

We are very happy that the reviewer likes our proposed method. We are looking forward to putting it into practice.

4.5   I appreciate the inclusion of a reflexivity section but would perhaps encourage the authors to remove the list of citations and instead reflect more upon the factors that have influenced the design, and will influence the implementation, of the work. For example, the fact that this is part of a larger project and with a bigger broader goal – what influence has that had upon the study? You already work in this space – what role will personal networks have in recruitment of participants and might your reputations and position in the field influence the results you get?

Thanks for your suggestion on improving the reflexivity/positionality section. We have indeed followed the suggestion of removing the list of citations, and added a reflection on the potential bias resulting from our reputation in the field. Though adding the context of this study in the larger research project here might also have been suitable, we believe it fits better in the introduction (see the last paragraph of the introduction), and we would like to minimise repetitions. Also, please note that we refer to individual positionality statements (available on https://osf.io/prvds), in which we explore our positionality in much greater detail than would be possible in the protocol. The revised section reads:

"Finally, we wish to be transparent about the contributions that we, as individuals and as a team, approach the subject of RCR. This allows the reader to evaluate our decisions with personal context in mind given the flexibility that exists in our design; especially as the dimensions are developed between rounds. Our team comes from a background in science and technology studies,

metascience and research integrity, and we have previously published on the topics of epistemic diversity, responsibility and quality. This means that the experts that we invite to participate in the panel may be familiar with our work. This contains a risk of biasing the study, since their decision of whether or not to participate may depend on their estimation of our work. However, the direction of this potential bias is uncertain. In addition, this risk is mitigated greatly by the fact that the study population is large and diverse, and its majority will have had no professional or personal interaction with any of the researchers. In our supplemental materials on OSF (https://osf.io/prvds), we provide individual statements about our positions in relation to the present study, structured using orienting questions proposed by Barry et al. (1999) and Olmos-Vega et al. (2023), to further highlight our link to the research our group is conducting" **(p. 23)**

4.6   In the second-to-last paragraph you suggest breaking down results by discipline. We already get a sense of disciplinary differences by the reporting and visualisation of median 'importance' scores so I'm interested to hear more about what is considered the purpose of being explicit about where each discipline stands on each specific dimension in this specific manuscript and whether any negative consequences could be anticipated and mitigated. I can see the value for it being in a supplementary or such but no case has been provided for why this should be a priority in the main manuscript given the primary research questions have been addressed by the previous analyses. There's the scope for this to be used to misrepresent fields etc. given small numbers and different practices/approaches and it'd be interesting to get your response to this potential for negative impact (I imagine this won't lead to any change to the manuscript but in the case that this might be a helpful thought I've left this here to be considered).

This is an important point, thank you for bringing it up. We agree that it is important to keep readers from inferring too much certainty from our small sample sizes. However, we do think it is valuable to present the full findings within the main manuscript. We will mitigate the risks mentioned above in a few ways:

- **In the paragraph mentioning the disciplinary breakdown analysis, we have explicitly made it clear that this is not the primary analysis, but rather a presentation of the full data for broader context. It now reads:**
  - *"In addition to the primary results above, we will also provide, for context, a table providing a complete view of the Delphi results broken down by discipline, reporting the median rated importance of each discipline category per dimension and the number of respondents per discipline category"* **(p 25).**
- **We also have added text advising readers not to make strong inferences based on these data, given the small sample sizes:**
  - *"Since our sample will include only a small number of panellists per discipline category, we must be cautious in making any inferences on the basis of these analyses, and advise readers from making any strong inferences either"* **(p 25).**
- **We will add text in the table description as well, highlighting the inability to make strong inferences from these small sample sizes**
- **We will also plan to discuss this issue at length in the discussion section. We will note that:**
  - **The small amounts of data are mitigated by the fact that 1) we are asking experts, whose views on their field should be well founded, and 2) in the methods we explicitly ask them to answer for their field, not their personal opinions**

- ○ **However, no experts are perfect, and therefore readers should take these data only as a starting point, not as a definitive reflection of these disciplines.**

Some very minor things:

4.7   Awkward wording on p4: "This broader remit of RCR includes dimensions that overlap with those of RRI, such as the responsibility research has for honest and transparent dealings with citizens and society."

We have reworded this slightly to make it clearer: "*On the other hand, RCR includes dimensions that overlap with those of RRI, such as the obligation of research to deal honestly and transparently with citizens and society*" (p. 4).

4.8   Some of the new content on p6/7 is a little awkwardly presented e.g., "lastly" which doesn't seem to follow a "firstly". It might be worth considering whether some of the context of the wider project might be better situated in the methodology (you later introduce it in p14) or reframed such that it helps flow from the established needs of the field as you discuss earlier.

Thanks for the suggestion. We considered moving the paragraph to the methods section. But since it serves to contextualise the *aims* of the study, and not the *methods* with which we will accomplish these aims, we concluded this would in fact be the best place for it. We hope you understand these considerations.

In addition, we agree that the added content was a bit awkwardly connected to the preceding paragraph. We have replaced the word "lastly", with "finally", which we hope flows better. The word itself refers to the fact that this paragraph wraps up the introduction.

4.9   You refer to dissensus in p15 when it's no longer discussed elsewhere so this may be confusing or seen as an inconsistency for readers.

Thanks for pointing this out. In our revision of the paragraph concerning sample size, following the editor's point IV above, we have deleted the entire sentence containing the reference to "dissensus". FYI, the paragraph now reads:

"*The ideal panel size in a Delphi study is largely dependent on the purpose of the particular study. We are using this Delphi study to map the disciplinary differences regarding RCR. As such, it is important to reasonably minimise the chance that differences in judgements between panellists stem from their personal convictions rather than their disciplinary backgrounds. One way to do so is to ensure we have a sufficient panel size, to balance out differing opinions. However, we will also mitigate this risk of noisy data by the fact that we recruit experts, whose views should already take into account their familiarity with the range of differing opinions in their field. Additionally, we explicitly ask participants to answer based on their knowledge of the field, rather than their personal beliefs. Finally, since the primary purpose of our study is to investigate diversity between disciplines in general, rather than making inferences about specific disciplines, we do not consider a minimum number of panellists per discipline to be necessary. Of course, the expected increase in reliability and representativeness of an increased panel size must be weighed against the feasibility of the study, both in terms of difficulty of recruitment and retention, and in terms of time and resources required to analyse (qualitative) data and draft the feedback reports. As such, we have set the minimum panel size at the start of the process to be 30 panellists.*" (p. 13).

4.10   In sum, I see the value of the proposed work and the focus and processes involved have been significantly refined from the previous version. I hope that these comments are helpful for the project, I have thoroughly enjoyed reading, reflecting and getting stuck-in with the different components of this research, and as before, I look forward to reading the next version of this work whether that be as reviewer or (hopefully!) reader!

It is great to hear that you enjoyed reviewing our work as much as we have implementing your suggestions!