Mapping ~~the Scope of~~ Cross-Disciplinary Perspectives on Responsible Conduct of Research ~~in Practice:~~

:

A ~~MAD (Modified ReActive Dissensus)~~ Delphi Study

Stage 1 Registered Report proposal

Sarahanne M. Field, Jackie Thompson, Tom van Drimmelen, Jennifer Ferrar, Bart Penders, Sarah de Rijcke, and Marcus R. Munafò

1

Abstract

Responsible conduct of research (RCR) is ~~essential to the trustworthiness of research.~~ generally agreed to be a laudable goal. It promotes high quality research practices, which should lead to more credible findings, and instill confidence in the research community. However, it is as yet unclear to what extent RCR differs across disciplines. Currently, many approaches to research and training in RCR are ~~discipline specific. For example, reproducibility is a concept that applies to quantitative disciplines, but less so qualitative disciplines and the social sciences, and even less so in the humanities.~~ either generalised across all disciplines, at a high level (e.g., international frameworks on research integrity) or at the other extreme, discipline-specific. Relying on the expertise and knowledge of a carefully selected multidisciplinary panel of RCR scholars and practitioners, this Delphi study aims to expand the current (underspecified) frameworks of RCR to develop a more diverse and comprehensive concept of what constitutes RCR across disciplines, along with a ~~framework that captures this updated understanding.~~ mapping that captures this updated understanding. The Delphi process will begin with participants refining a provisional list of dimensions of RCR collated from previous literature and interviews, then will proceed with several rounds of rating the importance of each dimension to particular disciplines. [After completion of the study, we will report the details of participant numbers, rounds of Delphi, and a summary of results here.]

CRediT statement: [to be completed in final submission]

The research enterprise is an impressive, powerful engine, capable of generating huge global change as well as catastrophe, as it deals with uncertain facts, disputed values, and high stakes in the context of a need for urgent decisions (see Funtowicz & Ravetz, 2018). When scientific research becomes misaligned with the needs and expectations of society, at best, it *becomes irrelevant*, failing to adequately meet society's needs, and falling short of the promises it makes. At worst, it *becomes unethical*, putting the lives of people and the environment at risk.

Over time, a growing awareness of the heavy responsibility the scientific enterprise carries has led to the development of frameworks which underpin individual and institutional codes of research conduct. Almost 30 years ago, 'ELSA' (**E**thical, **L**egal and **S**ocial **A**spects of emerging science and technology) was first introduced as a framework for research development and funding, to reorient scientific research practices to make ~~it~~them more effective and – crucially – more ethical and self-aware. Due to its origins, state Zwart, Landeweerd and van Rooij (2014), ELSA applied largely to disciplines such as genomics and other life sciences, bioethics, science and technology studies, technology assessment, philosophy of science, and science communication. Since then, the framework has been built upon and developed, making way for what is known as the Responsible Research and Innovation (RRI) framework, which is more widely applicable and more focused on addressing larger-scale socio-economic challenges than its predecessor ELSA was.

As with ELSA before it, RRI and, derived from it, responsible conduct of research (RCR)[1] is ~~institutionalized~~institutionalised as an instrument of policy, rather than being a

---

[1] Although the RRI framework involves both research and innovation, our focus is primarily on the research conduct aspect of the framework and will henceforth refer to RCR, the responsible conduct of research.

discipline in and of itself (Tallacchini, 2009). It refers to a methodological attitude to be applied to research conduct: a strategy to change how research is conducted in practice, regarding its responsibilities to society. Von Schomberg (2013) points out that there is no agreed-upon definition of what RCR is; rather, it holds an invitation. Rather, he invites readers to discussconsider what RCR as a top-down signifier might in fact denote, in relation to the disciplines and research processes with which it engages with. Despite this declaration, on page 9 of a 2012 article, Von Schomberg does explain RR(I) inpropose a somewhat concrete terms,definition of RR(I), describing it as "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)." Narrowing the scope of the definition for the current study, we consider responsible conduct of research (RCR) to be a topic that requires the synthesis of many disparate aspects. While it is distinct from concepts such as research integrity (RI) or responsible research and innovation (RRI), it is no doubt closely related. We argue that RCR casts a broader net than the typical definitions of research integrity, that is, promotion of confidence and trust in research and the research process. 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. However, while conceptualisations of RRI typically include impacts of technological innovations and research output on society, RCR concerns the subset of dimensions or responsibilities relating to the activities involved in conducting research.

　　Underpinned by a recent scoping review conducted by our research team (Field et al., 2023; the preprint is available at https://osf.io/6z4mx/),2024), in alignment with the arguments of many scholars and organizationsorganisations gone before, we contend that

4

RCR will promote the co-production of robust and reliable knowledge, can be used to foster a healthy research culture, and ultimately will lead to the public's trust in the research process, its output, and its implementation. In short, conducting research activity responsibly is crucial to the health of the knowledge generation enterprise going forward.

In that same scoping review study, with the view of eventually developing our own RCR framework, we made steps toward determining how RCR might vary across disciplines, and what elements ~~should stay constant.~~are shared. We found that although some dimensions of RCR are more applicable to some disciplines than others (e.g., anticipation and transparency), others were applicable across many fields (e.g., integrity). We concluded that communities of practice built around future RCR frameworks might benefit from updating and diversifying existing frameworks such that they allow room to accommodate different epistemological traditions and are sensitive to emerging fields which operate at the forefront of scientific advancement (such as genomics), or which are especially prone to ethical dilemmas (such as artificial intelligence) or both (such as nanotechnology).

This proposal concerns a study that builds upon the foundation laid by our scoping review. Namely, we plan to conduct a Delphi study which, with the help of a multidisciplinary expert panel, will help us pin down the most salient and crucial elements of a new ~~RCR framework that balances breadth and specificity with feasibility and practicality.~~mapping of RCR that descriptively lays out the key differences between the manifestations of RCR in different disciplines. The Delphi process will help us identify core tenets of RCR that go beyond the disciplines central to previous RRI frameworks, along with more niche, discipline-specific elements. We aim to develop a framework, using these RCR dimensions, that reflects the needs of the academic community, with the explicit intention of representing as many disciplines as possible. ~~We will use this input to establish our own RCR framework; however, the~~We aim for our mapping to fill the gap between the two extremes

5

that existing conceptualisations of RCR tend to fall under: either high-level frameworks designed to be universally applicable across all disciplines (e.g., the Singapore Statement on Research Integrity, the Australian Code for Responsible Conduct of Research, or the All European Academies European Code of Conduct for Research Integrity[2]), or prescriptive guides tailored to the practical instruction of researchers within a specific discipline or field (e.g., RCR training designed for members of a university department as part of a degree or continuing professional development, or mandated by funders such as the National Institutes of Health, or guidance from discipline-specific learned societies such as the Society for Improvement of Psychological Science). The findings of this Delphi study will also be relevant to subsequent research on the dimensions of RCR. In what follows, we outline our plans for developing this list of RCR dimensions and the subsequent framework. We note that while our goals involve the development of a novel RCR framework, the current registered report focuses only on the Delphi study. The outcome of the Delphi study will be the basis of the development of the later framework, which will be central to a separate, future article and details on how the framework will be developed are thus outside of the scope of this registered report.

Note that we do not share any explicit hypotheses at this juncture, where one might expect them. This is because the proposed study has the goal of developing anour list of RCR frameworkdimensions, rather than testing any specific theory, and is exploratory and descriptive as a result. Providing that our methodology is sound, and the Delphi carried out as planned, valid and reliable results could take many forms.

Lastly, it is also important to consider the role of this study as it forms a component of a larger multi-year project, which aims to broadly develop a diverse understanding of how

---

[2] The Singapore Statement can be found here: https://www.wcrif.org/guidance/singapore-statement and ALLEA's code of conduct here: https://allea.org/code-of-conduct/

RCR is conceptualised and applied across different research disciplines. The conceptual mapping we will have co-produced with the help of RCR experts during this Delphi study will form a scaffold for interaction with communities of practice in the project's latter half, helping them to contextualise where particular disciplines and practices sit in relation to others in the overall ecosystem of RCR. Its broad remit will also help to spotlight the perspectives of disciplines that have been more peripheral in discussions and evaluative frameworks on RCR so far. Thus, we aim for our mapping to stimulate a more nuanced understanding of cross-disciplinary conceptions of RCR within the communities that work to embed practices *in situ*. While our output can be used more broadly to assist other interested entities (such as individuals or research groups) in conceptualising and applying RCR principles for their own needs, that is a secondary purpose. Therefore in our aim to generate a tool that presents a wide perspective on the RCR sphere, we have cast our epistemic net broadly.

The Present Study

To try to develop a list of RCR dimensions that are practically applicable and relevant to a wide range of scientific disciplines without the explicit involvement of members of the scientific community would be remiss. In this study, we aim to progress that list of build upon previous literature and guidelines, crystallising distributed community knowledge into an updated collection of dimensions further, determining which of the of RCR and a mapping of how they apply to different research disciplines, including those traditionally underserved by existing already-considered elements of RCR should be included in our RCR framework, and which we might be missing (or which might be underemphasized).frameworks. To achieve these goals, we plan to assemble a disciplinarily-diverse Delphi panel of RCR experts to advise on a previously devised, a selected group of individuals with experience in

7

RCR frameworks, scholarship and/or practice. A pre-constructed, reference document outlining a rough draft of , comprising a new RCR framework (whose elements were established using information derived from both our scoping review, and interviews with a range of proposed list of RCR scholars and practitioners). dimensions, will be sent to our Delphi panel,. The panel will be invited to suggest additions to the list, and later asked to judge each RCR dimension in the revised list on its importance to RCR within the panellist's discipline of expertise.

The Delphi method (which some consider to be a sequential mixed methods study design) is a descriptive, exploratory approach commonly used in many scientific disciplines.The Delphi technique is a method of facilitating and structuring group communication processes (Linstone & Turoff, 2002, Linstone & Turoff, 2011). The procedure aims to get a refined set of participant evaluations on a set of 'target items' (in our case, potential RCR dimensions for inclusion in a practical framework) and is known for its 'iterative' nature. The participantsParticipating panellists provide judgmentsjudgements (depending on the specific approach, these might be quantitative, qualitative or both), which are collected by researchers, summarizedsummarised, and sent back to participants for another judgment round.subsequent judgement rounds. This evaluation-summary process is repeated for multiple rounds (a minimum of two rounds is recommended by Melander's 2018 review; however, depending on the specific approach, up to five or more may be required), and ends when researchers are content with the resulting item list, or when some other stopping criterion (such as a pre-registered number of rounds) is reached.

We have chosen the Delphi approach as it accommodates the inclusion of a broad range of participants, and is likely to yield results that reflect the whole group's opinions (rather than capturing just the perspectives of the most vocal participants). It also ensures relative anonymity of responses, which tends to allow participants to disagree with one

another more freely. ~~We plan to use a so-called 'reactive' Delphi method. In a reactive Delphi, participants respond to an *existing* reference document in contrast with other approaches in which the participants work to *construct* a list across the various rounds. We will modify the reactive Delphi somewhat, however, such that participants will both respond to a pre-constructed reference list and also will be encouraged to build upon it with their own diverging ideas. An initial reference document will be constructed by the authors prior to conducting this study for use in the proposed panel; a new, Delphi panel-informed list (upon which we will base our RCR framework) is the goal output.~~

~~At this juncture, we emphasize a further departure from the typically practiced Delphi. While typical~~We plan to use a so-called 'reactive' Delphi method, in which panellists respond to an *existing* reference document instead of creating one themselves (Salkind, 2007, p. 243). However, in order to make the best use of their expertise, our Delphi panel will be encouraged to build upon the initial reference list with their own diverging ideas, suggesting missing RCR dimensions.

While most Delphi approaches are consensus-based, meaning that they aim to converge on a selection of important elements of a reference document~~, our approach introduces elements of dissensus. That is, we wish to maximize the breadth of different ideas about salient RCR dimensions, with participants commenting on existing elements of the reference list as well as providing recommendations for new elements, as we mentioned above. Our pre-constructed reference document of RCR dimensions will be sent to our Delphi panel, a selected group of individuals with experience in RR(I) frameworks, scholarship and/or practice. The panel will be asked to judge each RCR dimension in the list based on its importance to the participant within their disciplinary contexts of research practice; panelists will also be invited to suggest additions to the list. Henceforth, we refer to the version of the Delphi that we have devised for this study as a *MAD Delphi*; a Modified~~

Formatted: Font: Times New Roman, Font color: Auto

Formatted: Font color: Auto

9

reActive Dissensus Delphi. This method is based on an earlier published Delphi study, in which SMF was an author (Pittelkow et al., 2023). Pittelkow and colleagues successfully used a modified reactive Delphi method to establish a checklist for communicating the rationale behind conducting replication studies and provides a kind of proof of concept for the methodology we set out in this proposal. (Diamond et al., 2014; von der Gracht, 2012), our approach aims to map and refine the existing breadth of perspective on various dimensions of RCR. It is important to note that though this does constitute a departure from the typically practiced Delphi, this actually is a reversion to how the Delphi process was originally intended. Multiple authors have commented that the value of a Delphi study lies exactly in mapping the distributions of opinions instead of generating consensus (Scheibe, 2002; Linstone and Turoff, 2011).

Our Delphi study is both based on existing literature and perspectives (which are captured in the reference document), as well as being highly exploratory in that it seeks to chart out expert perspectives on a broader, multi-disciplinary RCR not yet represented in frameworks and guidelines.

Method

Ethics approval for this proposal was granted by the University of Bristol's School of Psychological Science Research Ethics Committee (reference number 12071), in the United Kingdom (October 14th, 2022) and by the University of Leiden, the Netherlands (January 19th, 2023).

Initial Reference Document: A Proposed List of RCR Dimensions

Before writing this proposalconstructing the reference document, i.e., a proposed list of RCR dimensions, the authors conducted a scoping review of the existing RCR literature (Field et al., 20232024) and interviews with RCR scholars and practitioners. Both the (see

the interview guide at https://osf.io/xv98y). The interviews consisted of 10 one-on-one interviews and one focus group which included two moderators and eight participants. The articles included in the scoping review and ~~the~~ interview transcripts were subject to a thematic analysis~~. SMF,~~ conducted ~~these for the literature review and the interviews. They yielded a series of~~ by SMF, which involved coding topically salient sections of text and combining related codes into themes (see the appropriate appendix for a detailed description of this process at https://osf.io/jrf47). This analysis generated a series of overarching themes which reflected salient dimensions of RCR from the literature on the topic~~, as well as from~~. These dimensions are the core of the ~~opinions and experience of researchers, research support staff and others involved in RCR practice. The reference document contains these dimensions, along with~~ initial reference list, to which dimensions taken from existing, older RCR frameworks ~~(~~were added (i.e., the Singapore Statement on Research Integrity~~,~~ ~~the Montréal Statement,~~ and the ~~All European Academies European~~Australian Code ~~of~~ for Responsible Conduct ~~for~~of Research ~~Integrity).³~~.).

It is important to note that the choice of dimensions, and their respective definitions, are not intended to represent an authoritative list of dimensions of RCR, nor the only way to carve up these concepts. Rather, they were designed to maximise the information gain from the Delphi process, by covering a broad range of concepts and reducing redundancy. For example, the dimension "integrity" carries many different connotations and facets, and many of these are already covered by other proposed dimensions gleaned from the research literature and interviews, such as "rigour," "transparency," and others. Therefore, we defined the dimension of "integrity" to cover a more constrained facet not already mentioned,

---

³ ~~The Singapore and Montréal Statements can be found here: https://www.wcrif.org/guidance/singapore-statement and here: https://www.wcrif.org/guidance/montreal-statement; the Hong Kong Principles here: https://www.wcrif.org/guidance/hong-kong-principles, and ALLEA's code of here: https://allea.org/code-of-conduct/~~

concerning the possession of and adherence to strong moral principles. We also aimed to avoid vague or overly broad definitions, as these would not allow us to know which aspect of a multi-faceted dimension the panellists were responding to. The purpose of this study is not to come up with a consensus definition of any of these concepts; instead, the definitions are intended to make sure the concept space is covered adequately, and that participants are clear about the concepts they are rating. This caveat is included in the instructions to participants.

Similarly to the Singapore Statement, this ~~document~~initial list of RCR dimensions contains both ~~a short list of~~ core principles, or 'metaresponsibilities' (which we defined in Field et al~~.~~., 2024, as somewhat more diffusely defined principles that guide RCR practice and relate to~~/~~ or influence many other dimensions of RR), as well as ~~a list of~~ more concrete, practical, and specific responsibilities. We will not make distinctions between these different 'levels' of dimensions in the reference document, however. We do not wish to prime or ~~sensitize~~sensitise participants, influencing them to weigh items more or less heavily than they would have done in isolation. While some frameworks such as the Singapore Statement do make these distinctions (and we expect that most participants will be at least aware of these), much of the literature, including articles by some participants, does not. The relevant scholarship on the topic is far from concluding that there even *are* different levels of applicability or concreteness of responsibilities. Thus, we argue that it makes the most sense to keep the ~~reference document~~ structure of the initial list simple and allow the participants to judge dimensions as naturally as possible. Any structure we impose on the output list (i.e., the document resulting from the Delphi process) will occur as a part of later work on producing a novel RCR framework, and further discussion of it is outside the scope of this study.

Participant Sample ~~and Recruitment Strategy~~

12

We plan to recruit individuals who are well-versed in the theory and/or practice of RR.[4] These people may have published on the topic, developed frameworks and codes based on RR, may be involved in hands-on RCR training or community activity, or support researchers and research teams in RR-related research projects and institutions. Recommendations and empirical studies on dissensus Delphi methods vary on desirable sample size. For instance, 20-30 participants seems to be a sufficient panel based on Melander (2018; note that this is more than typical consensus Delphi panels tend to require), while Turoff (2002) suggests that between 10 and 50 panelists is sufficient for a dissensus Delphi. As our objective is to capture a wide range of perspectives, we aim to recruit – and ideally retain across the course of various rounds – a very diverse and somewhat large participant sample, in terms of disciplinary, geographical, and institutional contexts. We will aim to recruit 40 panelists.

As a guideFor our research aims, we are looking for a diverse – in terms of academic discipline – expert panel. As a guide for this disciplinary diversity, we use the 2021 Research Excellence Framework (REF) Units of Assessment (UoA). There are 34 UoA listed (Table 1), which in our estimation provides a cross-section of research disciplines sufficiently granular to provide a very diverse participant pool (for which a categorizationcategorisation with only 6 or 7 elements would not be sufficient), without having so many categories that we would have difficulty finding a participant from each category (some lists contain hundreds of small discipline categories). We have collected the names and contact details of approximately 95 individuals, with each of the 34 UoA

---

[4] Note that while many researchers interested and involved in open/transparent research and reform science topics have shown interest in being involved in our RCR research including this Delphi panel, we are selective when it comes to the individuals we will invite to participate. This is because although open and reform research topics are adjacent and relevant to RR, we are considering RCR in terms of its formal frameworks (for instance in terms of the framework officially used by the European Commission's Framework Programs) and require input from individuals with expertise in this specific RR(I) framework and its use in research and policy.

categories containing at least one potential Delphi participant, and most containing 2 or more as fallback options for when participants inevitably decline to participate or cannot be contacted. Potential participants are *not* required to be a practicing scholar within the UoA they have expertise on, although many may be; others may be an expert by dint of studying practices in that field through a different paradigm. For instance, a medical anthropologist may be recruited to be an expert in medicine rather than anthropology for the purposes of our study.

These potential panelists were identified using a combination of authorship lists in the scholarly literature as well as online researcher profiles. People were put on the candidate list based on a) their individual RRI/RCR activity, or b) their involvement in RRI/RCR on the community level. They were selected if they occupied a key role in an authorship list (i.e., being first, second or last author on an article), or if they include the keywords "responsible research", or "RRI" on their institution's website or their own personal webpage. They were also selected if they were involved in a project focusing on RRI (such as the European Commission's NewHoRRIzon, MoRRI or SUPER MoRRI projects: https://newhorrizon.eu; https://super-morri.eu/morri-2014-2018/), or, finally, if they are a part of a RCR/RRI network or working group (such as the RRING network: https://rring.eu, or the UKRI: https://www.ukri.org).

Panel	Unit of Assessment
A	1 Clinical Medicine
	2 ~~Pub.~~Public Health, Health Services and Primary Care
	3 Allied Health Professions, Dentistry, Nursing and Pharmacy
	4 Psychology, Psychiatry and Neuroscience
	~~5 Biology~~5 Biological Sciences
	6 Agriculture, Food and Veterinary Sciences
B	7 Earth ~~/Environ. Science~~ / Environmental Sciences
	8 Chemistry
	9 Physics
	10 ~~Math~~Mathematical Sciences
	11 ~~CS~~Computer Science and ~~informatics~~Informatics
	12 Engineering
C	13 Architecture, Built Environment and Planning
	14 Geography and Environmental Studies
	15 Archaeology
	16 Economics and Econometrics
	17 Business and Management Studies
	18 Law
	19 Politics and International Studies
	20 Social Work and Social Policy
	21 Sociology
	22 Anthropology and Development Studies
	23 Education
	24 Sport and Exercise Sciences, Leisure and Tourism
D	25 Area Studies
	26 Modern Languages and Linguistics
	27 English Language and Literature
	28 History
	29 Classics
	30 Philosophy
	31 Theology and Religious Studies
	32 Art and Design: History, Practice and Theory
	33 Music, Drama, Dance, Performing Arts, Film and Screen Studies
	34 Communication, Cultural and Media Studies, Library, and Information Management

Table 1. Research Excellence Framework 2021 Units of Assessment (retrieved from:

https://www.ref.ac.uk/panels/units-of-assessment/)

From these disciplines, we plan to recruit individuals who are well-versed in the theory and/or practice of RCR.[5] These people may have published on the topic, developed frameworks and codes based on RCR, may be involved in hands-on RCR training or community activity, or support researchers and research teams in RCR-related research projects and institutions. For our recruitment strategy, we have operationalised the above concepts into the following inclusion criteria: a participant must have (co)authored at least one peer reviewed article (in articles of more than two co-authors, their position in the author list must indicate leadership in the project in terms of its content; i.e., being in first or second author position, or being corresponding author) including the following keywords: "RCR", "RRI", "responsible research and innovation" "research integrity" or "responsible research" (as per our previously-published scoping review, see Field et al., 2024) AND/OR include one or more of these keywords in their personal institutional webpage AND/OR have taught RCR/RRI to researchers, AND/OR have been involved in a project focusing on RCR/RRI (such as the European Commission's NewHoRRIzon, MoRRI or SUPER MoRRI projects: https://newhorrizon.eu; https://super-morri.eu/morri-2014-2018/), AND/OR, finally, have been part of a RCR/RRI network or working group (such as the RRING network: https://rring.eu, or the UKRI: https://www.ukri.org).

While ~~this method does~~these operationalisations do not ensure that all participants on the list will be RCR experts, we ~~think that this is~~consider them a valid proxy~~.~~ for the purpose of this study. If people ~~conduct~~lead research projects on RRI or RCR, list these topics as being their areas of expertise, or train and support others in RCR/RRI activities, they will

---

[5] Note that while many researchers interested and involved in open/transparent research and reform science topics have shown interest in being involved in our RCR research including this Delphi panel, we are selective when it comes to the individuals we will invite to participate. This is because although open and reform research topics are adjacent and relevant to RR, we are considering RCR in terms of its formal frameworks (for instance in terms of the framework officially used by the European Commission's Framework Programs) and require input from individuals with expertise in this specific RR(I) framework and its use in research and policy.

likely have a sufficiently deep knowledge of the currently established tenets of RCR/RRI and some of the existing scholarly literature, and perhaps the application of RCR/RRI principles in policy to engage meaningfully with the stimulus provided as part of the Delphi process.[6]. Additionally, in the contact emails, we will be clear about the kinds of participants we are looking for for this study, and that while our recruitment methods are reasonable, we might have made errors in judgment. Should the participants we approach feel that they are not suitable for purposes of participating in our study, they may also exclude themselves on these grounds.

Where possible, we also aimed to balance the sample with a selection of different genders and regions of the world. The United Kingdom, the Republic of Ireland, Northern, Eastern, Central and Western Europe, the Nordic countries, the Middle East, countries in the Pacific such as Australia, North America, Canada and China are all represented in our candidate list. We will work to ensure that as much of that diversity as possible filters into the final sample.

We also note that despite our goal of developing output that is diverse in terms of the scientific disciplines that are represented in it, this output will represent only a limited selection of countries, regions and cultures. While the broader project within which this study is situated concerns RCR in the UK and regions of Europe, and a Euro-centric approach is appropriate to those ends, we emphasise that our findings will be produced with the input of a largely Western participant sample. We discuss the impact of this on our findings further in the limitations section in the discussion.

---

[6] We recognise that while assembling a panel of RCR experts is appropriate for the aims of this Delphi study, and for the wider aims of the project the study is part of, this approach risks leading to a somewhat homogenous set of dimensions based predominantly upon pre-existing frameworks and models. We recognise that other strategies may lead to a more substantially different or transformative framework in comparison with what already exists.

Panel Size

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. General interest in our project so far would indicate that we should expect a higher-than-average response rate. We received enthusiastic responses to our recruitment calls for general involvement in our interviews, the Delphi and any other potential empirical studies relating to our RRPractice project[7], which went out on social media and via various relevant networks in both the UK and Europe in mid-2022. Despite this, it is largely agreed (e.g., see Murphy et al., 1998, or Keeney, Hasson & McKenna, 2006) that the overall response rates of Delphi panels tend to be low, and attrition rates high (retention rates vary widely from study to study, however they decrease sharply as a function of the number of rounds in most studies).

We will therefore attempt to recruit at least 40 potential participants, which provides us with a buffer. We hope that recruiting this many panelists will lead to having a sample of no fewer than 20 participants by the close of the final round.

Such a targetRecommendations and empirical studies on Delphi methods vary on desirable sample may size. For instance, 20-30 participants seems to be a sufficient panel based on Melander (2018; note that this is more than typical consensus Delphi panels tend to require approaching many individuals.), while Turoff (2002) suggests that between 10 and 50 panellists is sufficient for a dissensus Delphi. Choosing a minimum number necessarily contains an arbitrary factor, as well as a pragmatic one. Considering our goal of disciplinary diversity in the expert panel, we elected to use a minimum on the higher end of the average

---

[7] For purposes of clarity, what we refer to as the RRPractice project is a large research project that subsumes the current, proposed study. More information on RRPractice is available here:
https://www.cwts.nl/projects/current-projects/rrpractice

that these two sources suggest. As such, we have set the minimum panel size at the start of the process to be 30 panellists. 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 a minimum amount of disciplines present 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 3 participants from one single discipline.

We will initially approach ~~68 (i.e., two~~ three persons ~~for~~from each ~~of the 34~~ UoA~~)~~ ~~and~~, aiming for well above our minimum sample size. We will start the Delphi process when either 1. at least two persons from each UoA have agreed to take part, or 2. after three weeks of recruitment have elapsed, as long as our minimum panel size and disciplinary diversity requirements are met. If this is not the case we will continue ~~to contact possible candidates~~recruiting until the minimum numbers are met.

Should enough participants drop from the study such that the total N drops below 15 or the amount of disciplines represented drops below 10, we will resume recruitment until these minimum thresholds are met. In this case we will still maintain a maximum of three participants per discipline. While this is not ideal methodologically speaking, our goal of having the input of a diverse and large enough panel is more important than having continuity across rounds. Any attrition and replacement will be thoroughly and transparently documented in the final manuscript.

Recruitment Strategy~~we receive consent to participate from 40 people. We will first exhaust the aforementioned candidate~~

The recruitment strategy for this ~~list.~~ will be largely centered on the articles included in the scoping review that preceded the Delphi study (Field et al., 2024). First and second authors of included articles will be searched whether they satisfied the inclusion criteria.

Where a specific discipline is missing potential participants, we will carry out google searches with combinations of the following keywords: "RCR", "RRI", "responsible research and innovation" "research integrity" and "responsible research" and the names of the specific UoA's, until we have a list of eligible participants per discipline. Should more recruitment be necessary, we will then actively recruit people through our own networks, ~~and then resort to snowball sampling (in which the same inclusion criteria will be used as listed above). Snowball sampling is a common method of sampling in qualitative research studies and can be especially effective for studies where the participant sample is required to have certain expertise or knowledge (Noy, 2008). In this case, we will request that each existing panelist provides us with other possible participants, and then approach~~including those ~~also, for names of people they recommend.~~identified in a previous, more general call for participants for the wider project through professional contacts of the authors, social media, and local university networks.

Potential recruits will be emailed to ask whether they are willing to participate~~.~~ as experts in our Delphi panel. We will send consent forms, participant information statements and study information (once again, via email) to those people that consent to participation in the Delphi. In addition, we will also ask for recommendations of other participants who fulfill the above-mentioned criteria, from those who declined to participate. Potential participants pointed out by declining participants will be vetted as to whether they meet the inclusion criteria before they are invited to participate.

~~Once we have reached the target sample and are content that the sample is as diverse as possible (we will be monitoring this aspect as we approach candidate panelists), we will proceed with the Delphi panel. If one month has elapsed and we have not successfully recruited more than 30 individuals, we will go ahead with the Delphi; however, we will also need to explicitly discuss the possible threat to our findings' validity that an insufficient~~

sample would pose. We consider this a remote possibility, however. As we mentioned earlier, we have received much interest in the RRPractice project, and have experienced a surprisingly high rate of agreement to our requests for interviews (though, we do recognize that agreeing to a one-hour interview is probably easier for participants than agreeing to participate in a multi-round Delphi panel).

Procedure

In order to acknowledge the panellists' effort in this research project, we will offer authorship to each participant who participated in at least three rounds (for these purposes, the initial modification phase is also counted as a round) and is willing to contribute to the drafting of the final manuscript. We will also offer to list all panellists as contributors in a statement in the final manuscript, if they wish. Because we are recruiting experts, and this is necessarily a small and select population, we will allow participants to skip rounds if need be, although this will be discouraged.

Delphi procedure

The goal of our Delphi process is to establish as broad a range of perspectives on our pre-constructed reference document, and as many suggestions for additional an inclusive list of various dimensions to this list as possible, as well as to establish the level of relevance RCR, and to estimate how important these dimensions haveof RCR are to the various represented research disciplines. The modified reference document, with the changes and additions provided by the panel, will form the basis for a contemporary multidisciplinary RCR framework.

The Modified reActive Dissensus (MAD) Delphi

The Delphi process, for any study, takes, including as broad a range of perspectives on this as possible. To do so, the Delphi process will take place over two phases, the latter of

21

which is divided into multiple, iterative rounds. At the start of the process, participants will be presented with the pre-constructed reference document. Then, they will have the opportunity to add to the list. Each following round involves participants evaluating the importance, or salience of several items that our pre-constructed reference document contains, with successive rounds involving response to feedback on the previous round.

Our MAD Delphi will proceed as follows. First, the authors develop the reference document (see the first Methods subsection of this article, *Initial Reference Document*, for details). Once the reference document has been established, the Delphi Figure 1 summarises the process goes ahead. This Delphi process, as we outlined earlier, involves both consensus and dissensus. Dissensus is operationalized as people adding new elements to the framework that they consider important yet are 'missing' from their ideal conceptualization, including the development of RCR. Consensus is operationalized as people rating how important dimensions on the initial reference document are, in by the opinion of authors (the participant, important to their research field or discipline.

Prior to Phase blue box in Figure 1 of the Delphi, we ask for a selection of participant demographics, including what field/discipline they will have in mind when they are engaging with the Delphi, their current (or primary) institute, and their job title or career stage. In Phase 1 of the Delphi, participants are presented with the reference document and asked to indicate what dimensions are missing from the reference document,.). This method is based on what's important to their practical or theoretical RCR 'needs'. We will ask them to consider their own sense of RCR (as developed through their embedment in their discipline), as well as the priorities of their institution. In cases where these are orthogonal, they are asked to explain the discrepancy and respond using their personal disciplinary sense of RCR. We reason that this will be a greater predictor of peoples' behavior (given that people often move between institutes to pursue academic jobs, but less readily between disciplines), and

will make for a more meaningful final framework. Ideally, they focus on their own disciplines an earlier published Delphi study, in which SMF was an author (Pittelkow et al., 2023). Pittelkow and colleagues successfully used a modified reactive Delphi method to establish a checklist for communicating the rationale behind conducting replication studies, and the protocol described and followed in Pittelkow and colleagues' paper provides a kind of proof of concept for the methodology we set out in this proposal.

The process will start with an initial modification phase, *Phase 1* (the green box in Figure 1), in which participants can suggest additions to the proposed list of dimensions. Here the goal is to broaden the scope of the initial list of dimensions, capturing the various disciplinary perspectives of the panel. After the research team incorporates these suggestions and updates the dimension list, the second phase, *Phase 2* (the red box in Figure 1) will involve multiple rounds in which participants rate the importance of these dimensions to RCR in their discipline. In this phase, the goal is to probe which items are more broadly appreciated by the sample (i.e., which might be universally valuable in RCR practice), versus which might be more discipline-specific. In this way, the list of RCR dimensions can be 'weighted' by importance across disciplines.[8]

---

[8] Note that the weighting information captured in Phase 2, comparing importance of dimensions across disciplines, will help the research team structure the RCR cross-disciplinary mapping we aim to develop in a further stage in the broader project.
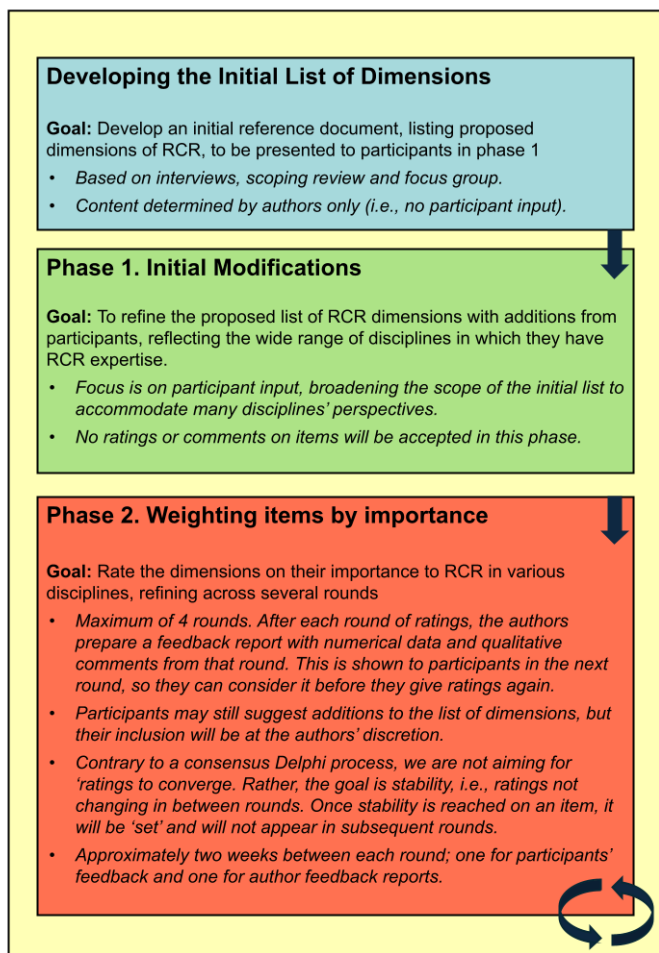
Developing the Initial List of Dimensions

Goal: Develop an initial reference document, listing proposed dimensions of RCR, to be presented to participants in phase 1
- *Based on interviews, scoping review and focus group.*
- *Content determined by authors only (i.e., no participant input).*

Phase 1. Initial Modifications

Goal: To refine the proposed list of RCR dimensions with additions from participants, reflecting the wide range of disciplines in which they have RCR expertise.
- *Focus is on participant input, broadening the scope of the initial list to accommodate many disciplines' perspectives.*
- *No ratings or comments on items will be accepted in this phase.*

Phase 2. Weighting items by importance

Goal: Rate the dimensions on their importance to RCR in various disciplines, refining across several rounds
- *Maximum of 4 rounds. After each round of ratings, the authors prepare a feedback report with numerical data and qualitative comments from that round. This is shown to participants in the next round, so they can consider it before they give ratings again.*
- *Participants may still suggest additions to the list of dimensions, but their inclusion will be at the authors' discretion.*
- *Contrary to a consensus Delphi process, we are not aiming for 'ratings to converge. Rather, the goal is stability, i.e., ratings not changing in between rounds. Once stability is reached on an item, it will be 'set' and will not appear in subsequent rounds.*
- *Approximately two weeks between each round; one for participants' feedback and one for author feedback reports.*

Figure 1. Flowchart showing the phases featuring in this modified Delphi protocol.

Each phase or round of this Delphi will be hosted asynchronously in an online environment that panellists can access through a link provided in an email. When panellists follow this link, they will arrive at a Qualtrics survey containing the Delphi questionnaire. Each questionnaire will start by asking which discipline a participant identifies with for the

24

purposes of the Delphi, which they will have confirmed with the researchers beforehand during recruitment. Prior to Phase 1 of the Delphi, we also ask for two participant demographics, namely their geographical region, and years of expertise in their discipline; this is merely to understand the overall makeup of our panel, and will be reported separately to any analysis of the main data.

In Phase 1, we will present the experts with the full list of the proposed dimensions, and ask whether they deem any dimension missing from the list. They will be encouraged to add any dimensions they think should feature in the list. Panellists can answer this question in a textbox with unlimited characters. After Phase 1, the research team will refine the list (see "Feedback reports and analysis plans" section below for details) and then dimensions (including their names and definitions) will be set for all future rounds of the Delphi process. However, if, from analysis of the qualitative data in Phase 2, it is overwhelmingly clear that two or more groups of panellists are making different interpretations of a particular dimension, the research team may have to change or split that specific dimension for future rounds. All changes to the dimension list will be noted in the results, and reported in detail in the supplemental materials.

In Phase 2, which will involve up to four rounds, we will ask for ratings of importance for each dimension. The survey will present each proposed dimension consecutively, and will ask the participants how important they consider the dimension of RCR to their specific discipline on a 7-point rating scale ranging from "extremely unimportant" to "extremely important". In addition, the participants will be encouraged to motivate their answers in a textbox with unlimited characters, though a motivation is not required to move to the following dimension. Note that, other than a brief explanation as part of the initial survey instructions, we will not attempt to define to the participants in any great detail what "important" means. Although in general it is advisable to be as precise as possible

in elicitation of survey measurements, we believe that in this case trying to prescribe particular aspects of the concept of "importance" would counterproductively narrow our measurement, when an intuitive, broad understanding of the word may more closely capture the essence of what we wish to measure, i.e., the sense that a dimension "matters" in that discipline. In round 1 of Phase 2 (i.e., the first round of ratings), only the name and definition of each dimension will be presented to participants. In all subsequent rounds, a feedback report (see section below) summarising data from the previous round will also be presented for each dimension. In Round 3 and 4, participants will only be presented with dimensions that have not reached stability in the prior rounds. We will conclude Phase 2 after a maximum of four Delphi rounds. Melander's review suggests between two and three rounds is the average for a consensus Delphi; however, because we expect a particular diversity of disciplines and perspectives in our Delphi, we will allow up to four rounds if needed.

Note that in Phase 2, participants will not be asked to add *new* dimensions. Allowing the addition of new items after Phase 1 is complete will needlessly complicate the process and cause it to take longer than might be reasonable, risking higher attrition and discontinuity. Should participants still wish to add items, despite this, it may be indicated in the response boxes provided, and will be considered by the research team. In extreme cases (where many participants suggest the addition of the same kinds of items), new additions will be considered though we believe this to be unlikely. What is and is not required of participants for each phase will be clearly communicated to them.

In all stages of the Delphi, panellists are asked to answer based on their expert understanding of the view from their discipline as a whole, not simply their own personal opinions. Ideally, they focus on a single discipline in this exercise, but some respondents (such as those that are more senior or more embedded in the RCR/RRI sphere, or work across multiple fields) might have more interdisciplinary input. We will ask them to answer

in relation to their primary field of expertise (that is, the one we recruited them for), however we will also ask that they add any additional insights from other fields they're familiar with into ~~the open text boxes. Further, we will ask them to note discrepancies or conflicting suggestions across fields if they are aware of them. These 'bonus' insights will not necessarily be used to develop the framework, but will certainly be useful for us to further furnish the broader picture of RCR across disciplines that we are attempting to capture in this study.~~ an open text box at the end of the survey.

~~The authors will pool the information derived from the first round, construct a feedback report for the participants, and revise the reference document in preparation for Phase 2. The feedback report will include, for instance, the calculated median and IQRs per item, a depiction of the distribution of the responses per item, and a report of what items will be excluded based on low importance ratings. The revised reference document will reflect the participant's suggested dimensions. In Phase 2, Round 1 the participants are presented with the *updated* reference list and the feedback on what was added from the previous round, and asked to rate how important each dimension is for their sense of RCR on a 9-point scale (where 9 corresponds to *Highly Important* and 1 to *Unimportant*). We will also provide an 'I don't understand' option, with a prompt to explain in the text box to avoid participants potentially guessing or providing a meaningless answer for items they aren't sure of how to answer. This will be a forced-response format, with participants prompted to enter an answer if they fail to provide a rating. They will also be asked to provide any (optional) extra contextual information or explanations per item in open response boxes. Note that in this phase, participants rate both pre-existing reference document dimensions, as well as those that were suggested by other panelists. As in Phase 1, the authors pool the information provided, and construct feedback and revise the reference document in preparation for a following round (if relevant), as described above. In the feedback report and the revised~~

reference document, we will also include information about what items were added and which were dropped.[9] Participants' 9-point scale responses will be pooled for each item, and interquartile range (IQR) and median will be calculated resulting in two values per item on the reference document. Subsequent rounds will focus on validating items in an increasingly refined reference list (see the following subsection *Developing the Framework* for details on this), as participants (hopefully) converge on the most important items, modifying their previous responses based on others' ratings and feedback.

We will conclude the process after a maximum of 4 Delphi rounds. Melander's review suggests between 2 and 3 rounds is the average for a *consensus* Delphi, therefore, since we are including an initial dissensus round, we will conduct a maximum of 4 rounds in Phase 2 (for a possible maximum of 5 rounds including the one round in Phase 1, where participants suggest dimensions). We contend that a 4-round process strikes the balance between feasibility for participants, and what the literature judges to be necessary for valid output. We will conclude the process earlier if no change is observed in the IQR and Median calculations for all items between two given subsequent rounds, or if the author team agrees that little enough change (i.e., so little change as to render the difference conceptually meaningless) has occurred between two rounds.

Developing The Framework

We expect that items will reach consensus to widely varying degrees. This is not only expected but desirable for our purposes. Recall that while most consensus Delphi studies pursue pure consensus, we intend on capturing diversity in perspectives (as is in line with the main goal of the overarching RRPractice project) and encourage dissensus. It is through

---

[9] Recall that we will not validate all dimension suggestions, only the ones that are considered relatively important in the rounds of Phase 2.

ranking the final item importance ratings that we will arrive at a concept of our contemporary multidisciplinary framework of RCR. The framework, if depicted visually, resembles an archery target, with a bullseye and concentric circles ringing it. We will arrive at a preliminary concept of the framework after this study, depending on how fruitful the Delphi panel—and how useful the output—is. It will be constructed as follows.

The items with the greatest consensus and relevance (for instance, items with an $IQR \leq 2$ and $Median \geq 7$) will form the *core* or 'bullseye' of the framework, representing more or less *universal* dimensions of RCR. Such items, which by their high convergence and relevance we will take to be very important to most, if not all researchers, should be central to the framework. Items with moderate consensus and relevance (IQR 3-5 and Median 4-6) will be the basis of the second circle of the target. These items are *peripheral*, to a degree, in that while they are generally considered quite important to many participants, they are not deemed central to an understanding of RCR by *all* participants. Finally, items with low consensus and relevance (i.e., $IQR \geq 5$ and $Median \leq 3$) will either form the final ring, or ultimately be dropped from the framework. Most of these items will be considered *niche*, that is, important to some participants but only within certain disciplinary categories (or which may signal dissensus within a category), yet nevertheless important within those niche research areas. Items where only one person considers them important will fail validation and be excluded from the framework. Items that many people consider somewhat important (though less than moderately) will be considered for inclusion in the niche category of the framework. In this way, we will use the final IQR and median calculations for each item as a sort of centrifuge, such that the 'heavier' items

will be separated from the 'lighter' items by the IQR and median, arranged thus in the framework.

We re-emphasize at this juncture the exploratory nature of this Delphi study. We will need to see the distributions of each item's data before determining for certain whether these quantitative categories (i.e., the median and IQR threshold ranges defined earlier) are valid and applicable. If they are not, we will redefine our categories and transparently report the change and its motivation.

For each round, as well as the initial modification phase, panellists will have one week to fill out the questionnaire. All rounds will have one week for the research team to complete that round's interim analysis, after which the following round will commence.

These weeks will be staggered and run from Wednesday to Tuesday of the consecutive week, to accommodate for panellists' potential leave weeks.

Feedback reports and analysis plans

After Phase 1, during which participants suggest additions or changes to the dimensions in the initial proposed list, the research team will screen these suggestions based on whether they are 1) sensible, i.e., relevant to RCR, coherent, and factually correct, and 2) different enough from existing dimensions to add value to the list. If suggested additions are merely narrower versions of existing dimensions, we may simply add the suggestion as an example to the definition of the applicable dimension. At least two researchers will screen all the participant-suggested additions and changes, recording their recommendations individually, then these will be compared across the team and any discrepancies will be resolved by group discussion. The team will then amend the proposed dimension list with all changes or additions agreed by the team to be both sensible and adding value. All suggested additions and changes will be collated, along with the decisions the research team makes on them, and recorded in supplemental materials.

After each round of the rating phase, we will analyse the panellists' answers and draft a feedback report. The feedback reports will consist of the descriptive statistics of the aggregated ratings (median and interquartile range, IQR) and an analysis of the qualitative feedback per dimension (as was done in Pittelkow et al., 2023). All qualitative feedback – the panellists' input – will be organised per item and placed in an appendix to the feedback report, accessible to participants by clicking a web link. This feedback will be anonymised if necessary, but otherwise left untouched.

Qualitative feedback will be bundled per dimension and analyzed by the research team using a simple form of thematic analysis. Though we encourage the participants to be as

31

detailed in their responses as time allows them to, we do not expect the qualitative data to be highly complex, as the participants will be answering a specific question. This is why we will use a "small q" or "coding reliability" version of thematic analysis, focusing on structured codebooks, independent coders, and consensus between coders (Braun & Clarke, 2023; Finlay, 2021; Kidder & Fine, 1987). For each round of the Delphi study, two researchers on the team will independently open code the data, after which they will meet to establish a final codebook with which the data will be coded. The result of this analysis will be summarised and given to the experts in the feedback report for the subsequent round, along with an appendix with the 'raw' qualitative data. The key to our analysis method is transparency, as we will provide access to the 'raw' qualitative data, as well as the coding books with which they were analysed. In addition, since our analyses will be distributed to all participants/experts before every new round, along with their own data, interested participants/experts are welcome to, and invited to, check the analysis as the Delphi process continues.

In subsequent rounds, the participants will be presented with the feedback report, and will be asked again to score the list of dimensions. This will continue until we reach stability on all dimensions, though for the sake of feasibility we will conduct no more than four rounds of ratings.

Since we selected our participants in this study on the basis of their individual expertise – which surpasses our knowledge of their disciplines – we will not perform any stringent quality checks on the content of the data, quantitative or qualitative. Data quality should be aided by the fact that participants are encouraged to write down free-text justifications to every question, which should prompt them to answer questions thoughtfully. If we notice suspicious patterns in the data, however (such as all items answered with the same choice) we will contact participants individually to check that they meant these.

Stability

An important aspect of the Delphi process is the concept of stability, or when we consider the answers to be similar enough between two or more subsequent rounds that we can consider the answer "settled" or "definite". Since panel responses can vary greatly between rounds - as per the explicit aim of the Delphi process - it is important to assess whether the panel's response on any given dimension is still developing or whether it can be considered settled. In fact, different authors have argued that assessing the level of consensus in a Delphi study is meaningless without having assessed stability of responses, since the response may not be an accurate reflection of the conclusive judgment of the panel (Dajani et al., 1979, Scheibe et al., 2002). To reduce participant burden, where stability is reached, the item will be considered 'set', and not feature in subsequent rounds.

We use a simple metric for stability: for each dimension, we will take the absolute value of the change in ratings for each participant. If the mean of these absolute-value change scores is less than the equivalent of 1 point on our rating scale (i.e., 16.66% of the total breadth of the rating scale), we will consider the dimension stable. This tracks closely with the recommended cutoff for stability of 15% difference recommended by Scheibe et al. (2002), which is based on an empirical estimation of the random change between rounds. However, we will also temper this quantitative stability judgement with qualitative analysis: if the qualitative data contain novel arguments for the importance or unimportance of a dimension that we have reason to believe may sway the panel substantially in the following round, we will not consider the dimension to be stable.

Reflexivity/positionality

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, and as the framework is built. Our team comes from a background in science studies, metascience and research integrity, and we have previously published on the topics of epistemic diversity, responsibility and quality (e.g., Field & Derksen, 2021; Muller & de Rijcke, 2017; Penders, de Rijcke & Holbrook, 2020; Penders & Goven, 2010; Valkenburg, Dix, Tijdink and de Rijcke, 2021; Van Drimmelen et al. 2023). We have also previously published the scoping review on which this Delphi study directly builds. As a result, we are aware of the literature on RCR and adjacent topics. Unavoidably, our decisions will be rooted in this knowledge, from the initial choices we have made to develop this study, to the choices we will collectively make as we co-construct an RCR framework with the input of our experts. 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.

Data Sharing

We believe in the importance of data sharing, both from the perspective of accountability, as well as the potential re-use of our data. As such, we will share all data and analysis, guided by the TOP Guidelines. We will do this by making the feedback reports openly available in a suitable repository. These feedback reports will include the pseudonymised 'raw' data, both quantitative and qualitative, along with our analyses of this data, subject to any redactions by study participants for their privacy. Final versions of all study materials will be uploaded to this page before the study starts.

Expected results

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. 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). However, our goal is different, as 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 (and expect we might realistically find) 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 (e.g., histograms) 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.

In addition to the main table containing the median importance ratings across the entire panel, we will also provide a complete view of the Delphi results broken down by

discipline, by reporting a separate table with the median rated importance of each discipline per dimension. Since our sample will include no more than three panellists per discipline, and likely less than that, we will be cautious in making any inferences on the basis of these analyses, though we do consider them worthwhile sharing. We will find it useful (for purposes of summarising and interpreting our findings) to pre-specify some simple labels to categorise these measurements of importance. Our categories of importance are easy to delineate: on our 1-7 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.

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. All of our results will be discussed in a discussion section.

References

Barry, C. A., Britten, N., Barber, N., Bradley, C., & Stevenson, F. (1999). Using Reflexivity to Optimize Teamwork in Qualitative Research. *Qualitative Health Research, 9*(1), 26-44. https://doi.org/10.1177/104973299129121677

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends

36

methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401-409. https://doi.org/https://doi.org/10.1016/j.jclinepi.2013.12.002

Dajani, J. S., Sincoff, M. Z., & Talley, W. K. (1979). Stability and agreement criteria for the termination of Delphi studies. Technological Forecasting and Social Change, 13(1), 83-90. https://doi.org/https://doi.org/10.1016/0040-1625(79)90007-6

Field, S. M., & Derksen, M. (2020). Experimenter as automaton; experimenter as human: exploring the position of the researcher in scientific research. European Journal for Philosophy of Science, 11(1), 11. https://doi.org/10.1007/s13194-020-00324-7

Field, S. M., Thompson, J., de Rijcke, S., Penders, B., & Munafo, M. (~~2023~~2024). Exploring the Dimensions of Responsible Research Systems and Cultures: A Scoping Review. Preprint: https://osf.io/6z4mx/

Funtowicz, S., & Ravetz, J. (2018). Post-normal science. In *Companion to environmental studies* (pp. 443-447). Routledge.

Grant, S., Booth, M., & Khodyakov, D. (2018). Lack of preregistered analysis plans allows unacceptable data mining for and selective reporting of consensus in Delphi studies. Journal of Clinical Epidemiology, 99, 96-105. https://doi.org/https://doi.org/10.1016/j.jclinepi.2018.03.007

Keeney, S., Hasson, F., & McKenna, H. (2006). Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *Journal of advanced nursing*, *53*(2), 205-212.

Linstone, H. A., & Turoff, M. (2002). The Delphi Method: Techniques and Applications. Addison-Wesley.

Linstone, H. A., & Turoff, M. (2011). Delphi: A brief look backward and forward. Technological Forecasting and Social Change, 78(9), 1712-1719. https://doi.org/https://doi.org/10.1016/j.techfore.2010.09.011

Melander, L. (2018). Scenario development in transport studies: Methodological considerations and reflections on Delphi studies. *Futures*, *96*, 68-78.

37

Müller, R., & de Rijcke, S. (2017). Thinking with Indicators. Exploring the Epistemic Impacts of Academic Performance Indicators in the Life Sciences. Research Evaluation, 26(4), 361-361. https://doi.org/10.1093/reseval/rvx033

Murphy, M. K., Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F., Askham, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment (Winchester, England)*, *2*(3), i-88.

Noy, C. (2008). Sampling knowledge: The hermeneuticsNational Health and Medical Research Council. (2018). Australian Code for Responsible Conduct of snowball samplingResearch. http://www.nhmrc.gov.au/guidelines/publications/r41

Olmos-Vega, F. M., Stalmeijer, R. E., Varpio, L., & Kahlke, R. (2023). A practical guide to reflexivity in qualitative research. *International*: AMEE Guide No. 149. *Medical Teacher, 45*(3), 241-251. https://doi.org/10.1080/0142159X.2022.2057287

Penders, B., de Rijcke, S., & Holbrook, J. B. (2020). Science's moral economy of repair: Replication and the circulation of reference. Accountability in Research, 27(2), 107-113. https://doi.org/10.1080/08989621.2020.1720659

Penders, B., & Goven, J. (2010). Nutrigenomics and the stewardship of scientific promises. Biotechnology, Journal *of social research methodology, 11*(4), 327-344., 5(9), 909-912. https://doi.org/https://doi.org/10.1002/biot.201000069

Pittelkow, M. M., Field, S. M., Isager, P. M., van't Veer, A. E., Anderson, T., Cole, S. N., ... & Van Ravenzwaaij, D. (2023). The process of replication target selection in psychology: what to consider?. *Royal Society Open Science, 10*(2), 210586.

Salkind, N. (2007). Encyclopedia of Measurement and Statistics. In. SAGE Publications, Inc. https://doi.org/10.4135/9781412952644

Scheibe, M., Skutsch, M., & Schofer, J. (2002). Experiments in Delphi Methodology. In H. A. Linstone & M. Turoff (Eds.), The Delphi Method: Techniques and Applications (pp. 257-281). Addison-Wesley Publishing Company.

Tallacchini, M. (2009). Governing by values. EU ethics: soft tool, hard effects. *Minerva*, *47*(3), 281.

Turoff, M. (2002). The Policy Delphi. In H.A. Linstone and M. Turoff (eds.*). The Delphi Method: Techniques and Applications*, Reading MA: Addison-Wesley, pp. 80-96.

Valkenburg, G., Dix, G., Tijdink, J., & de Rijcke, S. (2021). Expanding Research Integrity: A Cultural-Practice Perspective. Science and Engineering Ethics, 27(1), 10. https://doi.org/10.1007/s11948-021-00291-z

Van Drimmelen, T., Slagboom, M. N., Reis, R., Bouter, L. M., & Van der Steen, J. T. (2023). Decisions, Decisions, Decisions: an ethnographic study of researcher discretion in practice. MetaArXiv. https://doi.org/https://10.31222/osf.io/7dh3t

Von Schomberg, R. (2011). Towards responsible research and innovation in the information and communication technologies and security technologies fields. *Available at SSRN 2436399*.

Von Schomberg, R. (2012). Prospects for technology assessment in a framework of responsible research and innovation. *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden*, 39-61.

Von Schomberg, R. (2013). A vision of responsible research and innovation. *Responsible innovation: Managing the responsible emergence of science and innovation in society,* 51-74.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. Journal of Clinical Epidemiology, 67(4), 401-409. https://doi.org/https://doi.org/10.1016/j.jclinepi.2013.12.002

Grant, S., Booth, M., & Khodyakov, D. (2018). Lack of preregistered analysis plans allows unacceptable data mining for and selective reporting of consensus in Delphi studies. Journal of Clinical Epidemiology, 99, 96-105. https://doi.org/https://doi.org/10.1016/j.jclinepi.2018.03.007

39

Salkind, N. (2007). Encyclopedia of Measurement and Statistics. In. SAGE Publications, Inc.

    https://doi.org/10.4135/9781412952644

Williams, P. L., & Webb, C. (1994). The Delphi technique: a methodological discussion.

    Journal of Advanced Nursing, 19(1), 180-186.

    https://doi.org/https://doi.org/10.1111/j.1365-2648.1994.tb01066.x

Zwart, H., Landeweerd, L., & Van Rooij, A. (2014). Adapt or perish? Assessing the recent

    shift in the European research funding arena from 'ELSA' to 'RRI'. *Life sciences,*

    society and policy, *10*, 1-19.