

# BMJ Open Developing evaluation metrics to measure the US national institute of health plans to enhance diverse perspectives: a protocol for a consensus-based study

Yulia A Levites Strekalova <sup>1</sup>, Sanjay Jain,<sup>2</sup> Pinaki Sarder<sup>3</sup>

**To cite:** Levites Strekalova YA, Jain S, Sarder P. Developing evaluation metrics to measure the US national institute of health plans to enhance diverse perspectives: a protocol for a consensus-based study. *BMJ Open* 2025;15:e087739. doi:10.1136/bmjopen-2024-087739

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2024-087739>).

Received 17 April 2024

Accepted 07 February 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

<sup>1</sup>Health Services Research, Management & Policy, University of Florida College of Public Health & Health Professions, Gainesville, Florida, USA

<sup>2</sup>Department of Medicine, Washington University School of Medicine in Saint Louis, St Louis, Missouri, USA

<sup>3</sup>Department of Medicine, University of Florida College of Medicine, Gainesville, Florida, USA

## Correspondence to

Dr Yulia A Levites Strekalova; [yulias@ufl.edu](mailto:yulias@ufl.edu)

## ABSTRACT

**Introduction** The diversity of participating investigators representing diverse disciplines, career stages, stakeholder groups, regions and types of institutions is essential for the success of large-scale research programmes. In 2021, the National Institutes of Health introduced a requirement for some of its large grants to include a separate section that describes the project's plan for enhancing diverse perspectives (PEDPs). Our project aims to develop consensus-based PEDP evaluation metrics and instruments that can be systematically and sustainably collected across the projects.

**Methods and analysis** Evaluation work is organised into three objectives. First, shared knowledge about PEDP infrastructures, activities and outcomes will be elicited through the review of the PEDP texts of funded projects, with 15 as the target sample size. Data will be analysed using a cultural domain analysis approach and assessed for recurrence and salience of PEDP metrics. Second, consensus-based evaluation metrics will be developed using a three-round Delphi method. The descriptive statistics (mean, SD and IQR) and cultural consensus analyses will be applied to the first and last rounds of the Delphi survey. Third, metrics will be piloted for implementation and validation within one of the Human BioMolecular Atlas Programme sites. Work will be completed by Fall 2025.

**Ethics and dissemination** The long-term goal of the effort reported in this paper is to develop PEDP common metrics that are generalisable and feasible across diverse projects. This rigorous, focused evaluation development effort aims to inform scientific practices and policies around implementing the plans to enhance diverse perspectives.

## INTRODUCTION

Diversity of perspectives advances the translation, worldliness and impact of research.<sup>1–7</sup> Many large-scale and local programmes aim to improve the representation of groups and organisations traditionally underrepresented in research. Examples of the institutional-level programmes that promote institutional

## STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ It uses rigorous social scientific methods to develop consensus-based plan for enhancing diverse perspectives metrics. It grounds evaluation design in stakeholder engagement and systematic feedback.
- ⇒ It employs content validity assessment to ensure clarity of data capture instruments.
- ⇒ It conducts pilot implementation to assess the feasibility of data collection.
- ⇒ It is limited to pilot implementation within one project.
- ⇒ The project captures the shared knowledge of funded National Institutes of Health investigators but does not include perspectives of the general public.

and regional inclusion in research are the US National Institutes of Health (NIH) programmes like Research Centers in Minority Institutions and Institutional Development Awards.<sup>8</sup> At the stakeholder level, the Patient-Centred Outcomes Research Institute's efforts to support engagement in research of diverse stakeholder groups like patients, clinicians, hospital systems, industry and training institutions. These programmes also call for the advancement of scholarship in the science of diversity and inclusion using rigorous and validated social scientific methods to understand the drivers and outcomes of diversity efforts. Yet, there is a lack of consensus-based metrics and outcome indicators that can promote the systematic capture and evaluation of the efforts to promote diversity of perspectives in biomedical research.

## Plans for enhancing diverse perspectives

In 2021, the NIH Brain Research through Advancing Innovative Neurotechnologies Initiative introduced a requirement for large grants to include a separate section

that describes the project's plan for enhancing diverse perspectives (PEDPs). In the following years, multiple other NIH programmes adopted a similar requirement that remains active as of the development of the present evaluation plan. This initiative is supported by the growing evidence that diverse teams with varied perspectives outperform homogeneous ones,<sup>6</sup> fostering innovation and enhancing the impact of research. To guide the applicants, the NIH clarified that PEDP sections are expected to include examples of various strategies and the ways through which federally funded projects can ensure that diversity in its many aspects is represented in scientific projects.<sup>9</sup> Specifically, the NIH-guided investigators to consider the diversity of (1) participating investigators, (2) engagement of different types of institutions and organisations, (3) partnerships that enhance geographic and regional collaborations, (4) support of early-career advancement through project infrastructure, (5) training and mentoring of trainees, fellows and junior faculty from diverse backgrounds, (6) research collaborations across disciplines and fields of expertise and (7) inclusion of community-based partners to align research activities and community values. Overall, the NIH's requirement to enhance diverse perspectives underscores the importance of considering the contributions of different perspectives, geographies and demographics to enhance research outcomes.

The considerations for diversity of perspectives are particularly relevant to current NIH large-scale research programmes that aim to generate data tools and train the workforce skilled in artificial intelligence (AI) and machine learning approaches. One such programme is the Human BioMolecular Atlas Programme (HuBMAP).<sup>10</sup> The HuBMAP initiative was established to unravel the intricate relationship between tissue organisation and cellular behaviour at the molecular level. By understanding how tissues are structured, the programme aims to gain insights into organ function variations across the lifespan and the health-disease spectrum. Despite advancements in imaging and omics technologies, the current knowledge of tissue organisation at molecular and cellular resolution remains limited to a few microscopic structures. The programme seeks to bridge this gap by integrating imaging and omics analyses to create comprehensive profiles of biomolecular distribution and tissue morphology. These profiles are then mapped in three dimensions, allowing for modelling and exploration. The goal of HuBMAP is to enhance the understanding of inter-individual variability, tissue engineering and disease emergence at the biomolecular level. To achieve this, the HuBMAP focuses on critical scientific priorities, including sourcing high-quality tissue, data validation, annotation and community engagement.

The HuBMAP consortium includes over 400 researchers and 60 institutions with a wide range of expertise in single-cell genomics, proteomics, metabolomics, imaging, bioinformatics, statistics, ethics, informatics, image analysis, machine learning & AI, clinical medicine

and data visualisation. The programme is divided into five groups that focus on (1) infrastructure and visualisation, (2) technology implementation, (3) tissue mapping, (4) transformative technology development and (5) demonstration projects. The ability of this project to produce innovative technologies, foster open knowledge exchange and promote a robust learning requirement requires that diverse perspectives of participating investigators, trainees and stakeholders are recognised and enhanced. Similarly to other large NIH programmes, HuBMAP uses PEDPs as programme instruments. Therefore, it is critical to develop a robust consensus-based approach to evaluate the PEDP components, processes and expected outcomes.

## Objectives

The goal of our project is to develop consensus-based PEDP evaluation metrics and instruments that can be systematically and sustainably collected across the HuBMAP projects. As multiple NIH training and research programmes require PEDP efforts, it is essential to systematically study the PEDP components, learn from the early PEDP implementations, conceptualise PEDP programme theory and operationalise evaluation metrics. The underlying hypothesis for this project is that collaborative evaluation planning grounded in consensus development serves as an intervention for problem definition, action refinement, shared mental model development and programme capacity building. Three specific objectives guide the evaluation design work:

Objective 1. Elicit exemplars and outcomes of enhancing diverse perspectives within the HuBMAP consortium.

Objective 2. Develop a PEDP evaluation protocol for the HuBMAP consortium.

Objective 3. Pilot the implementation of the PEDP evaluation protocol with one of the HuBMAP projects, the Computational Image Analysis Platform (CIMAP), led by AI researchers from the University of Florida.

## METHODS AND ANALYSIS

The protocol for the development of the PEDP evaluation metrics is guided by the standards for core outcome set development. See appendix 1 for the COS-STAD checklist.<sup>11</sup>

### Objective 1. Elicit exemplars and outcomes of enhancing diverse perspectives within the HuBMAP consortium

#### Conceptual framework

PEDP requirement introduced a new shared practice for NIH-funded projects, and PEDP implementations create shared knowledge among investigators and peer reviewers about what constitutes possible infrastructure, activities and outcomes that enhance the diversity of perspectives in biomedical research. Shared practices create knowledge that can be systematically elicited and measured for the saliency of ideas and PEDP components. Conceptually and methodologically, objective 1 is grounded in the

cognitive theory of culture and cultural consensus analysis (CCA).

Cultural consensus theory was developed in the 1980s as a conceptual and methodological approach to examine and measure the consistency of cultural knowledge within groups with standard social or professional practices.<sup>12 13</sup> Cultural knowledge is assumed to be sharable and systematically distributed. Culture, therefore, can be defined as the set of acquired and shared beliefs and actions that are normative within a group.<sup>13</sup> Subsequently, cultural consensus reflects how individuals align their beliefs and actions with the cultural prototypes.<sup>14</sup> Methodologically, cultural consensus methods allow for eliciting and identifying the culturally salient knowledge and practices and assessing individuals' agreement with cultural knowledge.<sup>12</sup> Furthermore, consensus methods allow for identifying subgroups that have beliefs overlapping in some aspects and diverging in others.<sup>15 16</sup> CCA makes several assumptions. First, it assumes a shared cultural knowledge or belief among the informants. This shared knowledge is considered the 'cultural truth' and is the basis for determining consensus correct answers. Second, CCA assumes that there is heterogeneity in informant competence. This means that different informants may have various levels of knowledge or expertise, and various aspects of shared knowledge may be more or less salient. Third, CCA assumes that the responses of the informants can be elicited and modelled using item response models, such as True/False, ordered category or continuous response models. These models allow for the estimation of informants' competencies and biases, as well as the consensus correct answers and difficulty of the questions.<sup>13 17</sup>

### Data sources

The process of eliciting and drawing out shared knowledge among individuals or groups can be effectively accomplished through the identification and systematic analysis of various cultural artefacts, including policy documents, grant applications, program descriptions and other written texts.<sup>18</sup> Therefore, texts of PEDP sections of funded HuBMAP and other NIH projects will be collected and analysed. To acquire the data, primary investigators of HuBMAP projects that had the PEDP requirement will be contacted and asked for a copy of their PEDP section. Other NIH calls for proposals with PEDP submission requirements will be identified, and a representative sample of non-HuBMAP primary investigators will be contacted and asked to share their PEDP sections for analysis. Collected PEDP texts will be analysed to elicit the metrics of short- and long-term success and conduct a preliminary assessment of the saliency of the proposed PEDP activities and expected outcomes.<sup>18</sup> Methodological literature suggests that a sample size of 12–15 texts will be sufficient to reach the saturation of PEDP activities and outcomes, given the high information density and relative homogeneity of the phenomenon under investigation.<sup>19 20</sup> Saturation refers to the point at which data collection can be concluded because additional data

do not present new information. For the study under this protocol, the saturation will be assessed continuously as the sample PEDPs are collected and analysed. A recent policy engagement study using this methodology showed that saturation can be reached with fewer than 10 sources,<sup>21</sup> which further supports the target sample size.

Data will be triangulated with the NIH PEDP guidelines<sup>22</sup> and analysed to identify recurrent and divergent themes.<sup>23</sup> Final results will be presented as statements that include causal relationships that link PEDP components and activities to expected outcomes.<sup>24</sup> While it is expected that the application of the PEDP components will evolve as the scientific projects are implemented and new opportunities to enhance the diverse perspectives present themselves, the analysis proposed for this objective will aim to identify shared metrics that can guide the development of evaluation plans for HuBMAP and other large-scale scientific projects.

### Data analysis

Sections of the collected PEDPs will be analysed using content analytic<sup>25</sup> and cultural domain analysis<sup>18</sup> methods. First, the texts of PEDP sections will be read to identify descriptions of the infrastructures, activities and outcomes associated with the enhancement of diverse perspectives in biomedical research. Next, specific examples of PEDP infrastructures, activities and outcomes for each analysed text will be extracted verbatim into a list. PEDP component lists will be coded descriptively by two coders to identify recurring and unique components. Descriptive coding will focus on the repetition (use of the exact words) and recurrence (use of different words that communicate the same idea) of components.<sup>23</sup> The lists will be reviewed repeatedly for refinement and consistency. Resultant structured PEDP component lists will be organised in the order they appear in the text and analysed for the saliency of the infrastructures, activities and outcomes. The Free List Analysis under R Environment using Shiny (FLARES) software will be used to analyse the statements and assess the saliency of the individual PEDP components across the analysed texts.<sup>26</sup> FLARES will be used to conduct the frequency of mention, or saliency, saturation (number of PEDP texts after no new components emerge) and multidimensional scaling analyses to identify co-occurrence of the clusters of infrastructure, activity and outcome components. Finally, PEDP infrastructures, activities and outcomes will be presented as a causal diagram and common programme theory of HuBMAP PEDP efforts.<sup>27</sup>

## Objective 2. Develop a PEDP evaluation protocol for the HuBMAP consortium

### Methodology

The development of an agreed-upon set of elements requires that diverse perspectives on the importance and feasibility of the data elements are systematically elicited and effectively integrated across multiple primary investigators and project staff. Consensus development

entails systematically eliciting and capturing expert viewpoints, methodically incorporating various perspectives, discerning consensus through voting and addressing disagreements to guide final determinations. Known for their rigorous nature, consensus methods have become recognised and increasingly applied to produce dependable evidence about programme and policy priorities, expert agreement and areas of dissenting views.<sup>28</sup> The most well-known and established approach for reaching consensus for the development of shared guidelines, recommendations and common data elements is the Delphi technique. Initially devised by the RAND Corporation in the 1950s to predict the repercussions of nuclear warfare, the Delphi method has been since used across various academic domains, including health, science, technology, business, communication, policy analysis and education.<sup>28 29</sup> The Delphi method can be applied to facilitate the systematic emergence of a unified viewpoint<sup>30</sup> and to articulate policy alternatives and options.<sup>31</sup> The method enables the restructuring of a group communication process to effectively enable a group of individuals to tackle a complex issue.<sup>32</sup> Operationally, the application of Delphi involves an iterative procedure of completing a sequence of surveys over multiple iterations.<sup>28 32</sup> The explicit emphasis on soliciting input and providing a framework for the exchange of ideas renders the Delphi method particularly suitable for fostering meaningful engagement of stakeholders in interdisciplinary, large-team research. The activities under this objective will result in an evaluation plan that will include a description of evaluation goals, standard procedures, data analysis plans and a glossary.

#### Data sources

A three-round Delphi process will be applied to establish consensus around PEDP metrics for the HuBMAP consortium and articulate guidelines for data capture. Additional rounds of Delphi survey may be necessary should the consensus not be reached. A sample of HuBMAP consortium members will be invited to participate in the Delphi process. In line with the methodological guidelines and given the relative homogeneity of the sample, we will aim to recruit 10–15 participants<sup>33</sup> and aim for maximum variability of backgrounds (eg, contact Primary Investigators, PEDP leads, project managers and scientists). First, an online survey with PEDP activities, outcomes and associated evaluation metrics identified in objective 1 will be distributed to a sample of HuBMAP investigators and project managers to assess which metrics should be included, excluded or added.<sup>34</sup> Voting for the inclusion and exclusion of metrics will be implemented through a 6-point Likert-type scale, and suggestions for additional metrics will be captured through open-text comments. Participants will also be encouraged to comment on the formulations of the evaluation metrics and propose additional activities, outcomes and metrics relevant to the overall HuBMAP PEDP activities. The commenting will be open for 2 weeks and constitute the second round of

data collection. Quantitative results of the survey will be summarised to include vote statistics for each included item and reported back to HuBMAP investigators. Similarly, comments will be summarised and reported back to the HuBMAP investigators. Open-text comments will be analysed for recurrence, and conflicting opinions will be presented as separate options for voting in the next round. The results of the first two rounds of data collection will be used to formulate options and guidelines for data collection. The candidate guidelines will be entered into an online survey and distributed to the HuBMAP investigators and project managers to quantify the consensus and formally capture the cultural model for enhancing diverse perspectives in biomedical research.<sup>12 13 35</sup>

#### Data analysis

The descriptive statistics (mean, SD and IQR) and cultural consensus analyses will be applied to the first and last rounds of the Delphi survey. Analytical memos and process notes will be created throughout all Delphi rounds. Comments and analytical notes will be analysed thematically to identify recurrent and divergent themes.<sup>23</sup>

#### **Objective 3. Pilot the implementation of the PEDP evaluation protocol with one of the HuBMAP projects, the CIMAP, led by a team of AI researchers from the University of Florida**

##### Methodology

The goal of the pilot study implemented as objective 3 of the present evaluation design effort is to generate data for PEDP processes and outcomes, assess the content validity of data capture instruments and estimate the burden and feasibility of consistent capture of the PEDP common data elements. This pilot study also aims to provide best practices and evidence-based evaluation instruments that other programmes can consistently and systematically apply. Pilot studies in biomedical programme development research support the systematic programme design and implementation planning. Pilot studies have several methodological considerations for design, implementation and data analysis. When determining the sample size for a pilot study, it is essential to consider the ‘limit of interest’ that represents an effect of scientific or medical interest.<sup>36 37</sup> Process evaluation criteria, sample size estimation and qualitative methods considered in pilot studies can improve the design and hypotheses of full-scale studies.<sup>38</sup> In observational studies, pilot designs are used to gain information about outcome variations and improve study design, informing the instrumental variable designs and propensity score matching.<sup>39–41</sup> Additionally, pilot designs can help reduce within-set heterogeneity and improve sensitivity analyses.<sup>42</sup> It is also important to consider statistical analysis techniques and reporting transparency in pilot studies. Applied to the PEDP evaluation design, the pilot study will inform the feasibility and respondent burden for collecting process and outcome evaluation metrics and programme-level hypotheses around the contributions of PEDP to the infrastructures, activities and outcomes of biomedical research.



### Data sources

The CIMAP project will host the PEDP evaluation pilot. The feasibility of the PEDP evaluation plan will be assessed using observational, retrospective and prospective data collection.<sup>43</sup> Implementation notes will be used to refine and amend the evaluation plan as needed. For objective 3, the initial data capture forms will be developed and assessed using conduct content validity testing. Content validity is a systematic process that asks subject matter experts to assess the content and clarity of data capture instruments.<sup>44</sup> Data capture instruments will be distributed to the CIMAP investigators and project managers, who will vote on the clarity of the formulations and guidelines used within the instruments. Planned participants for the pilot implementation include three primary investigators, two faculty research scientists, two postdocs, two lab managers, five PhD students, three Master's students, four undergraduate research assistants and two research interns. Once data capture instruments are implemented for data collection, observation notes will be taken by the evaluation design team to assess the workflow and data capture burden.

### Data and analysis

For content validity, participants will be instructed to use a 4-point scale and assess each data capture item for its representativeness of the PEDP evaluation context and clarity of formulations. Participants will also be encouraged to comment on the rationale for their ratings. The content validity index will be calculated by dichotomising representativeness ratings (0=rating of one or two, and 1=rating of three or four). The sum of dichotomised ratings will then be divided by the total number of experts. The resultant number will represent the proportion of experts who deemed the item content valid. The content validity index for the data collection instrument will be estimated by calculating the average content validity index across all items. Following the existing methodological guidelines, a content validity index of 0.80 will be used to assess the readiness of the data collection forms for implementation.<sup>45 46</sup> Following content validity, an online survey will be used to pilot data collection. Process

observation notes will be analysed qualitatively to identify recurring issues. Qualitative analysis will be guided by the Normalisation Process Theory<sup>47 48</sup> to generate conceptually grounded yet practical recommendations for the broad application of the developed PEDP evaluation protocol. Emphasis and attention will be paid to the process of data collection, data burden, the existence of customarily project-created data and the need to collect additional metrics through surveys and interviews. The latter two data collection modes will require additional staff. Therefore, careful consideration will be given to the metrics obtained through regular project operations versus those that require additional data collection.

Figure 1 provides a summary of the activities under objectives 1–3.

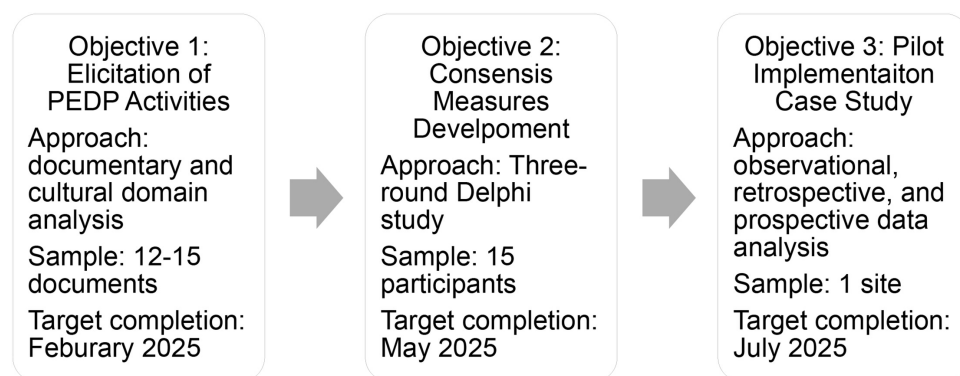
### Patient and public involvement

None. This project is limited to capturing and analysing emerging practices and knowledge among biomedical investigators. While the project does not involve human subjects or clinical trial design, we recognise that the involvement of citizen scientists would have enhanced the outcomes of this effort.

### ETHICS AND DISSEMINATION

The protocol for the objectives described above received a review by the University of Florida Institutional Review Board and has been approved as exempt under protocols ET00023350 (objective 1), ET00023593 (objective 2) and ET00023596 (objective 3). While the project does not involve patients, we recognise that the involvement of citizen scientists would have enhanced the outcomes of this effort.

The long-term goal of the effort reported in this paper is to develop PEDP common metrics that are generalisable and feasible across diverse projects. Therefore, the dissemination of the protocol and work-in-progress pilot data are integral to the completion of this work. Regular HuBMAP meetings and project demonstration days will be used to get ongoing stakeholder feedback and input. Once the initial metrics have been developed, the final set



**Figure 1** Summary of proposed activities. This figure illustrates the sequential objectives and methodologies of the PEDP study. The study comprises three key phases. The arrows between the objectives indicate the progression of the study from elicitation to consensus development and final pilot implementation. PEDP, plan for enhancing diverse perspectives.

of metrics will be made publicly available through publication and presentation at academic meetings. The evaluation design efforts and the protocol presented above can be generalised to other projects that implement their PEDPs. In sum, this focused evaluation development effort aims to inform scientific practices and policies that guide the implementation of the plans to enhance diverse perspectives.

Work is scheduled to be completed by Fall 2025.

**Contributors** YAS contributed to the writing—original draft, writing—review and editing, conceptualisation, funding acquisition, methodology, project administration and resources. PS contributed to the writing—review and editing, funding acquisition, project administration and resources. SJ contributed to the writing—review and editing and funding acquisition. YAS is responsible for the overall content as guarantor.

**Funding** This work was supported by the National Institutes of Health Common Fund award number 30T20D033753.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iD

Yulia A Levites Strekalova <http://orcid.org/0000-0002-6060-1233>

## REFERENCES

- Bennett LM, Gadlin H. Collaboration and team science: from theory to practice. *J Invest Med* 2012;60:768–75.
- Ledford H. How to solve the world's biggest problems. *Nature New Biol* 2015;525:308–11.
- Stokols D, Hall KL, Taylor BK, et al. The science of team science: overview of the field and introduction to the supplement. *Am J Prev Med* 2008;35:S77–89.
- Salas E, Reyes DL, McDaniel SH. The science of teamwork: Progress, reflections, and the road ahead. *American Psychologist* 2018;73:593–600.
- Nielsen MW, Alegria S, Börjeson L, et al. Gender diversity leads to better science. *Proc Natl Acad Sci USA* 2017;114:1740–2.
- Horwitz SK, Horwitz IB. The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography. *J Manage* 2007;33:987–1015.
- Levites Strekalova YA, Qin Y, McCormack WT. *Strategic Team Science*: Scaffolded training for research self-efficacy, interdisciplinarity, diversity, equity, and inclusive excellence in biomedical research. *J Clin Transl Sci* 2021;5:e195.
- Ofili EO, Tchounwou PB, Fernandez-Repollet E, et al. The Research Centers in Minority Institutions (RCMI) Translational Research Network: Building and Sustaining Capacity for Multi-Site Basic Biomedical, Clinical and Behavioral Research. *Ethn Dis* 2019;29:135–44.
- NIH Common Fund. Frequently asked questions | nih common fund. Available: <https://commonfund.nih.gov/HuBMAP/generalfaq> [Accessed 21 Mar 2024].
- Jain S, Pei L, Spraggins JM, et al. Advances and prospects for the Human BioMolecular Atlas Program (HuBMAP). *Nat Cell Biol* 2023;25:1089–100.
- Kirkham JJ, Davis K, Altman DG, et al. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. *PLoS Med* 2017;14:e1002447.
- Romney AK, Weller SC, Batchelder WH. Culture as Consensus: A Theory of Culture and Informant Accuracy. *Am Anthropol* 1986;88:313–38.
- Weller SC. Cultural Consensus Theory: Applications and Frequently Asked Questions. *Field methods* 2007;19:339–68.
- Dressler WW, Borges CD, Balieiro MC, et al. Measuring Cultural Consonance: Examples with Special Reference to Measurement Theory in Anthropology. *Field methods* 2005;17:331–55.
- Caulkins D, Hyatt SB. Using Consensus Analysis to Measure Cultural Diversity in Organizations and Social Movements. *Field methods* 1999;11:5–26.
- Smith CS, Morris M, Hill W, et al. Cultural consensus analysis as a tool for clinic improvements. *J Gen Intern Med* 2004;19:514–8.
- Batchelder WH, Anders R, Oravec Z. *Cultural consensus theory. stevens' handbook of experimental psychology and cognitive neuroscience*. Ltd 2018:1–64: John Wiley & Sons,
- Borgatti SP, Halgin DS. Elicitation techniques for cultural domain analysis. *The Ethnographer's Toolkit* 1999;3:115–51.
- Guest G, Bunce A, Johnson L. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods* 2006;18:59–82.
- Hamilton AB, Finley EP. Qualitative methods in implementation research: An introduction. *Psychiatry Res* 2019;280:112516.
- Levites Strekalova YA, Modjarrad L, Midence S. Mechanisms of young professional engagement in health policy development: a cultural domain approach. *Front Public Health* 2024;12:1389649.
- HuBMAP. General frequently asked questions. 2017. Available: <https://commonfund.nih.gov/hubmap/generalfaq>
- Owen WF. Interpretive themes in relational communication. *Quarterly Journal of Speech* 1984;70:274–87.
- Strong AE, White TL. Using paired cultural modelling and cultural consensus analysis to maximize programme suitability in local contexts. *Health Policy Plan* 2020;35:115–21.
- Saldaña J. Coding Techniques for Quantitative and Mixed Data. *The routledge reviewer's guide to mixed methods analysis routledge* 2021.
- Wencelius J, Garine E, Raimond C. FLARES. 2017.
- Gerritsen S, Harré S, Rees D, et al. Community Group Model Building as a Method for Engaging Participants and Mobilising Action in Public Health. *Int J Environ Res Public Health* 2020;17:3457.
- Jandhyala R. Delphi, non-RAND modified Delphi, RAND/UCLA appropriateness method and a novel group awareness and consensus methodology for consensus measurement: a systematic literature review. *Curr Med Res Opin* 2020;36:1873–87.
- Arakawa N, Bader LR. Consensus development methods: Considerations for national and global frameworks and policy development. *Res Social Adm Pharm* 2022;18:2222–9.
- Hohmann E, Cote MP, Brand JC. Research Pearls: Expert Consensus Based Evidence Using the Delphi Method. *Arthroscopy: the journal of arthroscopic & related surgery* 2018;34:3278–82.
- de Loë RC, Melnychuk N, Murray D, et al. Advancing the State of Policy Delphi Practice: A Systematic Review Evaluating Methodological Evolution, Innovation, and Opportunities. *Technol Forecast Soc Change* 2016;104:78–88.
- Turoff M, Linstone HA. The delphi method: techniques and applications.
- Trevelyan EG, Robinson PN. Delphi methodology in health research: how to do it? *Eur J Integr Med* 2015;7:423–8.
- McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm* 2016;38:655–62.
- Lacy MG, Snodgrass JG, Meyer MC, et al. A Formal Method for Detecting and Describing Cultural Complexity: Extending Classical Consensus Analysis. *Field Methods* 2018;30:241–57.
- Sim J, Lewis M. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. *J Clin Epidemiol* 2012;65:301–8.
- Muresherwa E, Loyiso CJ. The Value of a Pilot Study in Educational Research Learning: In Search of a Good Theory-Method Fit. *JESR* 2022;12:220.
- Bell ML, Whitehead AL, Julious SA. Guidance for using pilot studies to inform the design of intervention trials with continuous outcomes. *Clin Epidemiol* 2018;10:153–7.
- Hertzog MA. Considerations in determining sample size for pilot studies. *Research in Nursing & Health* 2008;31:180–91.
- Kim Y. The Pilot Study in Qualitative Inquiry: Identifying Issues and Learning Lessons for Culturally Competent Research. *Qual Soc Work* 2011;10:190–206.
- Aschbrenner KA, Kruse G, Gallo JJ, et al. Applying mixed methods to pilot feasibility studies to inform intervention trials. *Pilot Feasibility Stud* 2022;8:217.
- Aikens RC, Greaves D, Baiocchi M. A pilot design for observational studies: Using abundant data thoughtfully. *Stat Med* 2020;39:4821–40.

- 43 Teresi JA, Yu X, Stewart AL, *et al.* Guidelines for Designing and Evaluating Feasibility Pilot Studies. *Med Care* 2022;60:95–103.
- 44 Rubio DM, Berg-Weger M, Tebb SS, *et al.* Objectifying content validity: Conducting a content validity study in social work research. *Soc Work Res* 2003;27:94–104.
- 45 Rubio DM, Blank AE, Dozier A, *et al.* Developing Common Metrics for the Clinical and Translational Science Awards (CTSAs): Lessons Learned. *Clin Transl Sci* 2015;8:451–9.
- 46 Davis LL. Instrument review: Getting the most from a panel of experts. *Appl Nurs Res* 1992;5:194–7.
- 47 May C. A rational model for assessing and evaluating complex interventions in health care. *BMC Health Serv Res* 2006;6:86.
- 48 May CR, Albers B, Bracher M, *et al.* Translational framework for implementation evaluation and research: a normalisation process theory coding manual for qualitative research and instrument development. *Implement Sci* 2022;17:19.