

Philosophical Transactions A **Proposal for a theme issue**

Continuing its long history of influential scientific publishing, *Philosophical Transactions A* publishes high quality theme issues on topics of current importance and general interest within the physical, mathematical and engineering sciences, guest-edited by leading authorities and comprising new research, reviews and opinions from prominent researchers. Each issue aims to create an original and authoritative synthesis, often bridging traditional disciplines, which showcases current developments and provides a foundation for future research, applications and policy decisions.

Potential Guest Editors should consult our web page on **editing a theme issue**:

<https://royalsocietypublishing.org/rsta/guest-editors>

We have a page budget of 200 printed pages per issue, however we recognise that some issues will be longer and some shorter, and within an issue some articles will be longer than others. For this reason, we do not enforce a strict page limit, however if your proposal is accepted, guidelines will be sent to the authors, and you should try to ensure that authors are aware of this page budget. There is flexibility regarding the number of articles that comprise the theme issue, but the recommended range is 12-18 papers.

As Guest Editor, you are responsible for ensuring the quality of the scientific content of the manuscripts and for managing the peer review process via ScholarOne Manuscripts. In addition to this, the Guest Editors are expected to write an introductory article for the issue, which should provide both the expert and the general reader with the appropriate context for the topic.

In addition to the Introduction, each Guest Editor may only be the lead or final author for one paper, and be a listed author on one additional paper (for a total of 3 articles only). Other authors may also only be involved in 2 papers unless special justification is given and agreed by the Editorial Office.

This proposal will be reviewed by a member of the journal's Editorial Board and at least two independent referees. Following the review process, the Editor will recommend acceptance, rejection, or revisions. If revisions are required, you will be expected to submit a revised proposal within four weeks.

Please complete this form with as much information as you have available. When complete, the form should be submitted to <http://mc.manuscriptcentral.com/ptrsa>. Any queries should be addressed to Alice Power, Commissioning Editor, at philtransa@royalsociety.org.

PHILOSOPHICAL TRANSACTIONS A

Theme title (maximum 150 characters)

The title of the theme issue should be short, informative, and attract interest from both experts in the field and a general audience.

Workflow for Applied Data Analysis

Theme editor(s) with affiliations:

Andrew Gelman, Department of Statistics and Department of Political Science, Columbia University, New York

Aki Vehtari, Department of Computer Science, Aalto University, Espoo, Finland

Richard McElreath, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Elizabeth Wolkovich, Forest and Conservation Sciences, University of British Columbia, Vancouver, Canada

Summary of issue (100–200 words)

Please provide a brief synopsis of the subject area to be covered in the issue. The description should explain the importance of the topic, be concise, free of jargon and understandable to a non-specialist.

The topics of this issue will cover *workflow*—that is, the steps of building statistical models and connecting them to scientific theories, validating statistical methods, communication of results, scientific criticism, and other aspects of research that are central to building trust in the scientific process. This topic has become increasingly important in an era when open data are reshaping how we approach science at the same time that replication crises across fields are redefining the public's perception of science. The individual articles will address different conceptual and computational aspects of workflow in scientific and statistical modeling, including a range of areas of application in the social and biomedical sciences.

Explain why the topic is appropriate for a theme issue; indicate the new developments that will be presented and discussed, and explain why these are timely (maximum 500 words)

Use this section to explain why a theme issue will be appropriate, rather than papers published individually. Where necessary or appropriate, summarise any relevant background to the subject area, emphasising how the issue will build upon and extend previous work. Tell us why the issue is timely and whether there is anything particular happening in the field at this time.

Textbook presentations of the scientific method focus on theories, models, and formal statistical inference. In the past decade or so, things have changed. The replication crisis in science has revealed major problems with the system of peer review, publication, and publicity of science. New methods in machine learning have made it possible to make progress in many areas of science and engineering using minimal substantive theory.

The intention of our proposed theme issue is to bring all this together to provide a cohesive overview of the value of statistical data analysis workflow and how it can adapt to changes in science, in a way that is not possible through individual disparate articles. A themed issue will allow both overviews and opinions of data analysis workflow today, and provide new research with examples in the topic, to help researchers adopt these methods. The issue will include: reviews of recent developments in scientific and statistical workflow; new research on the application of modern workflow in biomedicine, economics, and other areas of natural and social science; discussions how to make these methods more widely available to researchers; and ideas on integrating developments of workflow in science and statistics, connecting ideas of measurement and replication to methods for building trust when integrating diverse sources of data.

What is the likely impact of the proposed Theme Issue? (maximum 300 words)

Please explain the benefits of this issue to the scientific community and how it could influence future research in this area. If the topic is likely to have wider public interest or influence policy decisions, please provide further details here.

We envision this theme issue of the journal as benefiting the academic areas of statistics and machine learning—including their applications across mathematics, physics, engineering and biology—by bringing a focus to ideas of workflow that have typically appeared in different places and in particular application areas without a clear overarching framework. This is important not just because of the new theoretical perspectives and worked examples presented in the articles in the journal but also because some of the papers particularly address questions of communication of workflow ideas to researchers who otherwise would not be aware of these ideas and methods.

This issue will have special impact from its eclectic set of methodological, theoretical, and applied articles. Within statistics and machine learning, the impact of the theoretical and methodological ideas (in particular in the first four and last two articles in the collection) will be enhanced through explorations of these ideas in computation and in a wide range of real applications in the physical, biological, and social sciences. From the other direction, the presence of articles by leaders in several different applied areas will raise the profile of the issue within these different fields. In this issue, we are ambitiously aiming for an impact across science, not just within statistics and computing.

Are these topics being covered in other journals/books, or in meetings or symposia? (maximum 300 words)

Please provide us with details of recent or forthcoming publications (books, theme issues) covering this topic. If the topic is being covered in other journals, how will the proposed theme

PHILOSOPHICAL TRANSACTIONS A

issue be distinct? Are these topics being covered in any meetings or symposia (either by you or by others)?

Three of the editors are currently completing a book on Bayesian workflow, but that is narrower than what we envision for this theme issue, which is not limited to the Bayesian paradigm in the statistical methods it will cover. There are also articles scattered throughout the statistics and engineering literatures on different aspects of workflow in modeling and computing. What is lacking is a big picture, linking specifics of statistics computational workflow to broader questions of replication, communication, and criticism in science. We envision this theme issue as a big step in bringing these ideas together and conveying their importance.

The topic of workflow has different meanings in different branches of science and engineering. We emphasize that the “workflow” discussed in this issue refers to statistical and scientific workflow—the traditionally-informal steps of building and checking models—not the software and hardware workflows discussed in the computing literature. We have tried to clarify this in the revised title of our proposal and we will also make it clear in the survey articles that will appear in the issue itself.

What fields are brought together in this proposal? (maximum 200 words)

Philosophical Transactions A covers the entire range of the physical, mathematical and engineering sciences and so is well-suited for topics that cover multiple disciplines. If the proposed issue will bring together a range of subject areas, please provide details here.

Statistics, anthropology / sociology of science, ecology, economics, political science, demography, education research, psychology

List of contributions

The majority of authors should have tentatively agreed to contribute to the proposed issue according to an approximate timetable that you may have provided. If any contributors have not yet indicated their willingness to contribute, please mention this.

The proposed contributions should provide a comprehensive coverage of the subject area. You should select authors on the basis of their scientific expertise and, without compromising this, please consider the balance of the authorship in respect of gender and nationality. In addition to the Introduction, each Guest Editor may only be the lead or final author for one paper, and be a listed author on one additional paper (for a total of 3 articles only). Other authors may also only be involved in 2 papers unless special justification is given and agreed by the Editorial Office.

As a guide, approximately half of the papers should present new research, with the other half consisting of review or review-type articles. The exact balance of reviews to research will differ by topic so this isn't a fixed requirement, but we would encourage you to commission a range of article types in order to tie the issue together. Visit the [journal website](#) for more information about the article types Philosophical Transactions A supports.

PHILOSOPHICAL TRANSACTIONS A

Please note that the journal will not put out a call for papers for your theme issue if your proposal is accepted. Only authors included in your proposal form will be invited to submit to the theme issue.

For each paper, please include the following information:

- A tentative title (this should be short and of broad interest)
- The name and affiliation of the lead author and the names of any co-authors if known at this stage
- The paper type (original research, review, opinion piece etc)
- A brief description of the information to be presented in the article. This should be substantial enough for reviewers to get a good sense of the likely content of the paper.

Title: More than one way to peel an orange? Exploring the similarities and differences of statistical workflows

Author: Lauren Kennedy, School of Computer and Mathematical Sciences, University of Adelaide

Paper Type: Opinion Piece / Original Research

Description: Statistical workflows help to characterise not just the end result of statistical analysis, but the process undertaken to come to this result. By exploring and discussing workflows formally, we can begin to discuss processes that help to produce reliable statistical results, including protecting against human biases in decision making. While this discussion isn't new, dating back to Tukey, recent years have led to a greater diversity of proposed workflows. To name just a few, this could include the workflow of Tukey, pre-registration practices advocated for in psychology, workflows that incorporate computational checking such as that proposed for Bayesian workflow and even workbench approaches common in data science. In this article I will broadly overview the origin and context that led to the development of these workflows. I then discuss whether workflow differences are driven by the data and research question of the field in which these workflows dominate, and the potential impact of sociological factors. Taking a meta-workflow perspective will give greater clarity in not only the how but the why of our statistical processes.

Title: Unsupervised machine learning for scientific discoveries

Authors: Genevera Allen, Professor of Statistics, Columbia University, Tiffany Tang, University of Notre Dame, Tarek Zikry, Columbia University

Paper type: Review / Best practices / Case study

Description: This paper will discuss how unsupervised machine learning methods are used to make scientific discoveries from large and complex unlabeled data. We will specifically focus on best practices for a reproducible and reliable scientific workflow when using unsupervised tools like clustering and dimension reduction. The paper will highlight both classical techniques like principal components analysis, K-means, and hierarchical clustering, as well as modern approaches like UMAP and autoencoders. We will also highlight and include examples from several application areas including omics and multi-omics, neuroscience, chemometrics, and astronomy. Our discussion of workflow will include best practices at several stages including data acquisition, data pre-processing, feature engineering, data exploration, unsupervised learning, validating results, and finally interpreting and visualizing results.

Title: Realistic elicitation of effect sizes

Author: Megan Higgs, Statistical Consultant, Critical Inference LLC

Paper type: Original research / Opinion

Description: We present an exercise to facilitate scientific workflow by helping people to more meaningfully connect the larger research context with the scale of the primary quantitative parameter of interest, in the design phase. The goal is to get people to go through the challenge of interpreting whatever intervals or posterior distributions on which they will eventually rely. This essentially means walking through hypothetical possibilities to define a “quantitative backdrop” to use for connecting conclusions to the quantitative scale (rather than relying on arbitrary cut-offs, defaults, etc.). The aim is to force researchers understand what a parameter of interest actually represents and go through the process of assigning meaning to different ranges of values—which can lead to identifying challenges early on and forcing discussions among stakeholders that wouldn’t otherwise happen—it also strengthens justification for conclusions if an interpretation plan has been laid out a priori.

Title: A preliminary data analysis workflow for meta-analysis of dependent effect sizes

Author: James E. Pustejovsky, Associate Professor of Statistics, School of Education, University of Wisconsin, Madison, Elizabeth Tipton, Professor of Statistics and Data Science, Northwestern University, Jingru Zhang, Ph.D. Student in Educational Psychology, University of Wisconsin, Madison

Paper type: Original research

Description: Meta-analyses of social science research findings often involve dependent effect sizes and hierarchical data structures. Statistical methods for analyzing dependent effect sizes are now well developed, but there has been less attention to the preliminary stages of data analysis, prior to formal modeling. We propose a generic workflow for initial, exploratory analyses of meta-analytic databases, which seeks to help validate the integrity of the input data and to inform decisions about formal statistical modeling. The workflow entails creating summaries and visualizations of features of the primary studies included in the meta-analysis in order to understand the distribution and structure of the data, especially with respect to between- and within-study variation. We illustrate the workflow using data from previously published meta-analyses and discuss connections between the preliminary analysis and subsequent statistical modeling strategies.

Title: Scientific workflow in social science lab and online experiments

Authors: Anna Dreber, Professor of Economics, Stockholm School of Economics, Séverine Toussaert, Associated Professor of Economics, Oxford University

Paper type: Review / Opinion

Description: Lab and online experiments are nowadays common in many parts of the social sciences, where the goal is typically to either test treatment effects or elicit parameter values for concepts like risk or social preferences. While the role of pre-registration of designs and analyses is increasingly discussed and embraced to different extent in different fields, many other parts of the workflow receive less attention in the communicated output. We will discuss what we believe is the common workflow for lab and online experiments in the social sciences. Some of this will be based on a recent survey we did with experimental economists, where they were asked about pilots and pre-registration, among other things. We will also discuss potential ways to test how the workflow could be improved in terms of what we can learn from a single study, in particular when something “fails” along the way.

Title: Veridical data science workflow with case studies in biomedicine

Authors: Bin Yu, Professor of Statistics, University of California, Berkeley, Jean Yang, Professor of Statistics, University of Sydney

Paper type: Original research / Review

Description: Globally, scientists now have the ability to generate a vast amount of high throughput biomedical data that carry critical information for health and medical discovery. The ability to have efficient, and purpose-specific data analysis approaches can assist in accelerating such scientific discovery. This paper applies and expands the veridical data science framework for responsible data analysis to develop a comprehensive workflow and documentation for practitioners to implement. It encourages users to consider and assess in context multiple statistical formulations for the same scientific problem, to consider and assess in context multiple data sources, data perturbations, model/algorithm perturbations, and their impacts on the final scientific conclusions. That is, the workflow development will be guided by the PCS (predictability-computability-stability) framework for reality checks, stability analyses and aggregations of PCS-vetted results for final conclusions. It will be demonstrated through two case studies in cancer detection, including one at single cell resolutions.

Title: Replicable workflow for online AI in digital health

Authors: Kelly Zhang, Daiqi Gao, Asim Gazi, Bhanu Gullapalli, Susan A. Murphy, Anna Trella, Ziping Xu, Susobhan Ghosh, Imperial College and Harvard University

Paper type: Original research/ Perspective

Description: Digital health promises to help individuals who are struggling with substantial health challenges such as cardiovascular disease and cancer recovery, learn and maintain healthy behaviors including adherence to medications, healthy eating and maintaining physical activity. These health interventions provide access to a variety of behavioral treatments, such as behavior monitoring, motivational support, via smart devices and wearables, as individuals go about daily life. AI algorithms, in particular online decision-making algorithms, are being deployed to determine when and in which settings different behavioral treatments will be delivered to individuals. This is a rapidly evolving area as digital technology including sensor technology, software advances and smart devices changes quickly. Thus, the development and deployment of digital interventions is ongoing, with deployments of the digital interventions, including the decision-making algorithm interspersed with (re-)development of the intervention and algorithm. The data from deployments is used both in the redevelopment of the algorithm as well as to aid in scientific discovery with the goal of improving understanding of the causal mechanisms at work in behavior change.

This rapidly changing area proposes multiple challenges to replicable scientific workflow, including collecting the appropriate data, documenting data provenance and documenting the AI decision-making algorithms used to collect that data. There is a tension between (a) ability to implement (storage and computation constraints), (b) AI algorithm optimization of behavioral treatment delivery, and (c) the utility of the data for scientific discovery, redevelopment of the AI algorithm, all with impact on replicability of the entire data pipeline. In this paper we discuss our work in improving the replicability of the data workflow, across multiple use-cases in digital health.

Title: Model formulation and monthly validation for the production of employment survey-based data products

Authors: Julie Gershunskaya, Office of Employment and Unemployment Statistics, U.S. Bureau of Labor Statistics, Terrance D. Savitsky, Office of Survey Methods Research, U.S. Bureau of Labor Statistics

Paper type: Original research and commentary

Description: The U.S. Bureau of Labor Statistics (BLS) publish employment estimates in domains indexed by industry and geography monthly from the Current Employment Statistics (CES) survey of business establishments. Our data consists of both point estimate and variance statistics formulated for each domain from underlying survey responses. Both the point estimates and variances reflect survey, measurement and approximation errors that are removed using models to uncover latent true values for both domain-indexed point estimates and variances. It is common for a subset of geographies within each industry to express underlying "true" employment changes that are notably different from other domains. We term these true deviations as "inliers" and we describe our model building workflow to accurately discover and differentiate inliers from outliers. The selected model is subsequently used on a monthly basis to produce published estimates for domains. A production validation workflow is described for nominating domains where the observed data values may not arise from the model-based estimator. We term such nominations as "discoveries" and human intervention is initiated to determine if each discovery is true and to uncover the root cause.

Title: Unlocking health coverage insights in low- and middle-income countries: Estimation and forecasting with diverse data sources

Authors: Shauna Mooney, Postgraduate Student in Mathematics and Statistics, Maynooth University, Niamh Cahill, Associate Professor of Mathematics and Statistics, Maynooth University, Leontine Alkema, Professor of Biostatistics, University of Massachusetts, Amherst

Paper type: Original research

Description: The Countdown to 2030 project aims to support low- and middle-income countries (LMICs) to track progress in improving Reproductive, Maternal, Newborn, Child and Adolescent Health and Nutrition (RMNCAH+N). Tracking progress of relevant health indicators, such as immunization coverage or coverage of care related to pregnancy and birth, is often challenging due to limited data availability and quality issues. While Bayesian statistical models have been successfully used for selected indicators in LMICs to triangulate information from different data sources to produce estimates and short-term forecasts, there is no standardized approach to develop such models for a wider range of indicators and populations, with varying data availability and temporal dynamics. We develop a workflow to standardize the development of Bayesian models to produce estimates and short-term forecasts of RMNCAH+N coverage indicators for LMICs at the national and subnational level. We build upon previous work that produces estimates and forecasts for family planning indicators using data from surveys and routine data collection systems. In collaboration with countries in the Countdown initiative, we apply the workflow to produce models for selected indicators.

Title: Machine learning in climate systems model development

Author: Tian Zheng, Professor of Statistics, Columbia University

Paper type: Review

Description: Climate systems models are crucial for understanding and projecting the complex dynamics of Earth's climate. Machine learning has opened new pathways for deriving scientific insights for improving such models and overcoming computational bottlenecks. This paper provides a comprehensive overview of how machine learning is used in the development and improvement of climate systems models. We discuss the design of model training and evaluation workflows that leverage machine learning algorithms to optimize performance and address uncertainties inherent in climate modeling. The paper also addresses the challenges and opportunities presented by the increasing availability of diverse observed climate data, and how machine learning techniques can be adapted to make the most of these resources. In this paper, we hope to showcase and explain transferable machine learning workflow modules that could guide future integration of machine learning into climate systems modeling research and shed light on some open challenges.

Title: "Measure twice, cut once": Scientific workflow for studying children's thinking

Authors: Lauren Girouard-Hallam, Postdoctoral fellow in Psychology, University of Michigan, Susan A. Gelman, Professor of Psychology, University of Michigan

Paper type: Overview

Description: As cognitive developmental psychologists, much of the focus of our scientific workflow is centered on the children that we study. A paper on our workflow would therefore include topics such as: (a) developing and validating measures with age-appropriate items in order to collect viable quantitative data for analysis; (b) the role of pilot testing in this process; and (c) inclusion/exclusion criteria, including the role of children's performance on warm-up trials. In the wake of COVID-19, this workflow has also included sampling considerations for online and hybrid experimental protocols.

Title: Closing the gap between statistical and scientific workflows for improved forecasts in ecology

Author: Elizabeth Wolkovich, Associate Professor of Forest and Conservation Sciences, University of British Columbia

Paper type: Opinion

Description: Increasing biodiversity loss and climate change have led to greater demands for useful ecological models and forecasts. Relevant datasets to meet these demands have also increased in size and complexity, including in their geographical, temporal and phylogenetic scales. While new research often suggests that accounting for these complexities variously increases, removes or otherwise alters major trends, I argue that the fundamental approach to model fitting in ecology makes it impossible to evaluate and compare models. These problems stem in part from continuing gaps between statistical workflows—where the data processing and model development are often addressed separately from the ecological question and aim—and scientific workflows, where all steps are integrated. Yet, as ecologists become increasingly computational, and new tools make it easier to share data, the opportunity to close this gap has never been greater. I outline how increased data simulation at multiple steps in the scientific workflow could revolutionize our understanding of ecological systems, yielding new insights. Combining these changes with more open model and data sharing—and developing new efforts to reuse the same data—could be transformative for ecological forecasting.

Title: The role of exploratory data analysis and statistical thinking in genomics workflows

Author: Rafael Irizarry, Professor of Biostatistics, Harvard University

Paper type: Review

Description: The 20th century's digital progress has significantly transformed scientific research. Clinical research in particular has benefited from new measurement technologies that make it possible to observe molecular entities we previously could not. A common feature of these new technologies is that they generate large and complex datasets. For example, instead of focusing on one gene at a time, we can now examine all genes together, shifting the approach from testing specific hypotheses to exploring and discovering new insights. However, the complexity of these data demand statistical expertise to discern meaningful patterns from chance or subtle systematic error, underscoring the critical role of statistics in biomedical research. Unfortunately, clinical education has not kept pace with the demand for data analysis skills. This review will showcase various examples to illustrate the vital role of statistical analysis and effective data visualization in the realm of genomics.

Title: Local robustness for Bayesian posteriors: Some examples and limitations

Authors: Ryan Giordano, Assistant Professor of Statistics, University of California, Berkeley, and Tamara Broderick, Associate Professor of Electrical Engineering and Computer Science, Massachusetts Institute of Technology

Paper type: Original research

Description: We will first review a simple expression for Bayesian local robustness measures and discuss its computation using probabilistic programming and MCMC samples. We will then derive a straightforward expression for the error term such local measures, providing intuition for what sorts of posterior perturbations can and cannot be effectively treated with local approximations. We will augment this intuition with extensive simple examples and experiments.

Title: Bayesian workflow with R-INLA

Authors: David Bolin, Associate Professor of Statistics, King Abdullah University of Science and Technology, Håvard Rue, Professor of Statistics, King Abdullah University of Science and Technology

Paper type: Review / Opinion

Description: R-INLA performs approximate inference for the class of latent Gaussian models (LGMs) and is highly popular in applied statistical work (www.r-inla.org). The key components in LGMs are sums of Gaussian processes, each representing each additive effect indexed by group, time, space, etc. In this paper we will discuss our ongoing work to validate and relax these Gaussian assumptions, using a wider class of latent non-Gaussian models derived from stochastic partial differential equations driven by non-Gaussian noise.

Title: Simulations in statistical workflows

Authors: Paul Bürkner, Professor of Computational Statistics, Technische Universität Dortmund, Florence Bockting, Ph.D. Student in Computational Statistics, Technische Universität Dortmund, Marvin Schmitt, Ph.D. Student in Statistics, University of Stuttgart, Stefan Radev, Assistant Professor of Computational Science, Rensselaer Polytechnic Institute

Paper type: Review / Opinion

Description: Simulations play important and diverse roles in statistical workflows, for example, in the context of model specification, checking, validation, and even directly for model inference. The application areas and overall potential of using simulations in statistical workflows have continued to increase in the past decades following the development of new simulation-based algorithms and steadily increasing computational resources. In the paper, we plan to review past and current trends in this area and provide perspectives on how simulations may change the way we will approach statistics in the future.

PHILOSOPHICAL TRANSACTIONS A

Title: Scientific workflow and statistical workflow

Authors: Aki Vehtari, Professor of Computer Science, Aalto University, Andrew Gelman, Professor of Statistics and Political Science, Columbia University, and Richard McElreath, Director, Max Planck Institute for Evolutionary Anthropology

Paper type: Review

Description: Statistics is the language of science, and statistical research has traditionally focused on mirroring the scientific processes of modeling, hypothesis testing, inference, and prediction. More recently, statisticians have adapted various principles of engineering workflow, including unit testing, fast failure, and iterative model building and comparison. In this paper we connect these developments to ideas in scientific workflow involving replication, measurement, and experimental design.

Diversity

The Royal Society is committed to increasing diversity in science, technology, engineering and mathematics. As part of this, a briefing on [unconscious bias](#) has been prepared, which we ask you to consider. Our expectation is that you will have considered the balance in terms of gender, geographic location and career stage when inviting authors, to ensure that your contributors are representative of your research field. Please provide an approximation breakdown of your invited authors below.*

	Male	Female	Non-binary
Lead author gender (provide number)	6	11	

	UK	Rest of Europe	USA & Canada	South & Central America	Asia	Australasia	Africa
Lead author current location (provide number)	1	4	10	0	1	1	0

	Early (<10 years postgrad)	Mid (10-20 years postgrad)	Senior (>20 years)
Lead author career stage (provide number)	5	6	6

*This will usually be one author per paper, but may be more if you are asking people to write joint contributions. Please just use your judgement to provide an overview of your key contributors.

How do you justify the choice of authors and titles? Please also comment on the diversity data given above if relevant.

PHILOSOPHICAL TRANSACTIONS A

Please explain your reasoning for the choice of contributors to this issue. Also, if relevant, please indicate any deliberate omissions and the reasons for them. We expect you to have made a deliberate effort to invite contributors outside of your own close network. Lastly, please provide any information on diversity if the information given above is not a good representation of the researchers in your field (e.g. poor gender representation for lead authors, few early career researchers, minimal representation from outside North America and Europe, etc.)

The authors are world experts in statistics, machine learning, and biomedical and social sciences. The titles address different aspects of statistical, computational, and scientific workflow and are relevant to efforts to understand and address the replication crisis in science. We have included collaborators outside our close networks and have put a special effort into covering a range of statistical and scientific topics.

Please summarise your contributions to the subject area and any previous editing experience you have.

Please note that decisions on proposals are not made based on the proposed editors' experience. Early career researchers are encouraged to propose issues. However, this information will help us to provide appropriate guidance.

Two of the editors (statistician Gelman and computer scientist Vehtari) are lead authors on an influential article on Bayesian workflow and are currently leading the writing a book with a dozen authors on this topic. The third editor (McElreath) is an anthropologist, author of a textbook linking scientific and data analysis workflows, and has published an influential paper, *The Natural Selection of Bad Science*, giving theoretical insight into the replication crisis. The fourth editor (Wolkovich) is an ecologist who has experience with large research groups bridging scientific theories and data modeling. In addition to their experience in various collaborative groups, Gelman has edited or co-edited three books (*Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, *A Quantitative Tour of the Social Sciences*, and *Handbook of Markov Chain Monte Carlo*). The editorial team is well qualified and ready to do the job.

Do you have a particular schedule in mind?

If you have already discussed a potential submission or publication date with the authors, or if you have a particular schedule in mind for the issue, please provide details. Although we will make every effort to accommodate your timeline, this may not always be possible as the journal's publication schedule is planned far in advance.

We have no particular schedule in mind. We have contacted all the authors listed above, they represent a range of research areas, they are all excited about the topic of the special issue, and they all have committed to contributing to this issue.