Supplemental Material:
Introduction to the Special Collection on the Fragile Families Challenge

Matthew J. Salganik, Ian Lundberg, Alexander T. Kindel, and Sara McLanahan

1   Peer review process for the Special Collection

Papers for the Special Collection were solicited with a call for papers that was circulated to all participants in the Fragile Families Challenge and was posted on the *Socius* website. This call is reproduced below. We received 14 manuscripts from participants, and these went through a process of peer review. They also went through a process of computational reproducibility review described in Liu and Salganik (2019). For the peer review, each manuscript was reviewed by the editors of the Special Collection and about three people. These three people were typically a participant in the Challenge who submitted a manuscript to the Special Collection, a participant in the Challenge who did not submit to the Special Collection, and someone who did not participate in the Challenge. The editors of the Special Collection also provided comments to standardize and harmonize the manuscripts. All manuscripts received a decision of revise and resubmit, which ranged from minor revisions to major revisions.

We received a revised version of all 14 manuscripts. Of these 14 revised manuscripts, 12 received a conditional acceptance based on a review of the editors. Each conditional acceptance came with three conditions: 1) revisions to the manuscript based on feedback, 2) computational reproducibility, and 3) timeliness. We had concerns that the other two manuscripts would not meet the criteria for the Special Collection, which were published in the call for papers. Therefore, we told the authors of these two manuscripts that they needed to submit a detailed revision plan before revising their manuscripts. Eventually, the authors withdrew these two manuscripts.

In addition to the manuscripts submitted by participants in the Fragile Families Challenge, our research team also submitted three papers about our experience running the Fragile Families Challenge (Liu and Salganik, 2019; Kindel et al., 2019; Lundberg et al., 2019). The review and acceptance of these papers was handled by the Editors of *Socius*.

2   *Socius* Call for Papers

Here we reproduce the Call for Papers that was used to solicit articles for the Special Collection.

Special issue on the Fragile Families Challenge
Guest editors: Matthew J. Salganik and Sara McLanahan

Socius, an open access journal published by the American Sociological Association, will publish a special issue on the predictive modeling phase of the Fragile Families Challenge. All participants in the Fragile Families Challenge are invited to submit a manuscript to this special issue.

A strong manuscript for the special issue will describe the process of creating a submission to the Challenge and will describe what was learned during that process. For example, a strong manuscript will describe the different approaches that were considered for data preprocessing, variable selection, missing data, model selection, and any other steps involved in creating the final submission to the Challenge. Further, a strong manuscript will also describe how the authors decided among the many possible approaches. Finally, some manuscripts may seek to draw more general lessons about social inequality, families, the common task method, social science, data science, or computational social science. Manuscript should be written in a style that is accessible to a general scientific audience.

The editors of the special issue may also consider other types of manuscripts that are consistent with the scientific goals of the Fragile Families Challenge. If you are considering submission a manuscript different from what is described above, please contact the editors of the special issue at fragilefamilieschallenge@gmail.com before submitting your manuscript.

All papers will be peer reviewed, and publication is not guaranteed. However, there is no limit on the number of articles that will be accepted in the special issue. All published papers must abide by the terms and conditions of the Fragile Families Challenge, and must be accompanied by open source code and a data file containing predictions.

Submissions for the special issue must be received through the *Socius* online submission platform by Sunday, October 1, 2017 Monday, October 16 at 11:59pm ET. If you have any questions about the special issue, please email fragilefamilieschallenge@gmail.com.

FAQ:

- *Do I need to describe an approach to predicting all six outcome variables in order to submit to the special issue?*

  No. We will happily consider papers that focus on one specific outcome variable.

- *Do I need to have a low mean squared error in order for my paper to be published?*

  No. Predictive performance in the held-out dataset is only part of what we will consider. For example, a paper that clearly shows that many common strategies were not very effective would be considered a valuable contribution.

- *What if I can't afford the Article Processing Charge?*

  Socius, like most open access journals, has an Article Processing Charge. This charge is required to keep *Socius* running, and it is in line with the charges at other open access journals. However, we strongly believe that the Article Processing Charge should not be a barrier to scientific participation. Therefore, the Fragile Families Challenge project will pay the Article Processing Charge for all accepted articles submitted by everyone except for tenure-track (or equivalent) faculty working in universities in OECD countries. In other words, we will cover the Article Processing Charge for undergraduates, graduate students, post-docs, and people working outside of universities. Further, we will pay the Article Processing Charge for all tenure-track (or equivalent) faculty working in universities outside the OECD.

If for any reason you think that the Article Processing Charge may be a barrier to your participation, please send us an email and we will try to find a solution: fragilefamilieschallenge@gmail.com.

- *How will you decide what manuscripts to accept for publication?*

  Articles in *Socius* are judged by four criteria: Accuracy, Novelty, Interest, and Presentation. In the case of this special issue, these criteria will be judged by the editors of the special issue, with feedback from reviewers and the editors of *Socius*. For the purposes of this special issue, here is how these criteria will be interpreted:

  - **Accuracy:** The key question is whether this analysis was conducted appropriately and accurately. Were the techniques used in the manuscript performed and interpreted correctly? Do the claims in the manuscript match the evidence provided?
  - **Novelty:** The key question is whether the manuscript will be novel to some social scientists or some data scientists. Because projects like the Fragile Families Challenge are not yet common, we expect that most submitted manuscripts will be somewhat novel.
  - **Interest:** The key question for the editors is whether the manuscript will be interesting to some social scientists or some data scientists. Will some people want to read this paper? Does it advance understanding of the Fragile Families Challenge and related intellectual domains?
  - **Presentation:** The key question is whether this manuscript communicates effectively to a diverse audience of social scientist and data scientists. We will also assess whether the figures and tables are presented clearly and whether the manuscript makes appropriate use of the opportunity for supporting online information. Because these manuscripts will be short, we expect that the supporting online information will play a key role.

- *Who is the audience for these papers?*

  All papers should be written for a general scientific audience that will include both social scientists and data scientists (broadly defined). In other words, when writing your paper you should imagine an audience similar to the audience at journals such as *Science* and *Proceedings of the National Academies of Sciences* (PNAS). We would recommend reading some articles from these journals to get a sense of this style. Manuscripts that use excessive jargon from a specific field will be asked to make revisions.

  Manuscripts should follow the length guidelines of a Report published in *Science*: 2,500 words, with up to 4 figures or tables. Additional materials should be included in supporting online materials. We will consider articles that deviate from these guidelines in some situation. Other aspects of the manuscript format will follow standard *Socius* rules.

- *Should we describe the Fragile Families Challenge in our paper?*

  No. There is no need to describe the Challenge in your paper. The special issue will have an introductory article describing the Challenge and data. You should assume that your readers

will already have this background information. [Note: This was a mistake and we later told participants to briefly describe the Challenge in their manuscript in order to make it more self-contained.]

- *Will the articles go through peer review?*

  Absolutely. All manuscripts will be reviewed by at least two people. Possible reviewers include: members of the board of the Fragile Families Challenge, qualified participants in the Challenge, members of the general reviewer pool at *Socius*, and other qualified researchers.

- *What are the requirements for the open source code?*

  The code must take the Fragile Families Challenge data files as an input and produce (1) all the figures and tables in your manuscript and supporting online materials and (2) your final predictions. The code can be written in any language (e.g., R, stata, Python). The code should be released under the MIT license, but we will consider other permissive licenses in special situations.

- *How long will the review process take?*

  We don't know exactly, but we are excited about having these results in the scientific literature as quickly as possible. Therefore, we will work as quickly as possible while maintaining the quality standards of the Fragile Families Challenge and *Socius*.

- *Will I have access to the holdout data when writing my paper?* (added July 20, 2017)

  No, but we will allow you to request scores for your models on the holdout as described in this blog post.

- *Will I have access to the Challenge data when writing my paper?* (added July 27, 2017)

  Yes. If you will submit to the Special Issue you can continue to use the Challenge data until the Special Issue is published. If you are not submitting to the Special Issue, then you should delete the Challenge data file on August 1. Finally, participants who want to continue to do non-Challenge related research with the Fragile Families and Child Wellbeing Study can, at any time, apply for access to the core Fragile Families data by following the instructions here: http://www.fragilefamilies.princeton.edu/documentation.

- *What if I want to submit to the special issue but I can't exactly reproduce my submission to the Challenge?* (added September 23, 2017)

  Everyone in the Challenge was supposed to uploaded their code. But there are several reasons why they might not be able to use their code to reproduce their submission such as forgetting to set their seed or changes to the packages that were used in the submission (if you are interested, here are some general tips for promoting reproducibility Sandve et al. (2013) "Ten Simple Rules for Reproducible Computational Research.")

  If there is a tension between making your paper reproducible and making it match the submission to the Challenge exactly, you should opt to make your paper reproducible. If the code and predictions that you submit with your paper don't exactly match what you submitted to the Challenge, you should include a note in your supporting online material explaining

these differences and why they occurred. If this note will require addition information from us such as the score of your reproducible results in the leaderboard data we will provide it to you. We are happy to help you with these issues on a case-by-case basis.

- *I have another question, how can I ask it?*

  Send us an email: fragilefamilieschallenge@gmail.com.

References

Kindel, A., V. Bansal, K. Catena, T. Hartshorne, K. Jaeger, D. Koffman, S. McLanahan, M. Phillips, S. Rouhani, R. Vinh, and M. J. Salganik 2019. Improving metadata infrastructure for complex surveys: Insights from the Fragile Families Challenge. *Socius*, this Special Collection.

Liu, D. M. and M. J. Salganik 2019. Successes and struggles with computational reproducibility: Lessons from the Fragile Families Challenge. *Socius*, this Special Collection.

Lundberg, I., A. Narayanan, K. Levy, and M. J. Salganik 2019. Privacy, ethics, and data access: A case study of the Fragile Families Challenge. *Socius*, this Special Collection.