STATEMENT TO PENNSYLVANIA REAPPORTIONMENT COMMISION REGARDING PROPOSED HOUSE PRELIMINARY PLAN

Kosuke Imai, Ph.D.

January 7, 2022

I. INTRODUCTION AND SCOPE OF WORK

1. My name is Kosuke Imai, Ph.D., and I am a Professor in the Department of Government and the Department of Statistics at Harvard University. I specialize in the development of statistical methods and computational algorithms and their applications to social science research. I am also affiliated with Harvard's Institute for Quantitative Social Science.

2. I have been engaged to analyze relevant data and provide my expert opinions on Professor Michael Barber's expert report, entitled "Memo on Proposed Redistricting Plan from PA Redistricting Commission." More specifically, I have examined how the consideration of race may alter the conclusions of the race-blind redistricting simulation analysis Professor Barber conducted regarding the evaluation of the State House preliminary plan approved by the Legislative Reapportionment Commission. I am continuing my study and look forward to appearing before the Commission next week.

3. Redistricting simulation analysis is a powerful methodology for the empirical evaluation of legislative districting plans. State-of-the-art redistricting simulation algorithms generate a representative sample of all possible plans under a specified set of criteria. This allows analysts to evaluate a proposed plan by comparing it against the simulated alternative plans. Statistical theory lets us quantify the degree to which a proposed plan is extreme relative to the ensemble of simulated plans in terms of partisan outcomes. Statistically significant differences in partisan outcomes between a proposed plan and simulated plans provide empirical evidence that the proposed plan may be a partisan gerrymander.

4. A primary advantage of the simulation-based approach over the traditional redistricting evaluation methods is its ability to account for the political and geographic features that are specific to each state, including spatial distribution of voters and configuration of administrative boundaries. Simulation algorithms can also incorporate each state's redistricting rules. These state-specific features limit the types of redistricting plans that can be drawn, making comparison across states and over different time periods difficult. The simulation-based approach, therefore, enables the comparison of a proposed plan against a representative set of alternate districting plans subject to Pennsylvania's constitutional and legal requirements. Given the importance of legislative

redistricting in representative democracy, I welcome Professor Barber's attempt to assist the Commission through the application of cutting-edge redistricting simulation methodology.

5. I have examined how the consideration of race can alter the conclusions of Professor Barber's race-blind simulation analysis regarding the proposed State House preliminary plan. Professor Barber conducted his simulation analysis without using any information about race. I investigate how the expected partisan outcomes under simulated plans change once race is incorporated into simulation algorithms. My analysis exploits the ability of redistricting simulation methodology to determine how a specific factor influences the types of redistricting plans one could draw while adhering to other redistricting criteria. The key implication is that analysts must carefully choose the inputs to redistricting simulation algorithms based on legal considerations.

6. I first conducted a *race-blind* simulation analysis. Like Professor Barber's race-blind analysis, this simulation analysis does not use any information about race but otherwise is designed to be consistent with the redistricting criteria specified in the Pennsylvania Constitution. I also conducted a race-aware simulation analysis that takes race into consideration when generating simulated plans. The race-aware simulation follows the same set of redistricting criteria, and so the only difference between race-blind and race-aware simulations is the consideration of race. The race-aware simulation analysis ensures that every simulated plan includes a certain number of majority Black and majority Hispanic districts reflective of Pennsylvania's demographics. As discussed below, my race-blind simulation analysis produced substantially different results than Professor Barber's simulation.

7. For each simulation analysis, I generated a representative set of 5,000 alternative plans that could be drawn under the corresponding set of redistricting criteria. I then evaluated the preliminary plan by comparing the likely number of Democratic districts under the preliminary plan with those under each set of 5,000 simulated plans. To make my results comparable with those of Professor Barber's report, I used the same three sets of elections to estimate the likely number of Democratic districts under each simulated plan. Finally, I examined whether and how the consideration of race alters the evaluation of the preliminary plan by comparing the conclusions of the race-blind simulation analysis with the race-aware simulation analysis.

II. SUMMARY OF OPINIONS

8. I find that the consideration of race alters the conclusions of the redistricting simulation analysis. Under the race-blind simulation analysis, the preliminary plan yields a greater number of Democratic districts than simulated plans. In comparison to the race-blind analysis, however, the race-aware simulation analysis reveals that the difference in the likely number of Democratic districts between the preliminary and simulated plans is substantially smaller and is not statistically significant, depending on the set of elections Dr. Barber selected to measure partisan outcome. In fact, the likely number of Democratic districts in the preliminary plan falls within the range of the 2012-2020 and the 2014-2020 simulations.

9. My race-blind simulation analysis also shows that the difference in the number of Democratic districts between the preliminary plan and the race-blind simulated plans is, while still statistically significant, approximately 3 to 4 districts smaller than Professor Barber's race-blind analysis. This finding contradicts with Professor Barber's conclusion that the preliminary plan generates an additional 8 to 10 Democratic districts more than the race-blind simulation.

III. QUALIFICATIONS AND EXPERIENCE

10. I am trained as a political scientist (Ph.D. in 2003, Harvard) and a statistician (MA in 2002, Harvard). I have published more than 60 articles in peer reviewed journals, including premier political science journals (e.g., *American Journal of Political Science*, *American Political Science Review*, *Political Science*), statistics journals (e.g., *Biometrika*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society*), and general science journals (e.g., *Lancet*, *Nature Human Behavior*, *Science Advances*). My work has been widely cited across a diverse set of disciplines. For each of the past four years, Clarivate Analytics, which tracks citation counts in academic journals, has named me as a highly cited researcher in the cross-field category for producing "multiple highly cited papers that rank in the top 1% by citations for field and year in Web of Science."

11. I started my academic career at Princeton University, where I played a leading role in building interdisciplinary data science communities and programs on campus. I was the founding director of Princeton's Program in Statistics and Machine Learning from 2013 to 2017. In 2018, I

moved to Harvard, where I am Professor jointly appointed in the Department of Government and the Department of Statistics, the first such appointment in the history of the university. Outside of universities, between 2017 and 2019, I served as the president of the Society for Political Methodology, a primary academic organization of more than one thousand researchers worldwide who conduct methodological research in political science. My introductory statistics textbook for social scientists, *Quantitative Social Science: An Introduction* (Princeton University Press, 2017), has been widely adopted at major research universities in the United States and beyond.

12. Computational social science is one of my major research areas. As part of this research agenda, I have developed simulation algorithms for evaluating legislative redistricting since the beginning of this emerging literature. At Harvard, I lead the Algorithm-Assisted Redistricting Methodology (ALARM; https://alarm-redist.github.io/) Project, which studies how algorithms can be used to improve legislative redistricting practice and evaluation.

13. Back in 2014, along with Jonathan Mattingly's team at Duke, my collaborators and I were the first to use Monte Carlo algorithms to generate an ensemble of redistricting plans. Since then, my team has written several methodological articles on redistricting simulation algorithms (Fifield, Higgins, et al. 2020; Fifield, Imai, et al. 2020; McCartan and Imai 2020; Kenny et al. 2021).

14. I have also developed an open-source software package titled redist that allows researchers and policy makers to implement the cutting-edge simulation methods developed by us and others (Kenny et al. 2020). This software package can be installed for free on any personal computer with Windows, Mac, or Linux operating system.