Seek a Paradigm and Distrust It? Statistical Arguments and the Representation of Uncertainty

Oliver L. Pescott[1]

[1]**UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, United Kingdom**

Bailey (2023a) does a very useful service to the polling and broader survey sampling communities by highlighting the insights that the Meng (2018) equation provides into current statistical practice in that area. Through the emphasis on the continuous nature of the independent and interactive effects of the data defect correlation (ρ) and population size on error (putting aside the 'problem difficulty'), we see clearly how relying solely on expectations (in both the statistical and everyday senses) can lead us astray in real-world problems. While I fully endorse the general message of Bailey (2023a) that the clarity of the Meng identity forces analysts to take departures from random (or any probability) sampling more seriously, and provides a neat framework for understanding the logical unity of existing weighting-type methods (as elegantly explained by Meng, 2022a), I have two small comments regarding his proposed paradigm shift. The first relates to the seemingly quite strong dismissal of the existing methods within the missing at random (MAR) or conditional ignorability toolbox; the second builds on this to consider the wider problem of the full communication of the potential uncertainty associated with descriptive inferences from nonprobability samples.

From 'Pollwashing' to Poll Cleansing

The insights of Professors Bailey and Meng point toward greater acknowledgment of the potential effects of often dubious assumptions in survey sampling, wherever these may occur. Rather than this leading toward an automatic rejection or distrust of some methods, and an embracing of others, this seems to fall well within the tradition of a wholesome skepticism of all model-based results presented without adequate context and humility (e.g. Freedman, 2010; Greenland, 2021; Hodges, 1996; Mallows, 1998; Saltelli et al., 2020; Swanson, 2023; Thompson, 2022). Bailey (2023a) suggests that the concept of representativeness is essentially a 'pollwashing' technique, in that the claims of a reweighted, or otherwise adjusted, nonprobability sample to represent the population of interest can never be unambiguously evaluated. I struggle to see how this argument can be admitted as a *fatal* weakness of the MAR methods undergirt by this concept, while statistical modeling in general still breathes. We can no doubt agree that some claims of representativeness are better than others, but surely this depends on evaluation(s) of the arguments put forward by those doing the adjusting for any given purpose? Statistical models do not (or should not; Thompson, 2022) exist in a vacuum, and the presentation of population inferences resulting from reweighted samples are nowhere required to be foisted on a public without self-criticism. Auxiliary variables can be evaluated and justified by researchers just as well, or as poorly, as any assumption (by, for example, using theory and qualitative research, or extra-statistical logic such as causal graphs), and targeted sensitivity analyses performed where these are uncertain. If variables considered to be important for adjustment are unavailable, the usefulness of proxy variables can be discussed and investigated.

If the objection of Professor Bailey to representativeness is one of precise boundaries allowing the analyst to claim success in representation, then one immediately thinks of all the other arbitrary, and largely unhelpful, sharp boundaries in everyday use in statistical methods (null hypothesis significance testing, anyone?), and the

extensively developed counterarguments for the importance of more nuanced continuous thinking (e.g. Greenland 2023b, 2023a). No doubt some assertions of representativeness by researchers are vague and meaningless, with little practical merit, but rejecting the concept as misleading or useless seems to brook no admittance of principled, well-theorized adjustments presented with appropriate sensitivity analyses or other caveats. The core concept of representativeness within the MAR framework, that is, whether or not adjusting for some variables that predict both survey response and outcome enables useful population inferences, seems little different in substance to the empirical reality of other key theoretical ideas in statistics, such as exchangeability (Draper et al., 1993). I suspect that none of this is in particular disagreement with Professor Bailey's thoughts, I merely seek to emphasize the sliding scale of error highlighted by Meng (2018), and the checks and balances that might be associated with this, as a more inclusive broadening of the traditional paradigm than one that singles out the potential weaknesses of MAR assumptions only. Data analysts should by now appreciate that all models are wrong; the more important point is whether the missing parts of reality that stalk us are mice or tigers (Box, 1976; Stark 2022). The habit of shining a light on these, and naming them as well as we are able, may allow us to move from accusations of pollwashing to a regular practice of 'poll cleansing.'

Field Guides to Mice and Tigers

Professor Bailey, both here and in his very readable and useful book (Bailey, 2023b), points to a number of ways in which this poll cleansing might be achieved, including his central points about investigating and attempting to minimize data missing not at random (MNAR) situations. As he recognizes more extensively in his book (Bailey, 2023b), other tools are available to ensure that we are not convicted of pollwashing, methods of better bounding our estimates with intervals that acknowledge potential biases and other sources of uncertainty, for example (he highlights Hartman & Huang, 2023, and Manski, 1990). Other techniques in this area have been suggested elsewhere, and these could be further explored by survey samplers (e.g. Coker et al., 2021; Meng, 2022b; Reichardt & Gollob, 1987). In a similar vein of exploring and exposing our assumptions to criticism, graphical plots of auxiliary variable balance before and after adjustment (e.g. Boyd et al., 2023; Makela et al., 2014), quantitative analogues of these in the form of *R*-indicators (Schouten et al., 2012), and transparent statements of the theory lying behind such adjustments in the form of causal graphs (Lee et al., 2023; Mohan & Pearl, 2021) or verbal models (Thompson, 2022), could also be included in reports. Potential indicators of nonignorable selection bias are also being rapidly developed (Andridge et al., 2019; Boonstra et al., 2021; Little et al., 2020). Perhaps more obviously, multiple model fits can also be presented: Little and Rubin (2020, pp. 402–403) describe such a case where both MAR and MNAR models were provided to a client. No doubt there is a rich vein of research here around how such model 'multiverses' (Steegen et al., 2016) can be most effectively communicated for different purposes (e.g. Liu et al., 2021).

The tradition of qualitative 'risk of bias' assessments for the internal validity of experiments and observational studies provides a further source of ideas in this area (Pescott et al., 2023), and I wonder whether reporting

standards could be erected for assessments of polling sample external validity (i.e., generalizability) in the same way that is being attempted for descriptive inferences in other disciplines (Boyd et al., 2022;Simons et al., 2017)? The argument has been made by Sterba and colleagues (2011) that the clear presentation of information relating to nonprobability sampling is an ethical issue, and leading survey sampling statisticians have already called for such considerations to become standard (Lohr, 2022;Valliant, 2023). There seems to be a practical question here for the polling community (in its broadest sense) that I have not seen addressed: if polling samples are fully evaluated for their potential weaknesses, and broader estimates of uncertainty presented, will the resulting polling products be of practical use and interest? I assume that such information, being more realistic, would be useful for betting markets, but it seems likely that the media (and partisan funders?) would find it less attractive. It would certainly be interesting to learn Professor Bailey's opinion on whether poll cleansing would be likely to survive and reproduce within the current environment.

Disclosure Statement

References

Andridge, R. R., West, B. T., Little, R. J. A., Boonstra, P. S., & Alvarado-Leiton, F. (2019). Indices of non-ignorable selection bias for proportions estimated from non-probability samples. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *68*(5), 1465–1483. https://doi.org/10.1111/rssc.12371

Bailey, M. A. (2023a). A new paradigm for polling. *Harvard Data Science Review*, *5*(3). https://doi.org/10.1162/99608f92.9898eede

Bailey, M. A. (2023b). *Polling at a crossroads: Rethinking modern survey research*. Cambridge University Press.

Boonstra, P. S., Little, R. J. A., West, B. T., Andridge, R. R., & Alvarado-Leiton, F. (2021). A simulation study of diagnostics for selection bias. *Journal of Official Statistics*, *37*(3), 751–769. https://doi.org/10.2478/jos-2021-0033

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791–799. https://doi.org/10.1080/01621459.1976.10480949

Boyd, R. J., Powney, G. D., Burns, F., Danet, A., Duchenne, F., Grainger, M. J., Jarvis, S. G., Martin, G., Nilsen, E. B., Porcher, E., Stewart, G. B., Wilson, O. J., & Pescott, O. L. (2022). ROBITT: A tool for assessing

the risk-of-bias in studies of temporal trends in ecology. *Methods in Ecology and Evolution*, *13*(7), 1497–1507. https://doi.org/10.1111/2041-210X.13857

Boyd, R. J., Stewart, G., & Pescott, O. (2023). *Descriptive inference using large, unrepresentative nonprobability samples: An introduction for ecologists*. EcoEvoRxiv. https://doi.org/10.32942/X2359P

Coker, B., Rudin, C., & King, G. (2021). A theory of statistical inference for ensuring the robustness of scientific results. *Management Science*, *67*(10), 6174–6197. https://doi.org/10.1287/mnsc.2020.3818

Draper, D., Hodges, J. S., Mallows, C. L., & Pregibon, D. (1993). Exchangeability and data analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *156*(1), 9–37. https://doi.org/10.2307/2982858

Freedman, D. A. (2010). *Statistical models and causal inference: A dialogue with the social sciences* (D. Collier, J. S. Sekhon, & P. B. Stark, Eds.). Cambridge University Press.

Greenland, S. (2021). Invited commentary: Dealing with the inevitable deficiencies of bias analysis—and all analyses. *American Journal of Epidemiology*, *190*(8), 1617–1621. https://doi.org/10.1093/aje/kwab069

Greenland, S. (2023a). Connecting simple and precise *P*-values to complex and ambiguous realities (includes rejoinder to comments on "Divergence vs. Decision *P*-values"). *Scandinavian Journal of Statistics*, *50*(3), 899–914. https://doi.org/10.1111/sjos.12645

Greenland, S. (2023b). Divergence versus decision *P*-values: A distinction worth making in theory and keeping in practice: Or, how divergence *P*-values measure evidence even when decision *P*-values do not. *Scandinavian Journal of Statistics*, *50*(1), 54–88. https://doi.org/10.1111/sjos.12625

Hartman, E., & Huang, M. (2023). Sensitivity analysis for survey weights. *Political Analysis*, 1–16. Advance online publication. https://doi.org/10.1017/pan.2023.12

Hodges, J. S. (1996). Statistical practice as argumentation: A sketch of a theory of applied statistics. In J. C. Lee, W. O. Johnson, & A. Zellner (Eds.), *Modelling and prediction honoring Seymour Geisser* (pp. 19–45). Springer. https://doi.org/10.1007/978-1-4612-2414-3_2

Lee, K. J., Carlin, J. B., Simpson, J. A., & Moreno-Betancur, M. (2023). Assumptions and analysis planning in studies with missing data in multiple variables: Moving beyond the MCAR/MAR/MNAR classification. *International Journal of Epidemiology*, *52*(4), 1268–1275. https://doi.org/10.1093/ije/dyad008

Little, R. J. A., & Rubin, D. B. (2020). *Statistical analysis with missing data* (3rd ed.). Wiley.

Little, R. J. A., West, B. T., Boonstra, P. S., & Hu, J. (2020). Measures of the degree of departure from ignorable sample selection. *Journal of Survey Statistics and Methodology*, *8*(5), 932–964. https://doi.org/10.1093/jssam/smz023

Liu, Y., Kale, A., Althoff, T., & Heer, J. (2021). Boba: Authoring and visualizing multiverse analyses. *IEEE Transactions on Visualization and Computer Graphics*, *27*(2), 1753–1763. https://doi.org/10.1109/TVCG.2020.3028985

Lohr, S. L. (2022). Comments on "Statistical inference with non-probability survey samples." *Survey Methodology*, *48*(2), 331–338.

Makela, S., Si, Y., & Gelman, A. (2014). Statistical graphics for survey weights. *Revista Colombiana de Estadística*, *37*(2), 285–295.

Mallows, C. (1998). The zeroth problem. *The American Statistician*, *52*(1), 1–9. https://doi.org/10.2307/2685557

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, *80*(2), 319–323.

Meng, X.-L. (2018). Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *The Annals of Applied Statistics*, *12*(2), 685–726. https://doi.org/10.1214/18-AOAS1161SF

Meng, X.-L. (2022a). Comments on "Statistical inference with non-probability survey samples" – Miniaturizing data defect correlation: A versatile strategy for handling non-probability samples. *Survey Methodology*, *48*(2). http://www.statcan.gc.ca/pub/12-001-x/2022002/article/00006-eng.htm

Meng, X.-L. (2022b). Double your variance, dirtify your Bayes, devour your pufferfish, and draw your kidstrogram. *The New England Journal of Statistics in Data Science*, *1*(1), 4–23. https://doi.org/10.51387/22-NEJSDS6

Mohan, K., & Pearl, J. (2021). Graphical models for processing missing data. *Journal of the American Statistical Association*, *116*(534), 1023–1037. https://doi.org/10.1080/01621459.2021.1874961

Pescott, O. L., Boyd, R. J., Powney, G. D., & Stewart, G. B. (2023). *Towards a unified approach to formal risk of bias assessments for causal and descriptive inference*. ArXiv. https://doi.org/10.48550/arXiv.2308.11458

Reichardt, C. S., & Gollob, H. F. (1987). Taking uncertainty into account when estimating effects. *New Directions for Program Evaluation*, *1987*(35), 7–22. https://doi.org/10.1002/ev.1456

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Nelson Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Pielke Jr, R., Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., … Vineis, P. (2020). Five ways to ensure that models serve society: A manifesto. *Nature*, *582*(7813), 482–484. https://doi.org/10.1038/d41586-020-01812-9

Schouten, B., Bethlehem, J., Beullens, K., Kleven, Ø., Loosveldt, G., Luiten, A., Rutar, K., Shlomo, N., & Skinner, C. (2012). Evaluating, comparing, monitoring, and improving representativeness of survey response through R-indicators and partial R-indicators. *International Statistical Review*, *80*(3), 382–399. https://doi.org/10.1111/j.1751-5823.2012.00189.x

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Stark, P. B. (2022). Pay no attention to the model behind the curtain. *Pure and Applied Geophysics*, *179*(11), 4121–4145. https://doi.org/10.1007/s00024-022-03137-2

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Sterba, S. K., Christ, S. L., Prinstein, M. J., & Nock, M. K. (2011). Beyond treating complex sampling designs as simple random samples: Data analysis and reporting. In A. T. Panter & S. K. Sterba (Eds.), *Handbook of ethics in quantitative methodology* (pp. 267–291). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9780203840023

Swanson, S. A. (2023). The causal effects of causal inference pedagogy. *Epidemiology*, *34*(5), 611–613. https://doi.org/10.1097/EDE.0000000000001639

Thompson, E. (2022). *Escape from model land*. Basic Books UK.

Valliant, R. (2023). Hansen Lecture 2022: The evolution of the use of models in survey sampling. *Journal of Survey Statistics and Methodology*, https://doi.org/10.1093/jssam/smad021

References

- Andridge, R. R., West, B. T., Little, R. J. A., Boonstra, P. S., & Alvarado-Leiton, F. (2019). Indices of Non-Ignorable Selection Bias for Proportions Estimated from Non-Probability Samples. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *68*(5), 1465–1483. https://doi.org/10.1111/rssc.12371 ↩

- Bailey, M. A. (2023). A New Paradigm for Polling. *Harvard Data Science Review*, *5*(3). https://doi.org/10.1162/99608f92.9898eede ↩

- Bailey, M. A. (2023). A New Paradigm for Polling. *Harvard Data Science Review*, *5*(3). https://hdsr.mitpress.mit.edu/pub/ejk5yhgv/release/2 ↩

- Bailey, M. A. (2023). *Polling at a Crossroads: Rethinking Modern Survey Research*. Cambridge University Press. ↩

- Boonstra, P. S., Little, R. J. A., West, B. T., Andridge, R. R., & Alvarado-Leiton, F. (2021). A simulation study of diagnostics for selection bias. *Journal of Official Statistics*, *37*(3), 751–769. https://doi.org/10.2478/jos-2021-0033 ↩

- Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*(356), 791–799. https://doi.org/10.1080/01621459.1976.10480949 ↩

- Boyd, R. J., Powney, G. D., Burns, F., Danet, A., Duchenne, F., Grainger, M. J., Jarvis, S. G., Martin, G., Nilsen, E. B., Porcher, E., Stewart, G. B., Wilson, O. J., & Pescott, O. L. (2022). ROBITT: A tool for assessing the risk-of-bias in studies of temporal trends in ecology. *Methods in Ecology and Evolution*, *13*(7), 1497–1507. https://doi.org/10.1111/2041-210X.13857 ↩

- Boyd, R. J., Stewart, G., & Pescott, O. (2023). Descriptive inference using large, unrepresentative nonprobability samples: An introduction for ecologists. *ecoEvoRxiv*, https://doi.org/10.32942/X2359P. ↩

- Coker, B., Rudin, C., & King, G. (2021). A Theory of Statistical Inference for Ensuring the Robustness of Scientific Results. *Management Science*, *67*(10), 6174–6197. https://doi.org/10.1287/mnsc.2020.3818 ↩

- Draper, D., Hodges, J. S., Mallows, C. L., & Pregibon, D. (1993). Exchangeability and Data Analysis. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *156*(1), 9–37. https://doi.org/10.2307/2982858 ↩

- Freedman, D. A. (2010). *Statistical Models and Causal Inference: A Dialogue with the Social Sciences* (D. Collier, J. S. Sekhon, & P. B. Stark, Eds.). Cambridge University Press. ↩

- Greenland, S. (2021). Invited Commentary: Dealing With the Inevitable Deficiencies of Bias Analysis—and All Analyses. *American Journal of Epidemiology*, *190*(8), 1617–1621. https://doi.org/10.1093/aje/kwab069 ↩

- Greenland, S. (2023). Connecting simple and precise P-values to complex and ambiguous realities (includes rejoinder to comments on "Divergence vs. decision P-values"). *Scandinavian Journal of Statistics*, *50*(3), 899–914. https://doi.org/10.1111/sjos.12645 ↩

- Greenland, S. (2023). Divergence versus decision P-values: A distinction worth making in theory and keeping in practice: Or, how divergence P-values measure evidence even when decision P-values do not. *Scandinavian Journal of Statistics*, *50*(1), 54–88. https://doi.org/10.1111/sjos.12625 ↩

- Hartman, E., & Huang, M. (2023). Sensitivity Analysis for Survey Weights. *Political Analysis*, 1–16. https://doi.org/10.1017/pan.2023.12 ↩

- Hodges, J. S. (1996). Statistical Practice as Argumentation: A Sketch of a Theory of Applied Statistics. In J. C. Lee, W. O. Johnson, & A. Zellner (Eds.), *Modelling and Prediction Honoring Seymour Geisser* (pp. 19–45). Springer. https://doi.org/10.1007/978-1-4612-2414-3_2 ↩

- Lee, K. J., Carlin, J. B., Simpson, J. A., & Moreno-Betancur, M. (2023). Assumptions and analysis planning in studies with missing data in multiple variables: moving beyond the MCAR/MAR/MNAR classification. *International Journal of Epidemiology*, 1268–1275. https://doi.org/10.1093/ije/dyad008 ↩

- Little, R. J. A., & Rubin, D. B. (2020). *Statistical Analysis with Missing Data* (3rd ed.). Wiley. ↩

- Little, R. J. A., West, B. T., Boonstra, P. S., & Hu, J. (2020). Measures of the Degree of Departure from Ignorable Sample Selection. *Journal of Survey Statistics and Methodology*, *8*(5), 932–964. https://doi.org/10.1093/jssam/smz023 ↩

- Liu, Y., Kale, A., Althoff, T., & Heer, J. (2021). Boba: Authoring and Visualizing Multiverse Analyses. *IEEE Transactions on Visualization and Computer Graphics*, *27*(2), 1753–1763. https://doi.org/10.1109/TVCG.2020.3028985 ↩

- Lohr, S. L. (2022). Comments on "Statistical inference with non-probability survey samples". *Survey Methodology*, *48*(2), 331–338. ↩

- Makela, S., Si, Y., & Gelman, A. (2014). Statistical graphics for survey weights. *Revista Colombiana de Estadística*, *37*(2), 285–295. ↩

- Mallows, C. (1998). The Zeroth Problem. *The American Statistician*, *52*(1), 1–9. https://doi.org/10.2307/2685557 ↩

- Manski, C. F. (1990). Nonparametric Bounds on Treatment Effects. *The American Economic Review*, *80*(2), 319–323. ↩

- Meng, X.-L. (2018). Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *The Annals of Applied Statistics*, *12*(2), 685–726. https://doi.org/10.1214/18-AOAS1161SF ↩

- Meng, X.-L. (2022). Comments on "Statistical inference with non-probability survey samples" – Miniaturizing data defect correlation: A versatile strategy for handling non-probability samples. *Survey Methodology*, *48*(2). https://doi.org/10.51387/22-NEJSDS6 ↩

- Meng, X.-L. (2022b). Double your variance, dirtify your Bayes, devour your pufferfish, and draw your kidstrogram. *The New England Journal of Statistics in Data Science*, *1*(1), 4–23. https://doi.org/10.51387/22-NEJSDS6

  ↩

- Mohan, K., & Pearl, J. (2021). Graphical Models for Processing Missing Data. *Journal of the American Statistical Association*, *116*(534), 1023–1037. https://doi.org/10.1080/01621459.2021.1874961 ↩

- Pescott, O. L., Boyd, R. J., Powney, G. D., & Stewart, G. B. (2023). *Towards a unified approach to formal risk of bias assessments for causal and descriptive inference*. https://doi.org/10.48550/arXiv.2308.11458 ↩

- Reichardt, C. S., & Gollob, H. F. (1987). Taking uncertainty into account when estimating effects. *New Directions for Program Evaluation*, *1987*(35), 7–22. https://doi.org/10.1002/ev.1456 ↩

- Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Nelson Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Pielke Jr, R., Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E.,

Sarewitz, D., Stark, P. B., … Vineis, P. (2020). Five ways to ensure that models serve society: a manifesto. *Nature*, *582*(7813), 482–484. https://doi.org/10.1038/d41586-020-01812-9↩

- Schouten, B., Bethlehem, J., Beullens, K., Kleven, Ø., Loosveldt, G., Luiten, A., Rutar, K., Shlomo, N., & Skinner, C. (2012). Evaluating, Comparing, Monitoring, and Improving Representativeness of Survey Response Through R-Indicators and Partial R-Indicators. *International Statistical Review*, *80*(3), 382–399. https://doi.org/10.1111/j.1751-5823.2012.00189.x ↩

- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A Proposed Addition to All Empirical Papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630 ↩

- Stark, P. B. (2022). Pay No Attention to the Model Behind the Curtain. *Pure and Applied Geophysics*, *179*(11), 4121–4145. https://doi.org/10.1007/s00024-022-03137-2 ↩

- Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637 ↩

- Sterba, S. K., Christ, S. L., Prinstein, M. J., & Nock, M. K. (2011). Beyond treating complex sampling designs as simple random samples: Data analysis and reporting. In A. T. Panter & S. K. Sterba (Eds.), *Handbook of ethics in quantitative methodology* (pp. 267–291). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9780203840023 ↩

- Swanson, S. A. (2023). The Causal Effects of Causal Inference Pedagogy. *Epidemiology (Cambridge, Mass.)*, *34*(5), 611–613. https://doi.org/10.1097/EDE.0000000000001639 ↩

- Thompson, E. (2022). *Escape from Model Land*. Basic Books UK. ↩

- Valliant, R. (2023). Hansen Lecture 2022: The Evolution of the Use of Models in Survey Sampling. *Journal of Survey Statistics and Methodology*. https://doi.org/10.1093/jssam/smad021 ↩