Father’s Analysis - Multiverse report

To arrive at our main results, we had to make many decisions that were not fully justified by theoretical or statistical considerations — other competent researchers could have chosen differently and justified their choices equally well. To illustrate the variety of choices we made and examine the robustness of our results to these choices, we conducted a multiverse analysis, reported in this document. A multiverse analysis involves specifying different choice points in the preparation and analysis of data with respect to some question and running and reporting all analyses rather than a single path (Steegen et al., 2016). In this document, we first describe the analysis options that comprise our multiverse analysis, then describe some conclusions based on the analysis, and finally present the multiverse results in several graphs.

# Analysis options

## Model for proportions

Our moderation analysis relies on models which predict a proportion (the mother’s IPV score, which can range from 0 to 24 when considering all items, and which can be turned into a proportion by dividing the score by the maximum possible score). This type of data is often modelled with a binomial GLM, where the outcome is the proportion itself, and a `weights` argument specifies how many “trials” the proportion is out of or, in our case, the maximum possible score.

Another option is to use a “linear probability model”, which just means to model a proportion (which necessarily cannot be smaller than 0 or greater than 1) as if it was normally distributed (which by definition is continuous and encompasses the entire real number line). These models come with advantages (e.g., interpretability of coefficients, ease of computation of mediation effects) and disadvantages (e.g., possibility of predicting data points outside the accepted range). In keeping with common practice for the Future of Families dataset, and because of the additional arbitrariness of binarizing these variables, our main manuscript reports the linear analysis. To complement our linear analysis, we also ran logistic regressions after first binarizing the variables. Binary IPV was 0 if the index was 0 and 1 if the index was above 0, and the same logic was applied to the IPV subscales.

Thus, we used three different models for proportions in this multiverse analysis: “binomial” (corresponding to a weighted logit regression), “continuous” (corresponding to a linear regression), and “binary” (corresponding to a logit regression on the binarized IPV index).

## Missing data

One way to deal with missing data in the Future of Families dataset is by analysing only complete cases. However, this approach can result in a big loss of information and less precise estimates. Thus, we used a dataset where missing values were imputed using the `missForest` algorithm for our main analysis and complemented it with a complete case analysis in the multiverse.

A further option that could be explored would be a more principled missing data analysis, where the imputation method (or lack thereof) is decided based on documented assumptions about the causes of missingness (add reference?).

## Sampling weights

The Future of Families survey has a complex design and is not representative of the US population, which is why the dataset comes with sampling weights which, in principle, should allow us to make the estimates nationally representative. It is clear that weights should be incorporated into the analysis when we are trying to estimate simple quantities (e.g, means or proportions) and use the sample to make inferences about the population. However, the use of survey weights for causal inference problems is not straightforward, and there is no guarantee that using the weights makes the estimate less rather than more biased Solon et al. (2015). Thus, we ran our models with and without using the national weights provided in the Future of Families data. For our main analysis reported in the manuscript, we followed convention when analysing Future of Families data and did not weight our multivariable regressions.

A more laborious but also more principled approach would be to generate graphical causal models that incorporate our assumptions about the causal structure of the variables as well as the sampling design in order to determine how best to account for the sampling design for any given causal inference question (Schuessler & Selb, 2023).

## Covariate adjustment

It is common in the social sciences to statistically adjust for potential confounders when analyzing observational data. However, this likely does more harm than good (Wysocki et al., 2022), as a causal effect can just as easily be confounded as deconfounded by the addition of a covariate. Thus, in addition to running models with covariate adjustment, which are common practice when examining effects of race with the Future of Families data, we examined the effects without covariate adjustment.

## Outcome

Our primary interest was women’s total IPV scores, but we were also interested in exploring three IPV subscales (emotional, controlling, and physical IPV). Thus we ran each analysis specification three additional times, using each IPV subscale as an outcome at a time.

All in all, our “multiverse” consisted of the following “universes”:

| universe | missing | proportion | weights | covariates | total |
| --- | --- | --- | --- | --- | --- |
| 1 | Imputed | Continuous | Weighted | Adjusted | Total |
| 2 | Imputed | Continuous | Weighted | Adjusted | Physical |
| 3 | Imputed | Continuous | Weighted | Adjusted | Emotional |
| 4 | Imputed | Continuous | Weighted | Adjusted | Controlling |
| 5 | Imputed | Continuous | Weighted | Unadjusted | Total |
| 6 | Imputed | Continuous | Weighted | Unadjusted | Physical |
| 7 | Imputed | Continuous | Weighted | Unadjusted | Emotional |
| 8 | Imputed | Continuous | Weighted | Unadjusted | Controlling |
| 9 | Imputed | Continuous | Unweighted | Adjusted | Total |
| 10 | Imputed | Continuous | Unweighted | Adjusted | Physical |
| 11 | Imputed | Continuous | Unweighted | Adjusted | Emotional |
| 12 | Imputed | Continuous | Unweighted | Adjusted | Controlling |
| 13 | Imputed | Continuous | Unweighted | Unadjusted | Total |
| 14 | Imputed | Continuous | Unweighted | Unadjusted | Physical |
| 15 | Imputed | Continuous | Unweighted | Unadjusted | Emotional |
| 16 | Imputed | Continuous | Unweighted | Unadjusted | Controlling |
| 17 | Imputed | Binary | Weighted | Adjusted | Total |
| 18 | Imputed | Binary | Weighted | Adjusted | Physical |
| 19 | Imputed | Binary | Weighted | Adjusted | Emotional |
| 20 | Imputed | Binary | Weighted | Adjusted | Controlling |
| 21 | Imputed | Binary | Weighted | Unadjusted | Total |
| 22 | Imputed | Binary | Weighted | Unadjusted | Physical |
| 23 | Imputed | Binary | Weighted | Unadjusted | Emotional |
| 24 | Imputed | Binary | Weighted | Unadjusted | Controlling |
| 25 | Imputed | Binary | Unweighted | Adjusted | Total |
| 26 | Imputed | Binary | Unweighted | Adjusted | Physical |
| 27 | Imputed | Binary | Unweighted | Adjusted | Emotional |
| 28 | Imputed | Binary | Unweighted | Adjusted | Controlling |
| 29 | Imputed | Binary | Unweighted | Unadjusted | Total |
| 30 | Imputed | Binary | Unweighted | Unadjusted | Physical |
| 31 | Imputed | Binary | Unweighted | Unadjusted | Emotional |
| 32 | Imputed | Binary | Unweighted | Unadjusted | Controlling |
| 33 | Imputed | Binomial | Weighted | Adjusted | Total |
| 34 | Imputed | Binomial | Weighted | Adjusted | Physical |
| 35 | Imputed | Binomial | Weighted | Adjusted | Emotional |
| 36 | Imputed | Binomial | Weighted | Adjusted | Controlling |
| 37 | Imputed | Binomial | Weighted | Unadjusted | Total |
| 38 | Imputed | Binomial | Weighted | Unadjusted | Physical |
| 39 | Imputed | Binomial | Weighted | Unadjusted | Emotional |
| 40 | Imputed | Binomial | Weighted | Unadjusted | Controlling |
| 41 | Imputed | Binomial | Unweighted | Adjusted | Total |
| 42 | Imputed | Binomial | Unweighted | Adjusted | Physical |
| 43 | Imputed | Binomial | Unweighted | Adjusted | Emotional |
| 44 | Imputed | Binomial | Unweighted | Adjusted | Controlling |
| 45 | Imputed | Binomial | Unweighted | Unadjusted | Total |
| 46 | Imputed | Binomial | Unweighted | Unadjusted | Physical |
| 47 | Imputed | Binomial | Unweighted | Unadjusted | Emotional |
| 48 | Imputed | Binomial | Unweighted | Unadjusted | Controlling |
| 49 | Complete cases | Continuous | Weighted | Adjusted | Total |
| 50 | Complete cases | Continuous | Weighted | Adjusted | Physical |
| 51 | Complete cases | Continuous | Weighted | Adjusted | Emotional |
| 52 | Complete cases | Continuous | Weighted | Adjusted | Controlling |
| 53 | Complete cases | Continuous | Weighted | Unadjusted | Total |
| 54 | Complete cases | Continuous | Weighted | Unadjusted | Physical |
| 55 | Complete cases | Continuous | Weighted | Unadjusted | Emotional |
| 56 | Complete cases | Continuous | Weighted | Unadjusted | Controlling |
| 57 | Complete cases | Continuous | Unweighted | Adjusted | Total |
| 58 | Complete cases | Continuous | Unweighted | Adjusted | Physical |
| 59 | Complete cases | Continuous | Unweighted | Adjusted | Emotional |
| 60 | Complete cases | Continuous | Unweighted | Adjusted | Controlling |
| 61 | Complete cases | Continuous | Unweighted | Unadjusted | Total |
| 62 | Complete cases | Continuous | Unweighted | Unadjusted | Physical |
| 63 | Complete cases | Continuous | Unweighted | Unadjusted | Emotional |
| 64 | Complete cases | Continuous | Unweighted | Unadjusted | Controlling |
| 65 | Complete cases | Binary | Weighted | Adjusted | Total |
| 66 | Complete cases | Binary | Weighted | Adjusted | Physical |
| 67 | Complete cases | Binary | Weighted | Adjusted | Emotional |
| 68 | Complete cases | Binary | Weighted | Adjusted | Controlling |
| 69 | Complete cases | Binary | Weighted | Unadjusted | Total |
| 70 | Complete cases | Binary | Weighted | Unadjusted | Physical |
| 71 | Complete cases | Binary | Weighted | Unadjusted | Emotional |
| 72 | Complete cases | Binary | Weighted | Unadjusted | Controlling |
| 73 | Complete cases | Binary | Unweighted | Adjusted | Total |
| 74 | Complete cases | Binary | Unweighted | Adjusted | Physical |
| 75 | Complete cases | Binary | Unweighted | Adjusted | Emotional |
| 76 | Complete cases | Binary | Unweighted | Adjusted | Controlling |
| 77 | Complete cases | Binary | Unweighted | Unadjusted | Total |
| 78 | Complete cases | Binary | Unweighted | Unadjusted | Physical |
| 79 | Complete cases | Binary | Unweighted | Unadjusted | Emotional |
| 80 | Complete cases | Binary | Unweighted | Unadjusted | Controlling |
| 81 | Complete cases | Binomial | Weighted | Adjusted | Total |
| 82 | Complete cases | Binomial | Weighted | Adjusted | Physical |
| 83 | Complete cases | Binomial | Weighted | Adjusted | Emotional |
| 84 | Complete cases | Binomial | Weighted | Adjusted | Controlling |
| 85 | Complete cases | Binomial | Weighted | Unadjusted | Total |
| 86 | Complete cases | Binomial | Weighted | Unadjusted | Physical |
| 87 | Complete cases | Binomial | Weighted | Unadjusted | Emotional |
| 88 | Complete cases | Binomial | Weighted | Unadjusted | Controlling |
| 89 | Complete cases | Binomial | Unweighted | Adjusted | Total |
| 90 | Complete cases | Binomial | Unweighted | Adjusted | Physical |
| 91 | Complete cases | Binomial | Unweighted | Adjusted | Emotional |
| 92 | Complete cases | Binomial | Unweighted | Adjusted | Controlling |
| 93 | Complete cases | Binomial | Unweighted | Unadjusted | Total |
| 94 | Complete cases | Binomial | Unweighted | Unadjusted | Physical |
| 95 | Complete cases | Binomial | Unweighted | Unadjusted | Emotional |
| 96 | Complete cases | Binomial | Unweighted | Unadjusted | Controlling |

# Multiverse results - Summary

The multiverse analysis reveals that our results are somewhat sensitive to different specifications. In the following sections, we present and describe some figures and tables that highlight this sensitivity.

Given the discussion above about the legitimacy of each analytic choice, we don’t think it sensible to summarize across specifications to arrive at an “overall” interaction effect of between race and employment. For example, summarizing the models which control for all covariates along with the models that don’t control for covariates does not make sense, as covariate adjustment changes the meaning of the estimate of the effect(s) of race on IPV, so there is no meaningful way they could be averaged. Instead, we consider this process as a pedagogical tool. As Julia Rohrer put it on the 100% CI blog (Rohrer, 2021):

In the end, the multiverse may prove to be most powerful as a pedagogical tool. […] Multiverse analysis can raise awareness of how data analytic flexibility can be exploited. It can also alert us to gaps in our knowledge. Gaps in our knowledge about the underlying causal web: Is including that covariate arbitrary or not? What does it mean if results hinge on it? Gaps in our knowledge about statistics: Can we expect these types of models to return the same answer? Under which conditions would they diverge? Gaps in our knowledge about measurement and conceptualization: Does it make sense to expect the same result for these different operationalizations of the outcome? What does it mean if results vary? We have now closed the loop and are [back to the original multiverse paper](https://journals.sagepub.com/doi/10.1177/1745691616658637), in which the authors write: “The real conclusion of the multiverse analysis is that there is a gaping hole in theory or in measurement.”

# Multiverse results - Visualizations

First, we visualize the full multiverse for each outcome (total, emotional, controlling, and physical IPV), looking at each effect in turn (marginal effect of unemployment for White men, marginal effect of unemployment for Black men, interaction effect), ordering by the estimate magnitude.

Each of the following three graphs has the same structure. The top panel shows the effect size (which can be interpreted as percentage point differences) with confidence interval for each specification, and is colored according to significance. The bottom panel shows which combination of analytical choices produced the corresponding effect size above.

First, we plot the effect of unemployment on White IPV perpetration (Figure 1), then the effect of unemployment on Black IPV perpetration (Figure 2), and finally we subtract the second from the first to show the interaction (Figure 3).

## White (Employed - Unemployed)

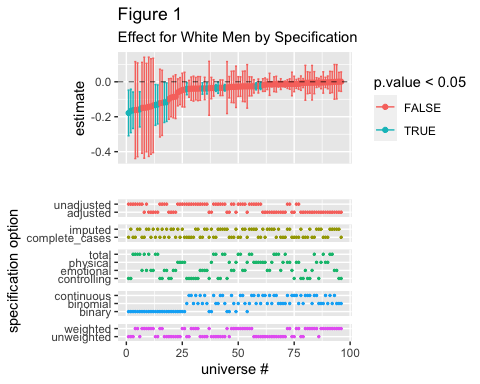


Figure 1 shows that, across specifications, among White men unemployment is associated with their partner experiencing more IPV. However, most estimates are very close to zero or not statistically significant. We find the largest effect magnitudes for specifications with a binary IPV outcome, as well as with models that include covariate adjustment. Further, unweighted models seem to result in narrower confidence intervals and thus also more estimates that are significantly different from zero. Results don’t seem to vary much whether imputed data or complete cases are used, nor based on whether the outcome is total IPV or one of the subscales.

Specifications for which there is a significant effect:

| missing | proportion | weights | outcome | covariates | estimate | *p* |
| --- | --- | --- | --- | --- | --- | --- |
| complete\_cases | binary | unweighted | controlling | unadjusted | -0.18 | .008\*\* |
| imputed | binary | unweighted | controlling | unadjusted | -0.17 | .009\*\* |
| complete\_cases | binary | unweighted | total | unadjusted | -0.16 | .002\*\* |
| imputed | binary | unweighted | total | unadjusted | -0.16 | .002\*\* |
| complete\_cases | binary | unweighted | total | adjusted | -0.13 | .026\* |
| imputed | binary | unweighted | total | adjusted | -0.13 | .023\* |
| complete\_cases | binary | unweighted | emotional | unadjusted | -0.12 | .045\* |
| imputed | binary | unweighted | emotional | unadjusted | -0.12 | .038\* |
| complete\_cases | binomial | unweighted | controlling | unadjusted | -0.04 | .021\* |
| complete\_cases | binomial | unweighted | emotional | unadjusted | -0.04 | .024\* |
| imputed | binomial | unweighted | emotional | unadjusted | -0.04 | .019\* |
| complete\_cases | binomial | unweighted | total | unadjusted | -0.04 | .001\*\*\* |
| imputed | binomial | unweighted | controlling | unadjusted | -0.03 | .036\* |
| imputed | binomial | unweighted | total | unadjusted | -0.03 | < .001\*\*\* |
| imputed | binomial | unweighted | physical | unadjusted | -0.03 | .012\* |
| complete\_cases | binomial | unweighted | physical | unadjusted | -0.02 | .021\* |
| *Note.* \* p < .05, \*\* p < .01, \*\*\* p < .001 | | | | | | |

Only models without sampling weights returned a significant result. Of note, the great majority of models with significant results did not adjust for covariates, thereby not reducing the effect of race through the inclusion of a post-treatment variable (which could be a mediator). Further, only binary or multiple-trial binomial models returned significant results.

## Black (Employed - Unemployed)

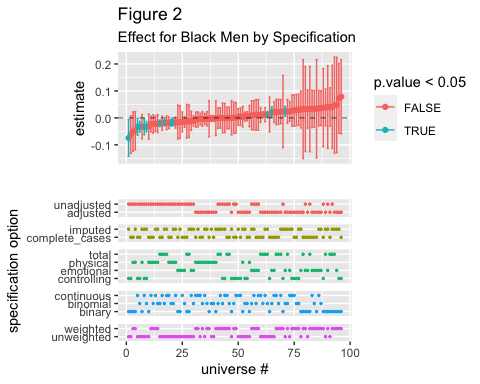


Figure 2 shows more variability regarding the direction of the impact of unemployment on IPV perpetrated by Black men: while most specifications suggest that unemployment is associated with lower IPV, almost half of the estimated effects and all of the significant effects are in the opposite direction. Covariate adjusted and survey-weighted models seem to result in more “positive” estimates (meaning that employment is associated with higher IPV), while the converse is true for unadjusted and unweighted models.

Specifications for which there is a significant effect:

| missing | proportion | weights | outcome | covariates | estimate | *p* |
| --- | --- | --- | --- | --- | --- | --- |
| imputed | binary | unweighted | controlling | unadjusted | -0.07 | .032\* |
| complete\_cases | continuous | unweighted | controlling | unadjusted | -0.03 | .043\* |
| complete\_cases | binomial | unweighted | controlling | unadjusted | -0.03 | .001\*\*\* |
| imputed | continuous | unweighted | controlling | unadjusted | -0.03 | .037\* |
| imputed | binomial | unweighted | controlling | unadjusted | -0.03 | .001\*\*\* |
| complete\_cases | binomial | unweighted | total | unadjusted | -0.02 | .001\*\*\* |
| imputed | binomial | unweighted | total | unadjusted | -0.02 | .002\*\* |
| complete\_cases | binomial | unweighted | physical | unadjusted | -0.02 | .004\*\* |
| imputed | binomial | unweighted | physical | unadjusted | -0.02 | .004\*\* |
| imputed | binomial | unweighted | total | adjusted | 0.01 | .031\* |
| complete\_cases | binomial | unweighted | emotional | adjusted | 0.02 | .031\* |
| imputed | binomial | unweighted | emotional | adjusted | 0.02 | .013\* |
| *Note.* \* p < .05, \*\* p < .01, \*\*\* p < .001 | | | | | | |

Here we once again only find an effect for unweighted models, and we can see that adjusting for covariates leads to significant effects in the opposite direction compared to unadjusted models.

## Interaction (White (Emp - Unemp) - Black (Emp - Unemp))

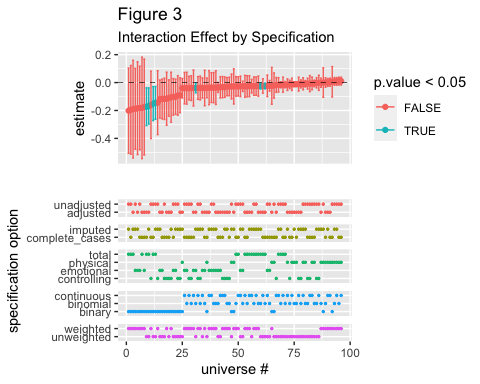


Figure 3 shows that a most specifications result in a “negative”, albeit not significant, interaction effect, such that the difference in IPV between employed and unemployed men is more negative for White men as compared to Black men. Here again we see that models using binarized IPV show the starkest negative interaction effect. The other specification options don’t seem to have a systematic effect on the estimated interaction.

Specifications for which there is a significant effect:

| missing | proportion | weights | outcome | covariates | estimate | *p* |
| --- | --- | --- | --- | --- | --- | --- |
| complete\_cases | binary | unweighted | total | adjusted | -0.17 | .012\* |
| imputed | binary | unweighted | total | adjusted | -0.17 | .011\* |
| complete\_cases | binary | unweighted | total | unadjusted | -0.15 | .015\* |
| imputed | binary | unweighted | total | unadjusted | -0.15 | .015\* |
| imputed | binomial | unweighted | emotional | adjusted | -0.04 | .029\* |
| imputed | binomial | unweighted | total | adjusted | -0.03 | .013\* |
| complete\_cases | binomial | unweighted | total | adjusted | -0.02 | .022\* |
| *Note.* \* p < .05, \*\* p < .01, \*\*\* p < .001 | | | | | | |

Again, only unweighted models return significant results. Further, we mostly find interactions when looking at total IPV and not controlling, emotional (except for in one specification), or physical IPV.

Next we home in on each outcome in turn (total, emotional, controlling, and physical IPV), without arranging by size of estimate. In each figure, the top plot shows the three effects (Top: effect of unemployment among Black men, Middle: effect of unemployment among White men, Bottom: interaction effect) while the bottom plot describes the corresponding specifications.

## Total IPV

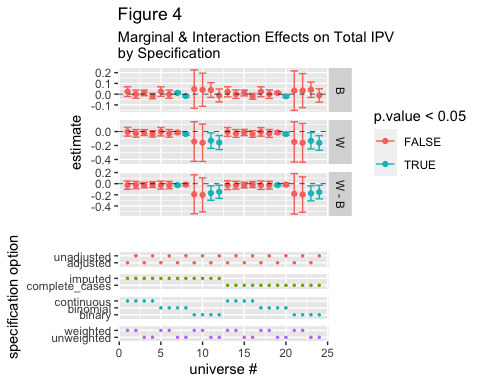
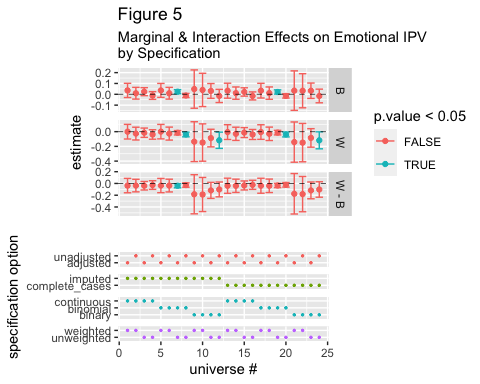


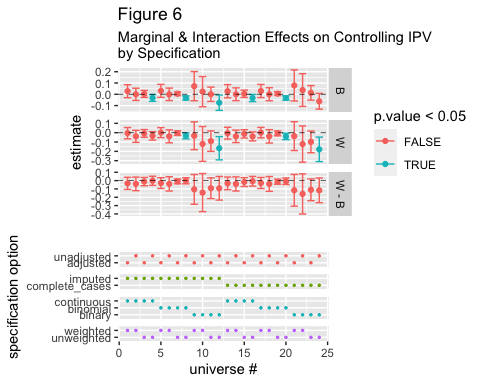
Figure 4 makes clear that when our outcome is a measure of total IPV, results vary most depending on the model specification (linear, binomial, or binary outcome), with the only significant results coming from binary or binomial unweighted models. Across specifications we see positive effects of unemployment (meaning, lower IPV perpetration) for Black fathers and negative effects for White fathers, though the effect sizes are very small.

## Emotional IPV



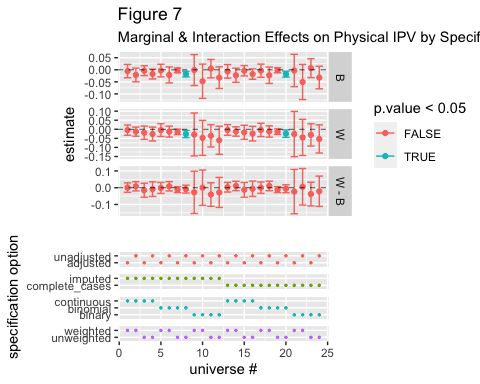
When we look at emotional IPV, the multiverse paints a picture of a null effect - all specifications result in estimates close to zero, the majority of which are not significant. Again, weighted binary models show the widest confidence intervals and unweigthed binomial models the narrowest.

## Controlling IPV



Controlling IPV also tells a story of no interaction, and the effects for Black and White parents also seem to fall on either side of the 0 line.

## Physical IPV



Looking just at physical IPV, the interaction becomes even less pronounced, with estimates for the difference between the effect of unemployment on Black and White fathers’ IPV perpetration hovering around 0. The marginal effects themselves are almost all small and nonsignificant, indicating that the true effects could be zero or lie on either side of zero.

# References

Rohrer, J. (2021). *Mülltiverse analysis*. <https://www.the100.ci/2021/03/07/mulltiverse-analysis/>

Schuessler, J., & Selb, P. (2023). Graphical causal models for survey inference. *Sociological Methods & Research*, 004912412311768. <https://doi.org/10.1177/00491241231176851>

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human Resources*, *50*(2), 301–316. <https://doi.org/10.3368/jhr.50.2.301>

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>

Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical control requires causal justification. *Advances in Methods and Practices in Psychological Science*, *5*(2), 1–19. <https://doi.org/10.1177/25152459221095823>