

Using Linked Parent–Child Data and Causal Mediation Analysis to Study Parental Influence on Turnout

Contributors: Elisabeth Gidengil, Hanna Wass & Maria Valaste

Academic Level: Postgraduate Publishing Company: SAGE Publications Ltd

City: London Online ISBN: 9781526464576

DOI: <http://dx.doi.org/10.4135/9781526464576>

© 2019 SAGE Publications Ltd All Rights Reserved.

Abstract

This case traces the steps that we took to bring our study from initial conception to publication. The study uses a unique dataset that links the official voting records of 18- to 30-year-olds and their parents to data about their social background characteristics. We explain how we went from testing a loosely connected set of expectations based on the political socialization literature to framing our study as a test of two major theories about the parent–child link in voting. This process necessitated a thorough re-thinking of our research design. We describe how we progressed from estimating straightforward logistic regression models to performing a causal mediation analysis. We emphasize the importance of understanding the assumptions that underlie causal mediation analysis and of incorporating robustness checks. Finally, our case illustrates how the peer-review process stimulated us to develop a stronger theoretical framework and a more sophisticated research design.

Learning Outcomes

By the end of this case, students should be able to

- Understand how to design a study using causal mediation analysis
- Evaluate critically studies that employ a similar design
- Understand how to respond constructively to feedback from reviewers

Project Overview and Context

Successful projects typically begin with a set of theoretical expectations that we want to evaluate. Then, we track down – or generate – suitable data to conduct a rigorous test. This project, though, started with a unique set of data, along with a shared interest in explaining why people do or do not exercise the most fundamental of democratic rights, the right to vote in free and fair elections. The dataset offered an unparalleled opportunity to explore the role of childhood socialization.

It linked socio-demographic data compiled by Statistics Finland to the official voting records of 18- to 30-year-olds and their parents in the 1999 Finnish parliamentary elections. These data enabled us to overcome one of the main weaknesses of much of the literature on voter turnout: the reliance on self-reports. Using actual voting records meant that our data were free of the biases that plague self-reported turnout, such as over-reporting due to social desirability and misreporting due to flawed recall. Flawed recall is even more of a problem when studying intergenerational transmission as young adults may not be able to report reliably on their parents' turnout. At the same time, the use of

administrative data meant that we had accurate information on the social backgrounds of both the young adults and their parents.

We also had a large enough sample that we could look at how the parent–child link in voting varied depending on factors such as the sex of the young adults and their parents, whether the young adults resided with their parent(s) or had left the nest, and the young adults' age. We began with a 10% random sample of all 18- to 30-year-olds. Statistics Finland used personal identification codes to match the young adults to their parents. This could be done with a very high degree of accuracy: The linkage rate was more than 99.5%. Personal identification codes were also used to link voting records with the administrative data. Our initial sample consisted of 83,727 young adults, but we ended up with an effective sample size of 67,465. The loss of cases reflected the impossibility of parent–child linking when the parent was deceased, unknown, not a Finnish citizen, or living in one of the few municipalities where voting records were unavailable. In other respects, though, we could assume that our sample was nationally representative. Importantly, we did not have to be concerned about the biasing effects of nonresponse and, with such a large sample, we could be confident in the observed effects.

Ideally, we would have had data on parental turnout during the children's formative years but those voting records are simply not accessible. Fortunately, we were able to draw on research showing that voting is a habit (e.g., Aldrich, Montgomery, & Wood, 2011). As a result, we could plausibly argue that many parents who voted in 1999 likely voted in earlier elections, too. Nonetheless, we had to acknowledge that the lack of past information on parental voting was a limitation of our study.

Research Design

The Initial Research Design

Our first task was to come up with some expectations about the parent–child link in voting. First and most obviously, we hypothesized that adult children are more likely to vote if their parent(s) vote. The rationale drew on social learning theory, which emphasizes the importance of observational learning and parents as role models (Bandura, 1977). As social learning is enhanced when the child is similar to the parent, we expected that the link would be stronger when the parent and the adult child were the same sex. We also predicted that the link would be stronger if both parents voted, given the importance of consistent cues to social learning. Drawing on prior work on political socialization, we expected that having a mother who votes would matter more than having a father who votes and that the influence of parental turnout would be weaker if the adult child no longer lived at home. We also expected it to become weaker as the adult child grew older. A final hypothesis focused on whether the strength of the parent–child link in voting varied depending on the parents' level of education.

The research design was rather simple. Given the dichotomous nature of the dependent variable (i.e., whether or not the young adult voted), we tested the hypotheses using logistic regression. To allow for the possibility that other factors might explain the parent–child link, controls were added to take account of social background characteristics that could confound the hypothesized relationships. When testing for sex-differentiated and age-differentiated effects, as well as the effects of leaving the nest, we simply added interaction terms to the model. In some cases, more than one adult child of the

same parent(s) was included in the dataset. This necessitated the use of robust standard errors, clustered by family identification number, to take account of intra-family correlation.

Coding the Variables

The key independent variables were parental voting, parental education, and the young adults' education. Parental voting was represented by three dummy variables: both parents voted, only the father voted, and only the mother voted. When variables are dummy coded in this way, there must always be an omitted category, which serves as the reference category for interpreting their effects. Neither parent voted served as our reference category. As the effects of education are not necessarily linear and there is no consensus in the literature as to which level of education matters for turnout, we also created three dummy variables to represent education. These variables corresponded to completing high school, an associate degree, and a university degree or equivalent. Our reference category was completing junior high school, which is the minimum level of education in Finland.

Estimating Marginal Effects

Given the size of the sample, it was likely that even small effects would achieve conventional levels of statistical significance. To ensure that the observed effects were not only statistically significant but also substantively important, we presented the results in the form of marginal effects. These were estimated based on the regression models, using the observed values method (Hanmer & Kalkan, 2013). Unlike logistic regression coefficients (which take the form of log-odds ratios), marginal effects also have the advantage of being easy to interpret. In our study, they represented the predicted change in the probability of voting when, for example, both parents voted as opposed to neither parent voted, keeping other covariates such as sex, age, and education fixed at their observed values.

Robustness Checks

The large sample size meant that we were able to conduct robustness checks by repeating the analyses on random split halves to see whether the results were replicated. A second robustness check was designed to enhance confidence that any observed parent–child link in voting was causal as opposed to being merely correlational. It involved a very different analytical approach that offers a more powerful way of achieving control than simply including potential confounders in a regression model (Blackwell, Iacus, King, & Porro, 2009). Matching mimics the logic of controlled experiments. It does so by comparing treatment and control groups that differ on the independent variable but have the same values on possible confounding variables. Given that the cases in both groups have the same values on those variables, they cannot account for any observed difference on the dependent variable. We conducted the matching analyses for each pair of values on the independent variable: both parents voted versus neither parent voted, both parents voted versus only the father voted, both parents voted versus only the mother voted, neither parent voted versus only the father voted, and only the father voted versus only the mother voted. We used the coarsened exact matching (CEM) algorithm as implemented in Stata. CEM is computationally efficient and fast, even for large datasets. It is also easy to use and has some attractive theoretical properties (Iacus, King, & Porro, 2012).

Matching, however, has its limitations. First and most obviously, the quality of the matching is constrained by the availability of data, leaving open the possibility that important potential

confounders have been omitted. Moreover, depending upon the comparison, there can be many unmatched cases. This raises issues for drawing inferences because the cases included in the matching will vary across the comparisons. For example, the group yielding matches for both parents voted differed depending on whether the cases were being matched with cases where neither parent voted or with cases where only the father voted. It is easy to understand why, given that education is a powerful predictor of turnout: There were few parents who voted but had only a junior high school education, just as there were relatively few parents who had a university degree but did not cast a ballot.

Back to the Drawing Board

We were very happy with the results. Our hypotheses were confirmed and our robustness checks increased our confidence in our findings. We optimistically submitted the resulting manuscript to a highly regarded political science journal that is aimed at a general audience. However, it turned out that we had been a little too optimistic. We were invited to undertake a major revision and to re-submit the manuscript. The reviewers saw our unique data, the robust results, and the ability of the findings to speak to the larger political socialization literature as key strengths of our study. However, they were unanimous in believing that the study needed a stronger theoretical framing to warrant publication in a high-level journal. On reflection, we had to agree with this assessment.

To this end, one of the reviewers made what turned out to be an extremely helpful suggestion. The advice was to use our data to adjudicate between competing theories. On one hand, the parent–child link might be explained by the intergenerational transmission of socioeconomic status. On the other hand, it might be better explained as a result of social learning. We realized that we were very well placed to test status transmission theory (Verba, Burns, & Schlozman, 2003). Proponents of this theory argue that social learning is of secondary importance when it comes to political socialization. They attribute the parent–child link primarily to parental education: Well-educated parents tend to have well-educated children, and people who are well educated tend to vote in higher numbers. It turned out that our Finnish data were very well suited to testing the theory. The theory was developed in the United States but the United States has an atypically high rate of intergenerational transmission of educational attainment for a post-industrial society. Finland, we discovered, was much more typical in this regard. As a result, we could assess whether the theory traveled beyond the American context. We expected that social learning theory would outperform status transmission theory in a setting where parental education had less of an effect on the educational attainment of their children.

Transforming our study into an evaluation of competing theories required a thoroughgoing revision. First, we had to dig deeper into social learning theory. This enabled us to reframe our original hypotheses as a series of logical implications of social learning theory. We also had to specify a testable hypothesis based on status transmission theory. Our new hypothesis predicted that the impact of parental education would be mediated by both parental turnout and the adult child’s education, but the child’s education would be the more important mediator.

A Modified Research Design

In order to test status transmission theory, we needed to assess the relative importance of the two paths through which parental education could influence the adult child’s propensity to vote, namely,

parental turnout and the adult child's education. Our initial approach was simple. We estimated a series of logistic regression models. The first model served as a baseline. It estimated the impact of parental education on the young adult's turnout controlling only for the adult child's sex, age, mother tongue, marital status, housing tenure, and residential mobility. The next model added parental voting while the third model substituted the young adult's education level for parental turnout. The final model included both mediators. Comparing the estimated marginal effects of parental education across the four models offered a way of assessing how much of the effect of parental education was mediated through parental voting and the young adult's own level of educational attainment, respectively.

Realizing that this was a rather crude approach, we also performed a causal mediation analysis as an additional robustness check. However, the method we chose (Imai, Keele, Tingley, & Yamamoto, 2011) was suboptimal for our purposes as it could not handle a mediator that consisted of unordered categories (we could not assume that having a mother who votes ranks higher than having a father who votes). This obliged us to dichotomize parental voting (neither parent voted vs. one or both parents voted). We were also unable to undertake a sensitivity analysis to assess how robust the estimates were to violation of the assumption that there were no unobserved confounders. Accordingly, we downplayed the results of this robustness check in our revised manuscript.

The results confirmed our expectation that the young adult's educational attainment was of secondary importance as a mediator of the impact of parental education. What mattered more was parental turnout. This was the case whether we looked at the regression results or the results of the causal mediation analysis. We thoroughly revised the manuscript and re-submitted it.

The Final Research Design

The reviewers raised concerns about our approach to assessing the mediating effects of parental turnout and the young adults' own educational attainment on their propensity to vote. One of the reviewers suggested that we make more of the causal mediation analysis while another suggested that structural equation modeling would be a better way of testing our theoretical expectations. The latter suggestion posed a problem. Our dependent variable—whether a young adult voted or not—was dichotomous. As authors like Imai et al. (2011) and Breen, Karlson, and Holm (2013) have explained, structural equation modeling can only appropriately be used to estimate causal paths when the dependent variable is continuous. If it is used to estimate a nonlinear model, it will not provide consistent estimates of the average causal mediation effect.

Once we realized that we could not follow this reviewer's advice, we scoured the literature to find an approach to causal mediation analysis that would be more appropriate to the unordered categories of our parental voting variable. We discovered the khb method. Developed by Breen et al. (2013), this method can appropriately be employed to decompose total effects into their direct and indirect components when the dependent variable is binary. Importantly, it was compatible with our hypothesized mediators.

Like any causal mediation analysis, this method requires a strong assumption, known as sequential ignorability. In fact, the same is true of structural equation modeling, though this is often overlooked. To satisfy this assumption, we had to assume that there was no confounding between the treatment

variable (parental education) and the mediating (parental turnout and the adult child's education) and outcome (turnout) variables. We also had to assume that there was no confounding between the mediating variables and the outcome variable. To satisfy the sequential ignorability assumption, we included controls for three potential confounders: the adult child's sex, age, and mother tongue. A final assumption was easier to satisfy. We could be confident that neither mediator confounded the relationship between the other mediator and the outcome as parental turnout could not influence a child's educational attainment and vice versa. It is important to realize, though, that no causal mediation analysis can definitively exclude the possibility that there are unobserved confounders. In our study, for example, shared genetic inheritance is a potential confounder. Children and their biological parents share about 50% of their genes, raising the possibility that genetic inheritance could explain the parent-child link in turnout. However, this would not explain why mothers have more influence than fathers on the turnout of their adult children. Moreover, attempts to identify a genetic basis for voting have yielded very mixed results.

Method in Action

The khb method enabled us to make a direct comparison of the two potential mediators. Contrary to status transmission theory, the results indicated that the child's educational attainment was of secondary importance. Parental voting proved to be the more important mediator. In fact, according to our estimates, parental voting mediated more than twice as much of the effect of parental education as the adult child's own educational attainment. In other words, status transmission theory could only explain a relatively small part of the connection between parental education and the turnout of their adult children. This confirmed our initial findings. As explained above, the latter had been based on a simple method that involved substituting first one and then the other mediator in a logistic regression model and comparing the marginal effects of parental education with a baseline model to see which mediator resulted in a bigger decrease in the estimated marginal effects of parental education. The khb causal mediation analysis enabled us to give our findings a much stronger basis. Not only had we been able to show that status transmission offered only a very partial explanation of the parent-child link in turnout but finding that parental turnout was the more important mediator lent additional weight to a social learning explanation. Moreover, we could be more confident that there was a causal connection between parental turnout and the turnout of their adult children.

Conclusion

Political scientists are taking issues of causality increasingly seriously. This project gave us the opportunity to learn more about the assumptions that are required to give mediation effects a causal interpretation. Sequential ignorability is a strong assumption that will rarely be fully satisfied. Accordingly, it is incumbent on users of causal mediation analysis to do their best to approximate the assumption while alerting their readers to the limits of their causal inferences.

Finally, bringing this project to fruition was an object lesson in the value of the peer-review process. The path to publication was not an easy one and took a good deal longer than we had anticipated but taking the reviewers' comments seriously really paid off. At the end of the day, we were able to make a much stronger contribution than would have been possible with our initial manuscript.

Exercises and Discussion Questions

1. What is the difference between mediating factors and confounding factors?
2. What are some potential confounders that would not be available from administrative sources like Statistics Finland?
3. Why do methods that mimic controlled experiments provide greater confidence in causal inferences?
4. What are some other methods that could be used to enhance confidence in causal inferences?

References

- Aldrich, J. H., Montgomery, J. M., & Wood, W. (2011). Turnout as a habit. *Political Behavior*, 33, 535–563.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. *Stata Journal*, 9, 524–546.
- Breen, R., Karlson, K. B., & Holm, A. (2013). Total, direct, and indirect effects in logit and probit models. *Sociological Methods & Research*, 42, 164–191.
- Hanmer, M. J., & Kalkan, K. O. (2013). Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models. *American Journal of Political Science*, 57, 263–277.
- Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20, 1–24.
- Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, 105, 765–789.
- Verba, S., Burns, N., & Schlozman, K. (2003). Unequal at the starting line: Creating participatory inequalities across generations and among groups. *American Sociologist*, 34, 45–69.