

Abstract Title Page

Title: Principal stratification as a framework for investigating mediational processes in experimental settings

Authors and Affiliations: Lindsay Page, Center for Education Policy Research, Harvard University

Abstract Body

Background / Context: Often, in the case of experimental evaluations of multifaceted interventions, researchers and policymakers alike are interested in asking not only whether a given intervention had an effect but also why. They seek answers to questions such as: What features of the intervention led to the impacts that we have observed, or what was the causal mechanism or pathway through which random assignment to the intervention resulted in an improved outcome or set of outcomes? A variable on a causal pathway between an initial condition (such as treatment assignment) and an ultimate outcome is referred to typically as a “mediator.” Conceptually, one regards the action of the initial treatment on the ultimate outcome as acting through the hypothesized mediator. Analytic strategies for assessing the extent to which the impact of an initial condition acts on the outcome in this way can be classified, broadly speaking, as approaches to mediation analysis. Quantitative methods for modeling mediational processes are an active area of exploration in the recent methodological literature (see, for example, Rubin, 2004; Bloom, 2006; Gallop et al., 2009; VanderWeele & Vansteelandt, 2009; Bullock, Green & Ha, 2010; Imai, Keele & Tingley, 2010). This paper contributes to the literature by illustrating an approach that capitalizes on the framework of principal stratification (Frangakis & Rubin, 2002) to address mediational questions in the context of an experiment. This use of principal stratification was proposed first by Rubin (2004) and has been applied and discussed recently by Gallop and colleagues (2009).

Purpose / Objective / Research Question / Focus of Study: Other methods proposed for exploring mediational hypotheses include multiple regression analysis (e.g., Baron and Kenny, 1986) and instrumental-variables (IV) estimation (e.g., Gennetian and colleagues, 2006). Despite their popularity, regression- and IV-based approaches rely on sets of untestable assumptions that may be too strong to yield valid inferences about the causal mechanism or mechanisms under investigation in most practical empirical settings. By highlighting the assumptions that must hold for one to achieve valid causal inferences using either of these two approaches, I motivate the particular benefits of investigating mediational processes using principal stratification, with which one is able to relax and test the assumptions underlying these other analytic approaches. Of course, no approach is without its limits. Therefore, I aim to highlight the requirements and limitations associated with this use of principal stratification.

Setting: To illustrate this use of principal stratification, I employ data from MDRC’s experimental evaluation of Career Academies, a secondary-education model in which curricula and student opportunities are organized explicitly around career themes. The intent-to-treat (ITT) evaluation revealed that the randomized offer of academy enrollment led to improved monthly earnings for males in the years after high school, while standard pathways to labor-market success (such as academic achievement, high school graduation and post-secondary attainment) remained unaffected. Therefore, the mechanism through which the academies led to improvements in later earnings is not well understood. I investigate the hypothesis that the career academies succeeded by providing students with greater exposure to the world-of-work and that this increased exposure translated to increased earnings in the years after high school.

Population / Participants / Subjects: MDRC conducted this evaluation of career academies in nine host high schools throughout the United States beginning in 1993. Several characteristics of

the participating sites are noteworthy. For example, the schools are all located in, or near, urban areas serving high percentages of African American and Hispanic students and where rates of high-school dropout, unemployment, and poverty are higher than the national average (Kemple, 2008). Across sites, 55 percent of interested students were assigned randomly the offer to attend their school's academy, while the remainder were offered the host school's regular program only. Among the nine sites, career-academy programs differed in length. Two spanned grades nine through twelve, and seven were intended for students in grades ten through twelve. In accordance with other aspects of this project, I focus on the subset of three-year academies, yielding an effective sample size of 1,306. Nevertheless, this paper's focus is primarily methodological, with the overarching purpose of illustrating the principal stratification approach in contrast to other strategies. Therefore, the analyses that I present are conducted on a subset of 405 male participants from the career-academy study for whom data were essentially complete.

Intervention / Program / Practice: The career-academy model has several important educational features. First, students are organized into small learning communities, distinct entities defined by structural and staff divisions within larger comprehensive high schools. The intention of these smaller units is to provide students with more individualized opportunities to learn and to foster more positive and meaningful relationships among teachers, staff and students. Other programmatic features deal explicitly with helping students draw connections to, and prepare for, the future world-of-work. Strategies implemented to meet these goals include the integration of academic and vocational curricula, centered typically on a career theme, and exposure to potential careers via internships, job shadowing, and in-school presentations, as promoted through school-employer partnerships. Each of the schools in the evaluation had both a career-academy and other academic programs. Thus, within each school, MDRC identified a sample of students interested in attending the academy and randomized them to either receive, or not receive, an offer to enroll at the beginning of their high-school careers. The research team followed all sampled students through high school and for eight years beyond their scheduled, on-time high-school graduation. ITT analyses revealed a positive, statistically significant impact of the offer to enroll in an career-academy, for males, on subsequent monthly earnings.

Significance / Novelty of study: This research furthers our understanding of the earnings impacts observed in MDRC's evaluation of career academy high schools. We have a better understanding of an aspect of the career academy model that appears to have contributed to that benefit – school-sponsored opportunities to learn about and participate in the labor market. Methodologically, this work illustrates the use of the principal stratification framework for unpacking causal impacts and their underlying mechanisms in light of complications typical of experimental work in the social sciences.

Statistical, Measurement, or Econometric Model: The data of primary focus for each student include levels of the potential mediators, subsequent earnings, treatment assignment, and covariates. I employ three different statistical models. The first is an ordinal probit model that describes the distribution of the level of the mediator under assignment to treatment, as a function of baseline covariates. The second is an analogous ordinal probit model that describes the distribution of the level of the mediator under assignment to control. The third is a linear regression model that describes the distribution of earnings as a function of stratum membership (as defined by the potential mediators, jointly), treatment assignment and baseline covariates.

Usefulness / Applicability of Method: Other methods proposed for exploring mediational hypotheses include multiple regression analysis and instrumental-variables (IV) estimation. Despite their popularity, regression- and IV-based approaches rely on sets of untestable assumptions that may be too strong to yield valid inferences about the causal mechanism or mechanisms under investigation in most practical empirical settings. By highlighting the stringent assumptions that must hold for one to achieve valid causal inferences using either of these two approaches, I motivate the particular benefits of investigating mediational processes using principal stratification, with which one is able to relax and test the assumptions underlying these other analytic approaches. Of course, no approach is without its limits. Therefore, I aim to be clear throughout on the limitations of this approach

Research Design: Given this finding, questions follow regarding the explicit mechanism – or mechanisms – through which the academy offer caused these impacts for males, especially as the academy offer did not lead to better standardized test scores, higher probability of high-school graduation, or greater post-secondary educational attainment. The question that emerges, therefore, concerns the aspects of the academy programs that led to the subsequent impacts on labor-market success, when they did not improve students' immediate educational success. One of the most salient aspects of the career-academy model is the connection that academy programs seek to build between students and the labor market. If students gained the skills and knowledge important for early labor-market success, then the impacts of the academy offer may be driven by participation in career-related programmatic activities. I use a principal stratification approach to test the hypothesis that school-sponsored exposure to the world-of-work mediates the positive effect of the career-academy offer on subsequent monthly earnings, among males.

Data Collection and Analysis: During the nearly 12-year time frame of the evaluation, MDRC surveyed participants multiple times: at baseline, after 1-2 years of high school, at the scheduled completion of high school, and one year, four years, and eight years after scheduled completion of high school. Because in-school labor-market exposure aspects of the academy programs were most prevalent in the later years of high school, the survey conducted at the scheduled completion of high school provides the most complete information regarding students' school-sponsored labor-market exposure. I conducted a principal components analysis of student responses to a selected set of survey items and utilize each student's score on the obtained first principal component as a continuous measure of school-sponsored exposure to the world-of-work. The outcome of interest, average monthly earnings, is calculated based on student survey responses from the period of four to eight years after scheduled high school completion.

In analyzing experimental data, researchers are often interested in looking beyond a simple estimate of the average effect of treatment. Here, interest is in evidence that supports or refutes the hypothesis that school-sponsored exposure to the world-of-work is a key mediator of the earnings impacts observed. The framework of principal stratification (Frangakis & Rubin, 2002) provides an approach for exploring this type of mediation hypothesis. In general terms, principal stratification is an analytic framework whereby the potential values of a post-randomization variable representing an analytic challenge or related to a question of interest (here, the mediator, school-sponsored exposure to the world of work) are used to stratify the sample so that pivotal analyses can be conducted within each stratum and the results utilized to inform the causal question of interest. In utilizing principal stratification to explore exposure to the world-of-work

as a mediator, I must engage the following set of questions for each student: (1) What was his level of school-sponsored exposure to the world-of-work, given his randomized treatment assignment? And (2) what would his level of school-sponsored exposure to the world-of-work have been, had he been assigned to the counterfactual experimental condition?

The permutations of the possible answers to these questions, together, define the principal strata. Therefore, I conceive of individuals as belonging to one of several latent classes (or principal strata) defined by the mediator's potential values under assignment to treatment and assignment to control, jointly. In utilizing the principal stratification framework, the analytic crux is that for no student do we observe all of the information necessary for determining individual stratum membership. For example, among those assigned to the treatment condition, we do not observe the level of the mediator that each would have exhibited under the control condition. Therefore, following the work of others (e.g., Hirano et al., 2000; Jin & Rubin, 2008, 2009; Gallop et al., 2009), I utilize a Bayesian approach to estimation and inference that relies on a set of assumptions, statistical models, and available information to infer each student's likely stratum of membership and to estimate stratum-specific treatment effects.

Findings / Results: My results indicate that for about 40 percent of students, the randomized academy offer would have no impact on exposure to the world-of-work: 17 percent of students would be exposed at a low level, 13 percent at a moderate level, and 10 percent at a high level, under assignment to either the treatment or control condition. Further, for these students, the estimated average effects of treatment assignment are closer (in absolute terms) to zero compared to those in strata where the mediator level is impacted by treatment assignment. This provides evidence that the direct effect of treatment, or that component of the effect that does not pass through school-sponsored exposure to the world-of-work, is minimal. A key feature worth emphasizing, however, is that this principal stratification estimate of the direct effect is obtained from the subset of strata where the the mediator would be unchanged by treatment assignment. For approximately 60 percent of students, the offer to attend an academy does impact exposure to the world-of-work: approximately 20 percent of students would be induced to obtain a moderate rather than low level of exposure; 16 percent would obtain a high rather than moderate level, and nearly one quarter would experience a considerable change in their level of exposure (moving from low to high), given the offer to enroll in an academy. In these groups where level of exposure does differ as a result of the academy offer, ITT estimates are comparatively larger. Further, the largest estimated treatment effect is among those students who would realize the largest shift in exposure to the world-of-work if offered the opportunity to enroll in an academy.

Conclusions: The elegance of the principal stratification approach, in general, is that because stratum membership is considered a pre-treatment covariate (albeit only partially observed), it is orthogonal to the academy offer, which was assigned at random. The implication is that within-stratum ITT estimates are unbiased estimates of stratum-specific treatment effects. In those strata where the level of the mediator would not be affected by treatment assignment, the principal stratification approach allows for the estimation of the direct effect of treatment. In those where the level of the mediator would be affected, this approach allows for the estimation of total effects of treatment. Results associated with this principal stratification analysis are consistent with the hypothesis that career academies generate positive effects by inducing greater exposure to the world-of-work and that this, in turn, yields higher average earnings, on average.

Appendix A. References

- Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173–1182.
- Bloom, H. S., ed. (2006). *Learning More from Social Experiments: Evolving Analytic Approaches*. New York: Russell Sage Foundation.
- Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology*, 98, 550-558.
- Frangakis, C. E. & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58, 21-29.
- Gallop, R., Small, D. S., Lin, J. Y., Elliott, M. R., Joffe, M., & Ten Have, T. R. (2009). Mediation analysis with principal stratification. *Statistics in Medicine*, 28, 1108-1130.
- Gennetian, L. A., Morris, P. A., Bos, J. M., & Bloom, H. S. (2006). Constructing instrumental variables from experimental data to explore how treatments produce effects. In *Learning More from Social Experiments: Evolving Analytic Approaches*, H. S. Bloom, ed., New York: Russell Sage Foundation, 75-114.
- Hirano, K., Imbens, G. W., Rubin, D. B., & Zhou, X-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, 1(1), 69-88.
- Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15, 309-334.
- Jin, H. & Rubin, D. B. (2008). Principal stratification for causal inference with extended partial compliance. *Journal of the American Statistical Association*, 103, 101–111.
- Jin, H. & Rubin, D. B. (2009). Public schools versus private schools: Causal inference with partial compliance. *Journal of Educational and Behavioral Statistics*, 34, 24-45.
- Kemple, J. J. (2008). *Career academies: Long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood*. New York: MDRC.
- Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, 31, 161-170.
- Robins, J. M. & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3, 143–155.
- VanderWeele, T. J. & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*, 2, 457-468.