Abstract Title Page
Not included in page count.

Title:

Addressing the Don't Ask, Don't Tell Practice in Observational Studies: Using Interviews to Understand the Assignment Mechanism

Author(s):

Jordan H. Rickles
University of California, Los Angeles

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

Valid causal inference depends on the process by which the units of analysis select, or are assigned, to the treatment condition(s) of interest. Rubin (1991, 2004, 2005) emphasizes this point and argues that statistical inference for causal effects "requires the specification of a posited assignment mechanism describing the process by which treatments were assigned to units" (1991, p. 403). Most scholars agree that randomized controlled trials and regression discontinuity designs can produce unbiased causal effect estimates because the assignment mechanism is known. Most inquiries into causal effects in education, however, do not utilize either of these two designs and thus depend on tenuous assumptions regarding the assignment mechanism.

Far too often, researchers simply take a "don't ask, don't tell" approach to questions surrounding the assignment mechanism and carry out an analysis armed with assumptions and veiled by the available data at hand. A more conscious effort to address the assignment mechanism can go a long way to addressing the confidence one places in causal effect estimates in an observational study. Cook, Shadish and Wong (2008) reviewed within-study comparisons to identify the conditions under which observational studies can produce estimates comparable to experiments. They find that "[k]nowledge of the selection process can significantly reduce selection bias provided the selection process is valid and reliably measured" (p. 740). Knowledge of the selection process is particularly challenging within education, where the multilevel nature of the educational system (students, teachers, schools, districts) and the possibility of multiple decision makers in the assignment process (students, parents, teachers, administrators) can result in complex assignment mechanisms. If research on program and policy effects is to play a prominent role in educational policy making, researchers must convincingly address these assignment mechanism complexities.

If treatment assignment depends on a set of factors (S) and these factors are associated with an individual's potential outcomes, the selection independence assumption breaks down (Holland, 1986). Researchers and evaluators trying to draw causal inferences from observational data must resort to research design and statistical adjustments to account for selection bias resulting from the true assignment mechanism. After these adjustments—which typically manifest in a regression model controlling for the observed factors in S or matching/stratifying based on the observed factors in S—the validity of any causal findings hinges on an assumption of "strongly ignorable treatment assignment" (Rosenbaum & Rubin, 1983). The assumption of "strong ignorability" states that the assignment mechanism is conditionally independent from the potential outcomes. If true, this implies that observed conditional treatment and control group differences in the outcome measure can be attributed to the treatment in question and not pre-existing differences between the two groups. Within economics this assumption is often referred to as "selection on the observables" (Heckman & Hotz, 1989).

The extent to which the assumption of strong ignorability holds depends on whether the observables controlled for in the analysis adequately capture the assignment mechanism, but far

too often this is a subjective exercise left to the reader. Despite the importance of the assignment mechanism explicit in Rubin's writings and advice from Morgan and Winship (2007) that "the first step in analysis [in an observational study] is to investigate the treatment selection mechanism" (p. 41), this process is often taken for granted in practice. Most researchers rely on the current knowledge base and theory to justify strong ignorability, but studies are often limited to the secondary data at hand and theoretical discussions of the assignment mechanism are a post hoc justification rather than an empirical investigation. Furthermore, researchers can draw on varying, and often conflicting, theories within the social sciences to describe the assignment mechanism. For example, economists may emphasize a rational choice model while sociologists may emphasize cultural and organizational factors influencing selection. Additional uncertainty arises when the general body of knowledge may not apply to the specific topic, population, or setting for the study in question.

Purpose / objective / research question / focus of study:
Description of what the research focused on and why.

This paper illustrates how information collected through interviews can develop a richer understanding of the assignment mechanism, which can result in more plausible causal effect estimates from observational studies and provides a roadmap for sensitivity analysis. Focusing on the issue of assignment to algebra in 8th grade, I show how a preliminary data collection effort aimed at understanding the assignment mechanism is particularly beneficial in multisite observational studies in education.

Rosenbaum and Silber (2001) provide an example of how ethnographic methods, namely thick description, can improve the use of quantitative data analysis. The purpose of this paper is to provide an example of how interviews can improve the use of quantitative data analysis. Conducting interviews about the assignment process has three main objectives: (1) to gather current and localized information about the factors associated with treatment assignment; (2) to gather information on not just which factors influence assignment but how they influence assignment; and (3) to identify heterogeneity in the treatment assignment process across schools.

Setting:
Description of where the research took place.

The data were collected in middle schools within a California school district spanning urban and suburban communities.

Population / Participants / Subjects:
Description of participants in the study: who (or what) how many, key features (or characteristics).

From a random sample of 20 middle schools within the district, I interviewed a key decision maker at ten schools covering six of the district's eight regional sub-districts. The decision maker was identified by the school principal as the most knowledgeable about the assignment process. The identified decision-makers were primarily assistant principals responsible for mathematics instruction or math coaches. At one school the principal self-identified as the appropriate interviewee and at another school the 8th grade counselor was identified as the interviewee.

Intervention / Program / Practice:
Description of the intervention, program or practice, including details of administration and duration.

The ultimate goal for the research project in question is to estimate the causal effect of taking 8th grade algebra on high school academic achievement. Historically, most students take their first formal algebra class in high school. However, an emphasis on 8th grade algebra has grown since the mid-1990s, with about half of all 8th graders now taking algebra. The issue of early access to algebra re-entered the policy debate in June 2008 when the California State Board of Education approved a policy to test all 8th graders on algebra content standards—effectively pushing all California 8th graders into algebra.

This paper examines the methodological practice of addressing causal effect research questions like the above without first addressing questions about the assignment mechanism. While the substantive details of this paper deal with 8th grade mathematics assignment, the findings are relevant to studies pertaining to broader topics such as ability grouping, tracking, and curricular intensity. More broadly, this paper exemplifies methodological practice applicable to situations where one wishes to make causal effect estimates under an uncertain assignment mechanism.

Research Design:
Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

This paper discusses the use of exploratory data collection and analysis to inform quasi-experimental or observational study research designs.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

To gather consistent information on the 8th grade mathematics assignment process I developed an interview protocol consisting of three parts. The first part is a series of six semi-structured, open-ended questions designed to elicit information on the decision-making process, types of students the school places in algebra and pre-algebra, and the decision maker's general philosophy about students taking algebra in 8th grade. The second part was designed to provide a standardized summary of the assessment mechanism across the schools. This section contains two scenarios describing two hypothetical students and asks the interviewee to rate, on a scale of 1 to 5, whether they would place the student in algebra or pre-algebra. Open-ended follow-up questions ask the interviewee to elaborate on why they chose the rating they selected and what other information they would like to know before making a final decision on each student. The last part of the protocol asks interviewees to rate, on a scale of 1 to 5, how important each of 16 pieces of information is in their decision to assign 8th graders to algebra or pre-algebra. This section was designed to provide a standardized assessment of data use across the schools, as well as, provide supporting or refuting evidence for the other two sections.

To analyze the data, I coded the schools based on four different characteristics of the assignment process: (1) whether more weight is given to objective criteria (e.g., standardized test scores) or subjective criteria (e.g., teacher recommendations) during the decision-making process; (2) whether decisions are primarily dictated by data systematically collected by the district (e.g.,

course grades) or non-systematic data (e.g., diagnostic tests); (3) whether course assignment is based on well-defined inclusion/exclusion decision rules or not; and (4) whether the decision maker's (or school's) philosophy regarding 8th grade mathematics assignment takes a more protectionist stance (e.g., protect students from failure) or a more laissez-faire stance (e.g., students have the right to try). These codes, along with summaries of close-ended questions in the interview, allow for a general description of the assignment mechanism, including: whether selection is on the observables; whether data should be modeled as continuous, linear measures or discrete, categorical measures; and how the assignment mechanism varies across schools.

Findings / Results:
Description of main findings with specific details.

The interviews indicate that schools draw from a common battery of sources to determine which students are ready for algebra in 8th grade and which should take a pre-algebra course. However, the data reveal two distinct and important ways in which use of these information sources vary across schools. First, schools place different weight on the sources they depend on during the decision making process. Second, schools use different decision rules (or cut-points) across the battery of sources to guide the decision making process. Additionally, I find evidence that the assignment mechanism at a school is associated with the school's (or decision maker's) underlying philosophical orientation regarding what is in a student's best interests.

In determining whether a student should take algebra or pre-algebra, decision makers in the district can draw on many different sources of data. However, as one interviewee put it, "you can only look at so much with 600 kids." The most common information cited as aiding the decision making process included objective measures of mathematics proficiency such as the annual standardized state mathematics test (CST) and an optional mathematics diagnostics test (MDTP), as well as more subjective measures such as grades in the 7th grade mathematics course and teacher recommendations. Some schools also mentioned looking at other mathematics assessments, course marks for work habits, and whether a student is in GATE or honors classes. Contrary to prior research on similar selection processes (Hallinan, 1994; Oakes, Gamoran & Page, 1992), I found little evidence that the assignment process was influenced by student and parent preferences (except in rare occasions), and none of the schools indicated that resource or staffing limitations affected course taking decisions. Research on the causal effects of course placement would ideally factor all these pieces of information into the research design and statistical models for effect estimation, some of which, however, are rarely observed and may require sensitivity analysis to examine omitted variable bias.

Research into the causal effects of course placement must also account for heterogeneous use of information sources in the assignment process. Two of the ten schools relied almost exclusively on objective measures of mathematics proficiency, while four of the schools relied almost exclusively on subjective information, particularly teacher recommendations and course grades. Even when schools relied on the same data sources to determine placement, they differed in the benchmarks employed to determine algebra readiness. For example, schools using the MDTP looked at the percent correct score for students, but the benchmark for algebra placement ranged from 60% to 80% correct depending on the school. Similarly, schools using the CST (which categorizes students into one of five performance bands) or course grades used different benchmarks to distinguish algebra students from pre-algebra students. Thus, statistical models of

the assignment mechanism and causal effects will produce biased estimates if these data are not properly modeled to account for discrete benchmarks and heterogeneity across schools.

During the course of data collection, it became apparent that decision makers—and by extension schools—had an underlying philosophy regarding course taking, and this philosophy was associated with the school's assignment mechanism. The philosophy was most apparent when discussing students who fall into a grey zone where, for example, a student meets one of the criteria for algebra placement but not another. Three of the ten schools exhibited a protectionist philosophy toward course placement, with a general notion to not "program a kid for failure" or "set a child up for failure." In these schools, the borderline students are more likely to get placed in pre-algebra instead of algebra. This is in contrast to schools with a more laissez faire philosophy toward course placement, with a general notion that "kids have a right to fail." As one interviewee in a laissez faire school described, "[i]t's looking for any area of hope, any one single piece of data that might lend itself to getting up to the algebra course."

Responses to the two interview scenarios provided further evidence for assignment heterogeneity across schools and the relationship between assignment and the school's philosophy toward placement. For both scenarios, schools were evenly divided between those favoring pre-algebra placement and those favoring algebra placement, although there was more uncertainty among the second scenario (see Figure 1). Schools with a protectionist philosophy all favored placing the two hypothetical students in pre-algebra while schools with a laissez faire philosophy generally favored placing the two students in algebra.

Conclusions:

Description of conclusions and recommendations based on findings and overall study.

In many situations, policy makers and practitioners desire research on policy or program effectiveness under conditions where random assignment is not feasible. When the assignment mechanism is unknown to the researcher, I argue that, whenever possible, formal investigation of the assignment mechanism should be built into the research design, possibly as part of a pilot study. The findings presented in this paper, based on ten interviews, show how collecting information about the assignment mechanism can aid statistical adjustment for causal effect estimation in observational studies. In particular, I was able to determine which types of data play a role in the assignment process, and therefore what should be accounted for to invoke the assumption of strong ignorability. If some factors are unobservable, then the interviews provide a better sense of what to investigate through sensitivity analysis. The interviews also suggest heterogeneity in the assignment mechanism across schools. This implies propensity score models or regression models for causal effects should be estimated separately for each school, or random effect, multilevel modeling techniques should be employed.

These findings likely extend beyond the current research project in question to broader educational policy issues pertaining to ability grouping, tracking, differential course taking, and curricular intensity. Prior research that did not account for the described assignment process complexities likely resulted in biased estimates.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: new findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27 (4), 724-750.

Hallinan, M. T. (1994). Tracking: from theory to practice. *Sociology of Education*, 67 (2), 79-84.

Heckman, J. J., & Hotz, J. V. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. *Journal of the American Statistical Association*, 84 (408), 862-874.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

Oakes, J., Gamoran, A., & Page, R. N. (1992). Curriculum differentiation: opportunities, outcomes, and meanings. In P. W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 570-608). New York: Macmillan Publishing Company.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41-55.

Rosenbaum, P. R., & Silber, J. H. (2001). Matching and thick description in an observational study of mortality after surgery. *Biostatistics*, 2 (2), 217-232.

Rubin, D. B. (1991). Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, 4, 1213-1234.

Rubin, D. B. (2004). Teaching statistical inference for causal effects in experiments and observational studies. *Journal of Educational and Behavioral Statistics*, 29 (3), 343-367.

Rubin, D. B. (2005). Causal inference using potential outcomes: design, modeling, decisions. *Journal of the American Statistical Association*, 100 (469), 322-331.

Appendix B. Tables and Figures

Not included in page count.

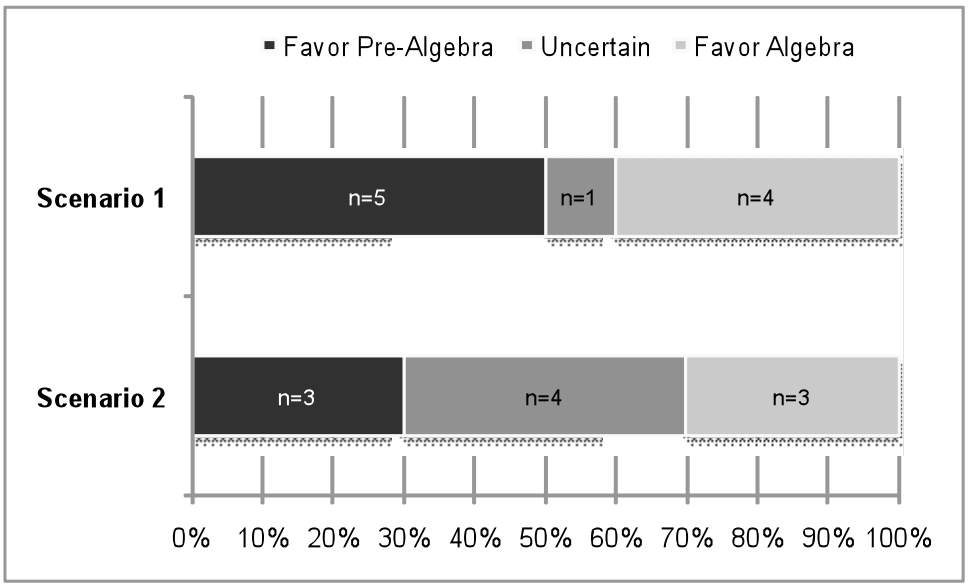

Figure 1. Responses to each Student Scenario's Likely Course Placement (N=10)

Note: tabulations based on collapsed responses to a 5-point question about likely course placement of each hypothetical student, where 1 = definitely pre-algebra and 5= definitely algebra. Responses of 1 and 2 were coded as "favor pre-algebra," responses of 3 were coded as "uncertain," and responses of 4 or 5 were coded as "favor algebra."

Scenario 1: *In 7th grade, Martin got a C in his first semester math class and a D the second semester. He received C's and B's in his other classes. He also received a mix of satisfactory and unsatisfactory marks for work habits and cooperation. In 6th grade, Martin's math grades were a little higher, with a C the first semester and a B the second semester. Similarly, he scored Basic on the 6th grade math CST and Below Basic on the 7th grade math CST. You heard a couple of Martin's 7th grade teachers mention that he started slipping behind and became more of a disruption in class as the year progressed.*

Scenario 2: *Maya moved to California from Mexico during her 6th grade year and started attending this school in 7th grade. She is an English learner and is struggling to keep up in most of her classes. She received mostly D's in 7th grade, but got a C in her second semester math class. Her work habits and cooperation marks are all satisfactory and there is no mention of any disciplinary problems in her records. She scored Far Below Basic on her 6th grade math and ELA CST tests but scored Basic on her 7th grade math test. You do not know much else about her except what is in her official record.*