Paper 3

Abstract Title Page

Not included in page count.

Title: Mediation and Spillover Effects in Group-Randomized Trials with Application to the 4Rs Evaluation

Author(s):
Tyler J. VanderWeele (Harvard School of Public Health) tvanderw@hsph.harvard.edu
Guanglei Hong (University of Chicago, contact author) ghong@uchicago.edu
Stephanie M. Jones (Harvard Graduate School of Education) jonesst@gse.harvard.edu
Joshua L. Brown (Fordham University) joshua.brown@nyu.edu

Abstract Body

Limit 5 pages single spaced.

Background / Context:

Description of prior research and its intellectual context.

A common setting in educational research consists of a randomized intervention at the school level, a mediator of interest at the classroom or teacher level and an outcome of interest at the child level. A common approach to addressing mediation in such settings consists of regressing the outcome on the treatment with and without the mediator variable. This approach to mediation analysis is subject to several limitations. First, the approach ignores selection into mediator levels; although the treatment is randomized, the mediator is not and thus analyses ignoring this selection issue are subject to potentially severe biases due to confounding (Judd and Kenny, 1981; Robins and Greenland, 1992; Pearl, 2001). The second issue with the standard regression approach is that potential interaction between the effects of treatment and the mediator on the outcome are typically ignored. Recent literature on causal inference has made clear that mediation analysis becomes considerably more complex when such interactions are present (Pearl, 2001). A third issue with the standard regression approach is that it ignores issues of interference and spill-over effects. Child level outcomes may depend not only on the characteristics of the child's own classroom but also on the characteristics of other classrooms because of social interactions among children from different classrooms. The issue is referred to as one of interference between units in the statistics literature (Cox, 1958). No interference between units is a component of Rubin's Stable Unit Treatment Value Assumption or SUTVA (Rubin, 1980, 1986). The assumption will be violated in settings in which social interactions allow one individual's exposure to affect the outcomes of other individuals. Such interference is part of the theoretical rationale of the 4Rs program which focuses on bringing together educators' collective efforts within a school. Analyses of causal effects are considerably more complex in the face of such interference.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

In this paper we extend recent work on mediation in a multilevel setting (VanderWeele, 2010) and on causal inference under interference among units (Hong and Raudenbush, 2006; Hudgens and Halloran, 2008; Rosenbaum, 2007; Sobel, 2006) to develop a template for the mediation analysis of group randomized educational interventions. The present work will contribute to the literature on interference, in particular on interference in the context of mediation analysis. We will show that not only does the total effect of the intervention decompose into a direct effect and an indirect effect mediated through classroom quality but also that the indirect effect itself decomposes into an effect mediated through the quality of a child's own classroom and a spillover effect from the quality of the other classrooms at a school. We will give some results for the identification of these direct, indirect and spillover effects and consider the consequences of ignoring interference when it is in fact present. We will then analyze the effects of the Reading, Writing, Respect and Resolution (4Rs) intervention in a group randomized trial.

Significance / Novelty of study:

Description of what is missing in previous work and the contribution the study makes.

Peer influence and social interactions can give rise to spillover effects in which characteristics of one individual unit may affect outcomes of other individual units. Evaluators who choose groups rather than individuals as experimental units in group randomized trials often anticipate that the desirable changes in targeted social behaviors will be reinforced through interference among individuals in a group exposed to the same treatment. Failure to account for such spillover effects can result in bias and problems with interpretation. Using a counterfactual conceptualization of direct, indirect and spillover effects, we provide a framework that can accommodate issues of mediation and spillover effects in group randomized trials.

Statistical, Measurement, or Econometric Model:
Description of the proposed new methods or novel applications of existing methods.

In this paper, we are interested in the extent to which the effect of the 4Rs intervention on child outcomes is mediated by classroom quality. Let T_k denote the school-wide randomized treatment for school k (1 for the 4Rs intervention; 0 for control). Let M_{jk} denote the classroom level mediator for classroom j in school k. In the 4Rs intervention study the mediator of interest is a continuous measure of classroom quality. Let J_k denote the number of classrooms in school k. Let Y_{ijk} denote the child-level outcome for child i in classroom j and school k. In the 4Rs study this outcome is a continuous score measuring depressive symptoms. Let $Y_{ijk}(t_k)$ denote the potential or counterfactual outcome that child i in classroom j and school k would have obtained if the school-level treatment, T_k, were set to t_k. Similarly, let $M_{jk}(t_k)$ denote the potential or counterfactual mediator that classroom j in school k would have obtained if the school-level treatment, T_k, were set to t_k. We assume that children do not change schools as a result of the treatment to which a particular school is assigned. Hong and Raudenbush (2006) referred to this assumption as that of "intact clusters." We also assume that there is no interference between schools (i.e. that the treatment received at one school does not affect the outcomes of the children at any other schools). To incorporate within-school interference into our potential outcomes notation, we let $Y_{ijk}(t_k, m_{jk}, \mathbf{m}_{-jk})$ denote the counterfactual outcome that child i in classroom j and school k would have obtained if the school-level treatment in school k were set to t_k, if the quality in classroom j of school k were set to m_{jk} and if the quality of all other classrooms in school k were set to the vector $\mathbf{m}_{-jk} = (m_{1k}, \ldots, m_{j-1k}, m_{j+1k}, \ldots, m_{J_k k})$. Following Hong and Raudenbush (2006) and Hudgens and Halloran (2008), we assume that the potential outcome $Y_{ijk}(t_k, m_{jk}, \mathbf{m}_{-jk})$ depends on $\mathbf{m}_{-jk}$ through some scalar function $G(\mathbf{m}_{-jk})$ of $\mathbf{m}_{-jk}$ so that we may express the potential outcome as $Y_{ijk}(t_k, m_{jk}, G(\mathbf{m}_{-jk}))$. For example, $G(\mathbf{m}_{-jk})$ may denote the average quality for all classrooms in school k other than classroom j. Here we let $Y_{ijk}(t, m, g)$ denote the outcome for child i in classroom j and school k if the school received treatment t, the child's classroom had quality m, and the scalar function of the quality of other classrooms, $G(\mathbf{m}_{-jk})$, took the value g.

The causal contrast $E[Y_{ijk}(1, m, g) - Y_{ijk}(0, m, g)]$ captures the direct effect of the 4Rs program but also intervening to fix the quality of the child's own classroom to level m and intervening to fix the average quality of other classrooms to g. This quantity is referred to as a controlled direct effect of treatment. Likewise the contrast $E[Y_{ijk}(t, m, g) - Y_{ijk}(t, m^*, g)]$ could be used to assess the effect of a child's own classroom quality (comparing levels m and m^*) on a child's outcome and to examine whether the contrast varies with t or g. Similarly, the contrast $E[Y_{ijk}(t, m, g) - Y_{ijk}(t, m, g^*)]$ could be used to assess the spillover effect of the quality of classrooms other than the child's own and whether the contrast varies with t or m.

When the classroom quality mediators are set to the levels they would have been at under the control condition, the natural direct effect is defined as $E[Y_{ijk}(1, M_{jk}(0), G(\mathbf{M}_{-jk}(0))) - Y_{ijk}(0, M_{jk}(0), G(\mathbf{M}_{-jk}(0)))]$. We can also define a natural indirect effect as $E[Y_{ijk}(1, M_{jk}(1), G(\mathbf{M}_{-jk}(1))) - Y_{ijk}(1, M_{jk}(0), G(\mathbf{M}_{-jk}(0)))]$. As in the case of non-clustered treatments without interference (Pearl, 2001), the total effect of the intervention on the outcome $E[Y_{ijk}(1) - Y_{ijk}(0)]$ decomposes into natural direct and indirect effects. The decomposition will hold even if there are interactions between the effects of the treatment and the mediator on the outcome. The natural indirect effect further decomposes into a within-classroom mediated effect $E[Y_{ijk}(1, M_{jk}(1), G(\mathbf{M}_{-jk}(0))) - Y_{ijk}(1, M_{jk}(0), G(\mathbf{M}_{-jk}(0)))]$ and a spillover mediated effect $E[Y_{ijk}(1, M_{jk}(1), G(\mathbf{M}_{-jk}(1))) - Y_{ijk}(1, M_{jk}(1), G(\mathbf{M}_{-jk}(0)))]$.

We will let $\mathbf{X}_{ijk}$, $\mathbf{W}_{jk}$, and $\mathbf{V}_k$ denote child-level, class-level, and school-level baseline covariates, respectively. We will use $\mathbf{X}_{-ijk}$ to denote the vector of child-level baseline covariates for children in school k other than child i in classroom j, $\mathbf{W}_{-jk}$ to denote classroom-level baseline covariates for classrooms in school k other than classroom j. We will consider certain functions of the baseline covariates of other children in the classroom (or even at the school), $\mathbf{h}_1(\mathbf{X}_{-ijk})$, and of baseline covariates of classrooms other than a child's own $\mathbf{h}_2(\mathbf{W}_{-jk})$. To simplify notation we let $\mathbf{L}_{ijk} = (\mathbf{X}_{ijk}, \mathbf{W}_{jk}, \mathbf{V}_k, \mathbf{h}_1(\mathbf{X}_{-ijk}), \mathbf{h}_2(\mathbf{W}_{-jk}))$.

Usefulness / Applicability of Method:
Demonstration of the usefulness of the proposed methods using hypothetical or real data.

We present four identification results, one for controlled direct effects, one for natural direct and indirect effects, one for the spillover and within-classroom mediated effects, and one for the consequences of ignoring interference when it is in fact present. For sets of random variables **A**, **B**, and **C**, we will use $\mathbf{A} \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C}$ to represent that **A** is independent of **B** conditional on **C**.

Theorem 1. If for all t, m, g we have that $Y_{ijk}(t, m, g) \perp\!\!\!\perp T_k \mid \mathbf{L}_{ijk}$ and that $Y_{ijk}(t, m, g) \perp\!\!\!\perp \{M_{jk}, G(\mathbf{M}_{-jk})\} \mid T_k, \mathbf{L}_{ijk}$, then we can identify the controlled direct effect of the treatment and that of each mediator.

Theorem 2. If in addition to the assumptions stated in Theorem 1, we also have that $\{M_{jk}(t), G(\mathbf{M}_{-jk}(t))\} \perp\!\!\!\perp T_k \mid \mathbf{L}_{ijk}$ and that for all t, t^*, m, g, $Y_{ijk}(t, m, g) \perp\!\!\!\perp \{M_{jk}(t^*), G(\mathbf{M}_{-jk}(t^*))\} \mid \mathbf{L}_{ijk}$, then we can identify the natural direct effect and the natural indirect effect.

Theorem 3. If in addition to the assumptions stated in Theorems 1 and 2, we also have that for $t' \neq t^*$, $M_{jk}(t') \perp\!\!\!\perp G(\mathbf{M}_{-jk}(t^*)) \mid \mathbf{L}_{ijk}$, then we can identify the within-class mediated effect and the spillover mediated effect.

Theorem 4. Suppose that the assumptions stated in Theorems 1, 2, and 3 hold. And suppose we also require that for all t, $M_{jk}(t) \perp\!\!\!\perp G(\mathbf{M}_{-jk}(t)) \mid \mathbf{L}_{ijk}$, then we can ignore interference while still obtaining an estimate of the within-classroom mediated effect and obtaining the sum of a spillover mediated effect and the actual natural direct effect. However, even if all the above assumptions hold, if the substantive question of interest is whether classroom quality mediates the effect of treatment, ignoring interference would lead to an underestimate of the actual importance of classroom quality since it will not include an assessment of the effect mediated through the quality of other classrooms.

Research Design:

Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).
(May not be applicable for Methods submissions)

The 4Rs program is a school-based intervention in literacy development, conflict resolution, and intergroup understanding. The study (Jones, Brown, & Aber, in press) involved a 3-year, 6-wave longitudinal experimental design with measurements in the fall and spring semester of each year. The eighteen New York City elementary schools in the study were fairly representative of the demographic characteristics of New York City schools and included 923 students in 82 classrooms. The schools were pair matched based on twenty school characteristics including size, reading achievement, race/ethnic composition, mobility/two-year stability, school lunch receipt, expenditures, attendance and organizational readiness. Within each pair, schools were randomly assigned to either the 4Rs treatment or the control group. The intervention was implemented school-wide from grades K-6 for 3 years. All 3rd grade children in each school were followed over three years through 5th grade. In the application here, we will consider the first year of the study for the children beginning in third grade.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.
(May not be applicable for Methods submissions)

Classroom quality was measured in the spring semester using the CLASS scoring system (Pianta, La Paro, and Hamre, 2005) which assesses instructional support, emotional support, and organizational climate with an overall score between 1 and 7. We dichotomized this measure using 4.4 as the cutoff. The child-level outcome was depressive symptoms scored on a scale of 0 to 1. Covariates in the model were chosen based on prior empirical work (Brown, Jones, LaRusso, & Aber, 2010). The covariates were at least marginally predictive of either the outcome or the mediator. The models also included pair fixed effects to control for school-level factors. We fitted a multilevel model for the effect of treatment on the depressive symptoms, a multilevel model for the effect of treatment on class quality, and finally a multilevel model for the effects of treatment, classroom quality, and quality of other classrooms on depressive symptoms with the interactions between these variables saturated. The parameter values and model-based standard errors were estimated via maximum likelihood in HLM 6.0.

Findings / Results:

Description of the main findings with specific details.
(May not be applicable for Methods submissions)

The estimated treatment effect of the 4Rs intervention on depressive symptoms is -0.052 (*s.e.* = 0.023, $t = -2.29$, $p = 0.05$), suggesting a marginally significant effect of the treatment in reducing child depressive symptoms. The estimated treatment effect of the 4Rs intervention on classroom quality is 0.45 (*s.e.* = 0.20, $t = 2.28$, $p = 0.05$). In the control schools it appears that depressive symptoms are highest for children in classrooms in which the quality of the child's own classroom is low but the quality of other classrooms at the school is relatively high. Apparently it is also only in this type of classrooms that the 4Rs intervention has a statistically significant direct effect on depressive symptoms (see Table 1). However, a χ^2 test failed to reject the null hypothesis of no interaction of treatment with either a child's own classroom quality or the quality of other classrooms. We then obtained an estimate of the overall controlled direct effect of treatment of -0.058 (p = 0.025). If we take the χ^2 test result as an indication that the assumptions in Theorem 2 holds, then this estimate would also correspond to the natural direct effect. We could obtain a natural indirect effect by subtracting the natural direct effect from the total effect of treatment which would give -0.052 – (-0.058) = 0.006. This provides very little evidence that any of the effect of the 4Rs intervention on depressive symptoms is mediated through either the quality of a child's own classroom or the quality of the classes other than the child's own. Because this particular evaluation of the 4Rs intervention was powered to be able to assess only the total effect, not direct and indirect effects, our results here are at best suggestive.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

In this paper we have made a number of contributions to allow researchers to address questions of mediation and spillover effects in group randomized trials. The approach we have described here constitutes an advance over the standard approach to estimating direct and indirect effects that is often used in group-randomized trials. Specifically, our approach (i) makes explicit the assumptions required for identification that will be important in study design and data analysis of group randomized trials, (ii) accommodates possible interactions that may be present, (iii) allows for interference between individuals in different clusters (e.g., classrooms in the 4Rs evaluation), and (4) allows for the definition, identification, and estimation of spillover effects. In particular, by relaxing the no-interference assumption, we have been able to investigate spillover effects that will often be of substantive and theoretical interests. Interference is not simply a problem that must be dealt with but in fact gives rise to research questions about spillover effects that are of interest in their own right. In addition, we have provided an analysis of the mediation and spillover effects in the 4Rs evaluation. The chief limitations of the analysis are: (i) a relatively large sample size may be required to draw reliable inferences about mediation and spillover effects; and (ii) relatively strong identification assumptions are required to empirically estimate these effects from data.

The approach that we have presented here could be extended in a number of directions. First, future work could consider accommodating longitudinal settings as the mediator and outcome changes over time. Second, work has been done on using weighting techniques (van der Laan and Petersen, 2008; VanderWeele, 2009; Hong, 2010) rather than regression to address confounding control in questions of mediation analysis; future research could attempt to extend these weighting techniques to estimate and distinguish spillover mediated effects and within-classroom mediated effects. Third, further research could develop sensitivity analysis techniques to assess the extent to which an unobserved variable affecting both the mediator and the outcome (and thus giving rise to confounding of the effects of both the mediator in a child's own classroom and that of the mediator in the other classrooms) might invalidate the inference about direct, indirect and spillover effects.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Brown, J.L., Jones, S.M., LaRusso, M., and Aber, J.L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4Rs program. *Journal of Educational Psychology*, *102*(1), 153-167.

Cox, D. R., (1958). *Planning of Experiments*. New York: John Wiley & Sons.

Hong, G. (2010). Ratio-of-mediator-probability weighting for estimating natural direct and indirect effects. In *JSM Proceedings, Biometrics Section*. Alexandria, VA: American Statistical Association.

Hong, G., and Raudenbush, S.W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, *101*, 901-910.

Jones, S.M., Brown, J.L., and Aber, J.L. (in press). The longitudinal impact of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research, *Child Development: Special Issue on Raising Healthy Children: Translating Child Development Research into Practice*.

Hudgens, M.G. and Halloran, M.E. (2008). Towards causal inference with interference. *Journal of the American Statistical Association*, *103*, 832-842.

Judd, C.M. and Kenny, D.A. (1981). Process analysis: estimating mediation in treatment evaluations. *Evaluation Review*, *5*, 602-619.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*. San Francisco: Morgan Kaufmann, 411-420.

Pianta, R.C., La Paro, K., and Hamre, B. (2005). Classroom Assessment Scoring System (CLASS). Unpublished measure. University of Virginia, Charlottesville, VA.

Robins, J.M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, *3*, 143-155.

Rosenbaum, P.R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, *102*, 191-200.

Rubin, D. B. (1980). Comment on: "Randomization analysis of experimental data in the Fisher randomization test" by D. Basu. *Journal of the American Statistical Association*, *75*, 591-593.

Rubin, D. B. (1986). Which ifs have causal answers? Comment on: "Statistics and causal inference" by P. Holland. *Journal of the American Statistical Association*, *81*, 961-962.

Sobel, M.E. (2006) What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *Journal of the American Statistical Association*, *101*, 1398-1407.

van der Laan, M.J. and Petersen, M.L. (2008). Direct effect models. *International Journal of Biostatistics*, *4*, Article 23.

VanderWeele, T.J. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, *20*, 18-26.

VanderWeele, T.J. (2010). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociological Research and Methods*, *38*, 515-544.

Appendix B. Tables and Figures
Not included in page count.

Table 1
Controlled Direct Effects of 4Rs Program by Class Quality Indicators

	Coefficient	Standard Error	*t*
M=0, G=0			
Intercept	0.58	0.03	12.28***
Direct effect	-0.05	0.05	-1.18
M=0, G=1			
Intercept	0.64	0.06	10.16***
Direct effect	-0.13	0.05	-2.63*
M=1, G=0			
Intercept	0.58	0.06	10.08***
Direct effect	-0.01	0.05	-0.14
M=1, G=1			
Intercept	0.59	0.09	6.36***
Direct effect	-0.04	0.07	-0.56

*$p < .05$; ***$p < .001$