The direct and indirect psychosocial consequences of neighborhood disadvantage on cognitive development: A novel application of the parametric g-formula

Nick Graetz, University of Pennsylvania
Courtney Boen, University of Pennsylvania
Irma Elo, University of Pennsylvania

Short Abstract

In studying the effects of neighborhood disadvantage on child development, many analyses find small or negligible associations upon controlling for contemporary characteristics known to vary over age. However, some proportion of the effect of place on child development is expected to operate indirectly through other characteristics that may have varying or cumulative effects over time, such as parental income, quality of schooling, housing security, exposure to violence, and other stressors. The present study seeks to disentangle the direct and indirect effects of neighborhood exposures throughout childhood on cognitive development. Taking advantage of the prospective cohort design of the Fragile Families & Child Wellbeing Study, we employ the parametric g-formula to address issues of time-varying confounding. By studying the ways in which disadvantage develops both directly and indirectly over time, we add to the literature on geographic disparities in child development and methods for longitudinal analysis in life-course research.

Extended Abstract

Place is an important factor governing an individual's risk profile over their life course.[1] Many studies have documented associations between neighborhood disadvantage and disparities in health and disease.[2–6] The life-course perspective is a particularly useful paradigm for this research, focusing on how disparities in adult health and socioeconomic status (SES) are the result of a complex, dynamic interplay of factors accumulating as an individual ages. There is a particular emphasis on the timing and duration of exposures as well as the accumulating disadvantage associated with chronic exposure to certain stressors. These exposures often revolve around health indicators, indicators of socioeconomic status, and the dynamic relationships between them that unfold as an individual grows older.

Analyses employing a life-course framework aim to explain the ways in which broad risk exposures, such as living in a very poor neighborhood, actually operate on individuals over time to produce diverging trajectories in health and development. In the literature on neighborhood effects, there are several theories on the pathways through which disadvantage operates. Resource theories focus on lack of important institutional resources like schools, daycare, pharmacies, and safe recreational areas. Epidemiologists frequently examine environmental theories that focus on direct health outcomes. For example, certain health risks may result from the differential exposure to air pollution or toxins related to dilapidated housing that is increasingly concentrated in poor, marginalized neighborhoods. Sociologists focus on theories of social isolation, paying attention to the ways in which certain neighborhoods inhibit individuals from accessing labor markets and social network opportunities.[2,7] Other sociological studies focus on the ways in which social organization and cohesion work to affect crime or deviant behavior.[8,9]

There is a statistical challenge that applies to almost all observational analyses using a longitudinal life-course framework, but particularly plagues the study of neighborhood effects: "time-varying confounding". This refers to a scenario where the exposure of interest varies over time, and time-varying confounding occurs whereby past exposure levels affect current levels of confounders as well as the exposure itself. Traditional methods, such as regression modelling, matching, and propensity scoring, attempt to control for confounding in observational settings and are still frequently used in the study of neighborhood effects. However, these approaches as well as more modern approaches such as mixed-effects models, generalized estimating equations, and time-dependent Cox proportional hazard models have been demonstrated to be inadequate in controlling for time-varying confounding.[10] In examining the effect of neighborhood disadvantage on children's educational attainment, Wodtke et al. point out that certain time-varying characteristics, such as parental employment and mental health, are likely to affect both children's educational attainment and neighborhood disadvantage (example in Figure 1).[11] In this way, different neighborhood environments are affected by prior neighborhood conditions, and past neighborhood conditions affect contemporary confounders such as parental employment. Researchers in these scenarios intuitively want to control for the confounding effects of parental employment on the neighborhood-education relationship. However, this can easily introduce "over-adjustment bias", whereby any *indirect* effect of past neighborhood conditions on current education operating through current parental employment will be "controlled out" of the total average neighborhood effect.

Several previous efforts acknowledging these limitations of traditional methods use marginal structural models to estimate the average effect of neighborhood factors.[11,12] With a form of propensity score weighting based on the inverse-probability-of-treatment (IPTW), these studies acknowledge and *control* for time-varying confounding in the dynamic longitudinal process of neighborhood disadvantage, ideally removing it as a source of bias in the *total treatment effect* (example in Figure 2). For example, Wodtke et al. use this approach with longitudinal data from the Panel Survey on Income Dynamics (PSID) to estimate how a high index of neighborhood disadvantage reduces the probability of high school graduation from 96 to 76 percent for black children.[11] Lee & Jackson use marginal structural models to show how SES and health work reciprocally and dynamically over the life-course to produce inequality in child development.[12] However, these studies do not attempt to actually estimate the direct and indirect effects of neighborhood disadvantage. These indirect pathways are more than simply a source of confounding that is difficult to overcome with traditional regression techniques. They are themselves an important theoretical aspect describing how exposure to different neighborhoods manifests as structural disadvantage in a myriad of ways, and how each pathway develops across the life-course.

Our study utilizes the parametric g-formula approach, a recently developed methodology in clinical epidemiology and causal inference to account for issues of time-varying confounding (example in Figure 3).[13,14] A recent methodological overview of epidemiological models for situations involving time-varying confounding presented g-formula approaches in contrast with marginal structural models as the two modern methods most appropriate in adjusting for this potential bias.[10] The parametric g-formula approach is an extremely flexible, generalizable extension of traditional mixed-effects models that allows for the direct estimation of both the direct and indirect effects of neighborhood disadvantage, rather than averaging over these to arrive at an unbiased estimate of the total average effect as in marginal structural models. Beyond clinical trial applications, several studies have demonstrated the utility of this computational framework for observational data in social epidemiology and demography, such as exploring the direct and indirect pathways of the social determinants of fertility.[15] In the present analysis, we use multiple waves of a nationally-representative prospective cohort study that follows children from ages 0 to 15 across large cities within the United States. We examine how neighborhood characteristics and their unequal distribution manifest as disparities in cognitive development by the end of adolescence, paying close attention to the indirect effects that operate through exposure to violence, parental incarceration, social cohesion, child welfare contact, housing insecurity and eviction, as well as mental health and sociodemographic characteristics of parents. By characterizing the relative magnitude of the direct and indirect pathways through which neighborhood disadvantage develops and affects cognitive development, we discuss how social policies can be better structured around this dynamic process of accumulating disadvantage in order to most effectively mitigate diverging trajectories early in life.

Data

The Fragile Families & Child Wellbeing Study (FFCWS) is a longitudinal cohort study of children living in cities across the US. The sample is designed to be nationally representative of US cities with populations over 200,000, and sampling is clustered within 75 hospitals across 20 cities at the time of each child's birth. The study oversamples children born to unwed mothers (75% of total sample), as this circumstance was a major policy focus in the planning of this sample. Most existing survey data on family conditions have high rates of missing fathers, and little information on how these fathers differ from those who are included. The Fragile Families survey is population-based with a very low rate of missing fathers. This dataset also heavily oversamples non-Hispanic Black unmarried parents, who comprise 69% of the sample compared to 32% of the unmarried national population.

The FFCWS collected data at birth and in the year each child reached ages 1, 3, 5, 9, and 15. Survey modules were included for the mother, father, and child (ages 9 and 15). Additionally, we merged restricted datasets obtained via permission from FFCWS containing medical records data for mothers at birth, a set of contextual characteristics based on the census tract at each wave, and a set of characteristics of the child's school based on statistics from the National Center for Educational Statistics.

Methods and next steps

The parametric g-formula approach is the logical extension of the quite complicated conceptual diagrams at the beginning of many life-course analyses. However, decomposing change over the entire predictor space as above allows for estimation of both the direct and indirect effects of any one predictor on the outcome of interest. This can be used to observe effects that may manifest over time primarily through other variables, such as how past neighborhood manifests primarily through contemporary differentials in income, social isolation, exposure to violence, exposure to housing insecurity, etc. Applied to observational longitudinal data, this framework at best provides causal estimators. However, at worst it provides a more nuanced decomposition of life-course processes that does justice to critical theory in the field around accumulating risk, critical periods, and how disparities between groups develop dynamically over time in complex ways.

The parametric g-formula is itself not a model specification, but a computational framework for estimating all direct and indirect effects over time. The process involves the following steps:

1. Fit models for all time-varying predictors and the outcome of interest using all contemporary and lagged values. These models have any functional form and can be different across variable, as long as you can sample from the full probability distribution. In the present analysis, we use mixed-effect generalized linear models with appropriate likelihoods given each dependent variable.
2. Sample parameters from the full joint probability distributions of all models.
3. Bootstrap empirical data in the first period.
4. Simulate sample forward through each subsequent period.

5. Repeat Steps 2-4 as necessary in order to reach convergence in Monte Carlo error (Step 2) and error from sampling variability (Step 3). We repeat each step 1000 times.
6. Validation and sensitivity analyses to ensure the "natural course" estimated via Steps 1-5 resembles the empirical data in all periods.
7. Simulate counterfactual scenarios to decompose direct and indirect effects of interest.

Employing this framework, we report how divergence in cognitive development arises across childhood and adolescence both directly and indirectly from neighborhood disadvantage related to hyper-concentrated poverty, incarceration, and housing insecurity. We discuss implications for future research on the effects of neighborhood disadvantage and quantitative life-course research more broadly.

References

1. Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H. & Subramanian, S. V. Painting a Truer Picture of US Socioeconomic and Racial/Ethnic Health Inequalities: The Public Health Disparities Geocoding Project. *Am. J. Public Heal. J Public Heal.* **9595,** 312–323 (2005).

2. Douglas S. Massey *et al.* Neighborhood Disadvantage and Telomere Length: Results from the Fragile Families Study. *RSF Russell Sage Found. J. Soc. Sci.* **4,** 28 (2018).

3. Kothari, C. L. *et al.* The interplay of race, socioeconomic status and neighborhood residence upon birth outcomes in a high black infant mortality community. *SSM - Popul. Heal.* **2,** 859–867 (2016).

4. Hayward, M. D. & Sheehan, C. M. Handbook of the Life Course. (2003). doi:10.1007/b100507

5. Yang, Y. C. *et al.* Social relationships and physiological determinants of longevity across the human life span. *Proc. Natl. Acad. Sci.* **113,** 578–583 (2016).

6. Yang, Y. C., Gerken, K., Schorpp, K., Boen, C. & Harris, K. M. Early-life Socioeconomic Status and Adult Physiological Functioning: A Life Course Examination of Biosocial Mechanisms. *Biodemography Soc. Biol.* **63:2,** 87–103 (2017).

7. Sampson, R. J. Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure 1. *Source Am. J. Sociol.* **114,** 189–231 (2008).

8. Haskins, A. R. & McCauley, E. J. Casualties of context? Risk of cognitive, behavioral and physical health difficulties among children living in high-incarceration neighborhoods. *J. Public Health (Bangkok).* 1–9 (2018). doi:10.1007/s10389-018-0942-4

9. Park, J. M., Fertig, A. R. & Allison, P. D. Physical and mental health, cognitive development, and health care use by housing status of low-income young children in 20 American cities: a prospective cohort study. *Am. J. Public Health* **101 Suppl,** S255-61 (2011).

10. Mansournia, M. A., Etminan, M., Danaei, G., Kaufman, J. S. & Collins, G. Handling time varying confounding in observational research. *BMJ* **359,** j4587 (2017).

11. Wodtke, G. T., Harding, D. J. & Elwert, F. Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. *Am. Sociol. Rev.* **76,** 713–736 (2011).

12. Lee, D. & Jackson, M. The Simultaneous Effects of Socioeconomic Disadvantage and Child Health on Children's Cognitive Development. *Demography* **54,** 1845–1871 (2017).

13. Keil, A. P., Edwards, J. K., Richardson, D. B., Naimi, A. I. & Cole, S. R. The parametric g-formula for time-to-event data: intuition and a worked example. *Epidemiology* **25,** 889–97 (2014).

14. Vangen-Lønne, A. M. *et al.* Hypothetical interventions to prevent stroke: an application of the parametric g-formula to a healthy middle-aged population. *Eur. J. Epidemiol.* **3456789,** 1–10 (2018).

15. Bijlsma, M. J. & Wilson, B. Modelling the socio-economic determinants of fertility : A mediation analysis using the parametric g-formula. **49,**

Figure 1. Causal diagram for a traditional mixed-effects model involving some correlation structure to capture repeated observations within individual (random intercept, latent growth curve, etc.). Red dotted arrows indicate potential relationships that would introduce time-varying confounding into the estimation of the relationship between neighborhood disadvantage and cognitive development.

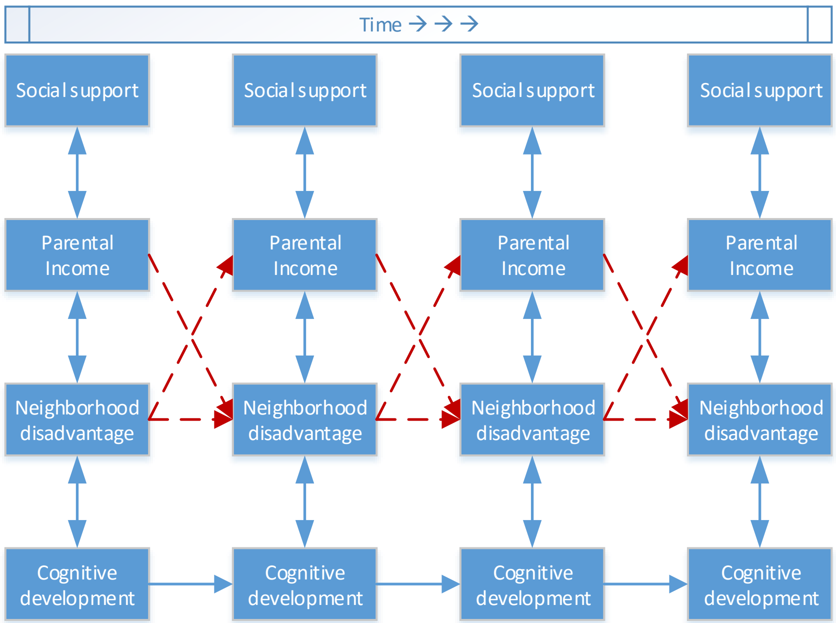


Figure 2. Causal diagram of the marginal structural model. This approach relies on weighting neighborhood exposure in each period, based on all confounders, in order to remove the bias of time-varying confounding whereby past characteristics affect future exposure to neighborhood disadvantage. In this way, the only factor that can affect current neighborhood disadvantage is past neighborhood disadvantage. This ideally results in an unbiased coefficient estimate for neighborhood disadvantage that is the sum of the direct effect and the indirect effects via other future confounders.

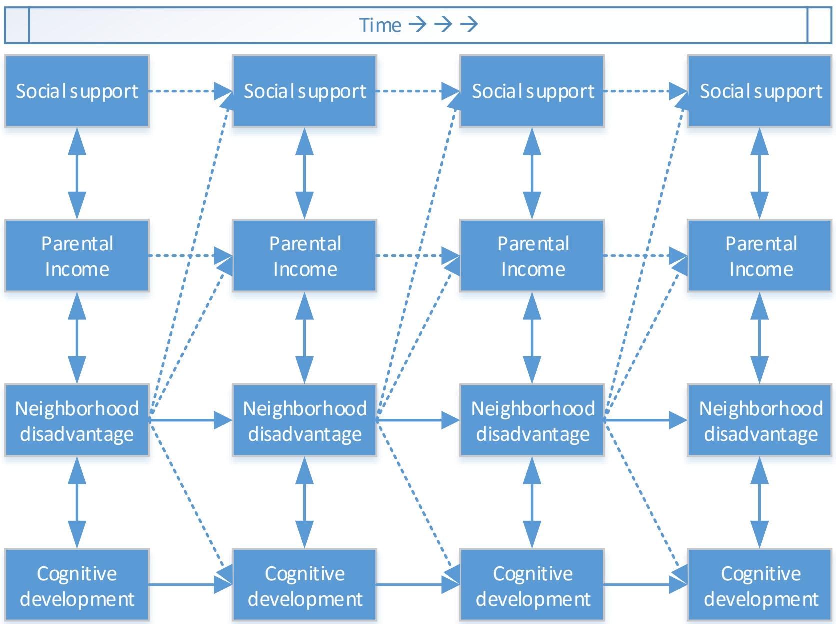
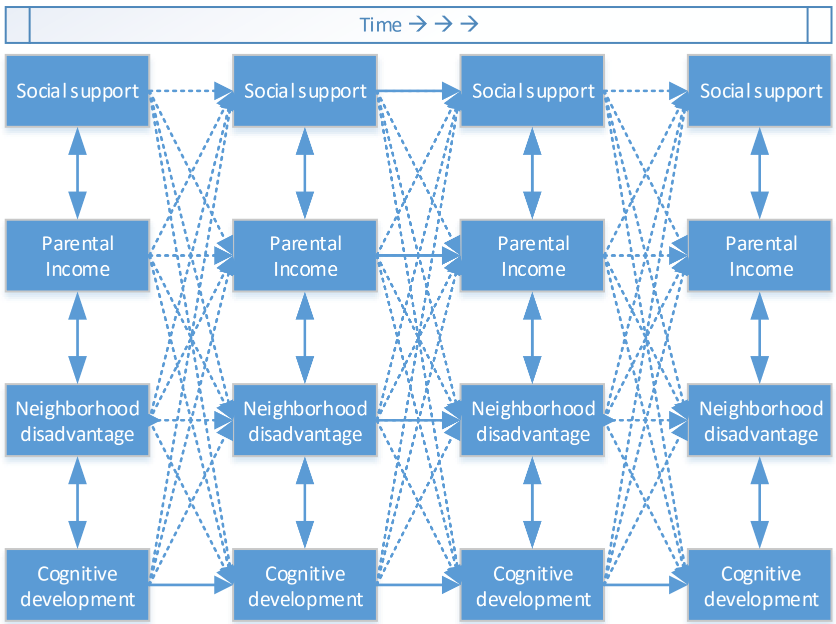


Figure 3. Causal diagram of the parametric g-formula approach. Combining longitudinal models for the exposure of interest (neighborhood disadvantage) and each time-varying confounder with a stochastic Monte Carlo simulation framework allows for the unbiased estimation of both direct and indirect pathways between neighborhood disadvantage and cognitive development.



$$Social\ Support_{i,t} = \beta \boldsymbol{X} + \alpha_i + \varepsilon_{i,t}$$

$$Parental\ income_{i,t} = \beta \boldsymbol{X} + \alpha_i + \varepsilon_{i,t}$$

$$Neighborhood_{i,t} = \beta \boldsymbol{X} + \alpha_i + \varepsilon_{i,t}$$

$$Cognitive_{i,t} = \beta \boldsymbol{X} + \alpha_i + \varepsilon_{i,t}$$