

On the Robustness and Validity of Groups

Robert J. Sampson,^{1,3} John H. Laub,² and Elaine P. Eggleston²

In response to Nagin's comment on our paper, "Methodological Sensitivities to Latent Class Analysis of Long-Term Criminal Trajectories," we reconsider the robustness and validity of group-based approaches to criminal trajectories and introduce additional issues for future research. We emphasize the limitations of typological approaches and the dangers of reifying the idea of distinct offender groupings for research and policy. We also clarify misunderstandings about the use of hierarchical linear models for studying trajectories of crime. Our basic conclusion is that methods in criminology need to be more tightly linked with theory.

KEY WORDS: trajectories; groups; longitudinal research methods.

The original intention of our article was to inject some critical thinking into the growing use of semiparametric, group-based methodology (SPGM) for the analysis of criminal trajectories. It was our feeling that SPGM, like many a method that has come before it (e.g., LISREL, HLM), was increasingly being used in criminology in a non-reflective or complacent fashion, often without concern for data quality or model assumptions. Our goal, then, was to examine the sensitivity of inferences drawn from SPGM to research-design and measurement variations in longitudinal data that are common in criminological research.

We are grateful that the leading expert on SPGM has taken the time to comment on our work, and we are even more heartened that Daniel Nagin agrees with our goal. His response to our effort demonstrates a deep understanding of the methodological issues involved, and he wisely places the debate in a larger context. Nagin's main argument is that the sensitivities we uncover are generic to all longitudinal research methods. We agree with the exceptions noted below, and if this exchange has the effect of increasing

¹Department of Sociology, William James Hall, 33 Kirkland Street, Harvard University, Cambridge, MA 02138.

²Department of Criminology and Criminal Justice, University of Maryland, College Park, MD.

³To whom correspondence should be addressed: E-mail: rsampson@wjh.harvard.edu

awareness among criminologists to the non-sexy but nonetheless vitally important issues of follow-up length, incarceration time, and disguised censorship owing to death, then we will have considered it a success. Our appeal applies not just to SPGM, as Nagin rightly observes, but to the users of event history analysis, growth curve models, sequence analysis, and more. Although we agree with Nagin in principle, it was our sense that SPGM, being newer, had simply not been vetted like the other longitudinal methods longer known to criminology. So in a very real sense we agree on first-order issues.

Nagin's response is useful in another way because it forced us to think harder about the unique aspects of SPGM that are good candidates for sensitivity analysis. In order to advance the discussion, we reconsider issues raised in the original paper that may be unique to SPGM, introduce some new concerns inspired by Nagin's logic, and conclude with some thoughts on the larger issue of the value of a group-based approach that Nagin focuses on in his comment. Nagin specifically poses the question: "Why use groups to approximate a continuous distribution of developmental trajectories?" This is indeed a question of considerable significance, and so we return to it below.

Whether a methodological issue like follow-up length or incarceration time is unique to SPGM turns out to be more complex than we first thought. Nagin is certainly right that all longitudinal methods suffer from the "extrapolation" problem. But in reconsidering our findings in light of his comment, we were struck by the specific form that variation took in our data. Our paper concluded that sensitivity was most acute in identifying a chronic, high-rate group of offenders, one that is typically small in the size of its membership. SPGM thus appeared vulnerable to the admittedly generic problem of variable follow-up length. (Recall that our results showed that inferences about the high-rate group differed depending on how you define the follow-up period.) If one peruses the journals of criminology and developmental psychology it becomes immediately apparent that SPGM is being applied on data sets with remarkably different observation windows in terms of age. For example, Nagin and Tremblay (1999) analyze data for 6 to 15 year olds, Bushway and colleagues (2003) focus on 13 to 22 year olds, and Piquero and colleagues (2001) examine data for 18 to 32 year olds. The majority of the 30 published papers we have identified to date use short lengths of follow-up in childhood and adolescence. If our results are any guide, it follows that premature calls are probably being made about the existence of a *stable* group of high-rate chronics.

Why does this matter? We believe the potential for danger stems not from the estimating method itself but from how it is interpreted. Through no fault of the method's designers, SPGM research is particularly vulnerable to

misappropriation by those pre-disposed to believe in the idea of a high-rate, chronic, life-course persister, superpredator group—especially policy makers who seek to intervene. As Travis Hirschi has remarked: “Typologizing is the first recourse of the treater” (2000). And indeed the extraction of chronic high-rate offender groups in SPGM has drawn considerable attention in policy circles, underscoring the connection between the search for groups and interventions. Although beyond the scope of this comment, numerous examples linking a high-rate offender group to interventions can be found in the literature.⁴ Because other longitudinal methods seek to make inferences about individual-level trajectories rather than groups, SPGM thus appears to bear a somewhat greater (albeit not unique) burden. The reason, as Nagin points out, is that methods like HLM do not start with the “a priori” assumption that there are discrete groups. This is a complex issue but we do see a methodological value in raising it, for if this discussion has the effect of sensitizing researchers and policy makers to the perils of making premature classifications that can have deleterious effects on people’s lives, we again consider it a success.

Following Nagin’s logic, we turn to an issue that *is* more clearly in the unique camp of SPGM and that we did not address in the original article. Namely, consider the variation in the number of groups that are estimated as a function of sample size. To get a handle on this issue, we used the same data in our article but selected random sub-samples of differing sizes and conducted new SPGM analyses for the purposes of this comment. Because these are the same men and random draws, we can safely assume that there are no material differences in underlying trajectory groups. To test sensitivity, we then estimated groups by applying the same objective criteria—the Bayesian Information Criteria (BIC) statistic. We varied the sample size from a low of 25 men to a high of almost 500 in the total delinquent group sample (480). For total crime the results yielded anywhere from 5 to 8 groups, with the 8-group models derived only for sample sizes of 200 and higher. The “bad” news is thus that the number of groups extracted is variable and partly a function of sample size—the more individuals the more groups one will find (see also Nagin and Tremblay, 2001, p. 30). This result is not surprising if indeed the underlying distribution is more or less continuous in nature.

On the other hand, the good news for SPGM is that the number of groups appeared to plateau at about sample size 200, suggesting a stabilizing of results. D’Unger and colleagues (1998) also found that the optimal number of groups did not vary with respect to sample size above this range; specifically, they found invariance in group number across three data sets

⁴For an extended discussion see Laub and Sampson (2003, chapter 10).

ranging from about 500 to 2000 subjects. Overall, our sense is that the sample size problem is modest and not likely to fundamentally alter the descriptive picture of criminal trajectories. We set aside the question of whether causal inferences about the effects of covariates are likely to be robust to assumptions about the number of groups (see also the related discussion in Bauer and Curran, 2003).

Another question that we would like to raise concerns the use of SPGM in accelerated longitudinal designs that are becoming more common in criminology. The National Youth Survey and the Project on Human Development in Chicago Neighborhoods, for example, both use sequential cohort designs that permit linking across cohorts. It is not immediately clear to us how to best estimate group-based trajectories in this type of research design. An example of using hierarchical linear models to estimate an age-crime curve using data on multiple cohorts is found in Raudenbush *et al.* (2003). Work in progress is extending this model to estimate period effects that can be cleanly separated from aging in estimating the effect of covariates in the study of change. In the context of developmental change following one cohort, the confounding of aging and history is, of course, ever present. Interestingly, however, in the SPGM literature trajectories are almost always referred to in developmental terms when history could have equally generated the differential patterns. Why is this so? It is our sense that many users of group-based developmental trajectory models, like developmental criminologists in general, are ahistorical, automatically attributing the unfolding of trajectories to developmental (aging) effects rather than historical change. For us this means that greater attention to age, period, *and* cohort is in order. We raise this concern not as a critique of SPGM, but as a pointer to a research agenda that the method might fruitfully be applied in future work, and as a call for greater clarity among users about causal interpretation. In particular, we would like to see the explicit application of group-based methodology to accelerated longitudinal designs.

We would also like to clarify some of the misunderstanding that has arisen in the criminological literature concerning the analysis of criminal trajectories. Some who advocate the use of SPGM have claimed that it is superior to hierarchical linear models because the latter are forced to assume “one” trajectory in the population and are subject to strict parametric assumptions (e.g., Bushway *et al.*, 2001, p. 501, note 12; p. 503). In point of fact, with the estimation of robust standard errors and population average models, HLM provides quite flexible methods for analyzing non-normal data without the imposition of strict parametric assumptions (Raudenbush and Bryk, 2002, pp. 276–280). Furthermore, HLM provides highly general strategies for estimating age-crime trajectories for different groups defined by observed covariates, and for allowing random effects that

reflect heterogeneity in the variance of age slopes. For example, one can easily specify age interactions with say, race or sex, to test whether or not there are different trajectory patterns through time. One need not, as has been claimed, assume one age-crime curve for all individuals; indeed each person has his or her own trajectory that can be estimated by empirical Bayes methods. (For a useful comparison of statistical models for analyzing personal trajectories, see Raudenbush, 2001.) We make these clarifications not to claim superiority for HLM, for obviously any decision on appropriate methods should come down to matching to the theoretical goal.

This brings us to perhaps the biggest issue that Nagin raises in his comment—the validity of grouping methods in general and what we see as a question about the very idea of groups. He asks, why use groups? We have approached this issue in recent work and refer the reader to our efforts for in-depth treatment (Laub and Sampson, 2003; Sampson and Laub, 2003). In this comment, we wish to highlight a distinction that Nagin himself seems to make, and that is the distinction between description and explanation. We agree that SPGM allows a very flexible method for exploring discontinuous patterns in longitudinal data, much in the same way that cross-sectional clustering techniques discover clumping of cases. We would further note the useful role of clustering methods and SPGM for the generation of hypotheses that can be tested on other data sets.

For us, the larger question concerns the causal theory of groups. We have yet to be convinced that offender groupings are meaningful in the sense of social ontology (see Sampson and Laub, 2003). The widely noted “life-course persister” group, for example, may be useful in terms of a heuristic device, but in terms of theoretical validity as a distinctive and replicable group that has decisive implications for etiological theory, we believe the evidence is rather thin. (It is interesting to note that SPGM research has by now firmly rejected the notion that there are only two groups of offenders, with the apparent result that life-course persisters have become subdivided into multiple chronic-offender groups.) One catch in all this is that SPGM begins with the methodological assumption that groups exist, often leading to the notion that a wide array of group configurations is possible. It is then easy for the naïve user to conclude (tautologically?) that groups exist because they are discovered, even though a model cannot be said to discover what it assumes. SPGM will estimate groups from an underlying continuous distribution, a fact that can bedevil even the most sophisticated user.

The solution, in our view, is to rely on strong theory. There is a long criminological tradition that does not posit the existence of materially different and prospectively valid groups. Based on considerable research, that is the view to which we now hold (Laub and Sampson, 2003). On the other hand, typological theories have a long history too, and examining them with

group-based methodology is a logical choice. Our point is simply that absent a theory of crime, the interpretation of any evidence is ambiguous. So whether HLM or SPGM, methods are just that—methods in service of explanation and understanding. HLM and SPGM should also not be viewed as competing methods, for they each have unique strengths and capacities.

The field of criminology stands to benefit from assessing the sensitivity and robustness of *all* methods, a stance on which we have reached agreement. We are hopeful this exchange hastens the process not only of improving methodology, but of linking methods more tightly with criminological theory.

REFERENCES

- Bauer, D. J., and Curran, P. (2003). Distributional assumptions of growth mixture models: Implications for over-extraction of latent trajectory classes. *Psychol. Meth.* 8: 338–363.
- Bushway, S., Piquero, A., Broidy, L., Cauffman, E., and Mazerolle, P. (2001). An empirical framework for studying desistance as a process. *Criminology* 39: 491–516.
- Bushway, S., Thornberry, T., and Krohn, M. (2003). Desistance as a developmental process: A comparison of static and dynamic approaches. *J. Quant. Criminol.* 19: 129–153.
- D’Unger, A. V., Land, K. C., McCall, P. L., and Nagin, D. S. (1998). How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses. *Am. J. Sociol.* 103: 1593–1630.
- Hirschi, T. (2000). Personal Interview by John Laub, Tucson, Arizona.
- Laub, J. H., and Sampson, R. J. (2003). *Shared Beginnings, Divergent Lives: Delinquent Boys To Age 70*. Harvard University Press, Cambridge, Massachusetts.
- Nagin, D. S., and Tremblay, R. E. (1999). Trajectories of boys’ physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Dev.* 70: 1181–1196.
- Nagin, D. S., and Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychol. Meth.* 6: 18–34.
- Piquero, A. R., Blumstein, A., Brame, R., Haapanen, R., Mulvey, E. P., and Nagin, D. S. (2001). Assessing the impact of exposure time and incapacitation on longitudinal trajectories of criminal offending. *J. Adolesc. Res.* 16: 54–74.
- Raudenbush, S. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Ann. Rev. Psychol.* 52: 501–525.
- Raudenbush, S., and Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, Thousand Oaks, California.
- Raudenbush, S., Johnson, C., and Sampson, R. J. (2003). A multivariate, multilevel rasch model with application to self-reported criminal behavior. *Sociol. Methodol.*, in press.
- Sampson, R. J., and Laub, J. H. (2003). Life-course desisters? Trajectories of crime among delinquent boys followed to age 70. *Criminology*, in press.