

ABSTRACT

Title: CAUSAL INFERENCE WITH GROUP-BASED TRAJECTORIES AND PROPENSITY SCORE MATCHING: IS HIGH SCHOOL DROPOUT A TURNING POINT?

Gary Allen Sweeten, Ph.D., 2006

Dissertation directed by: Professor Shawn D. Bushway
Department of Criminology and Criminal Justice

Life course criminology focuses on trajectories of deviant or criminal behavior punctuated by turning point events that redirect trajectories onto a different path. There is no consensus in the field on how to measure turning points. In this study I ask: Is high school dropout a turning point in offending trajectories? I utilize two kinds of matching methods to answer this question: matching based on semi-parametric group-based trajectory models, and propensity score matching. These methods are ideally suited to measure turning points because they explicitly model counterfactual outcomes which can be used to estimate the effect of turning point events over time.

It has been suggested that dropout is the end result of a process of disengagement from school. In order to assess the effect of the event of dropout, it is necessary to separate dropout from the processes that lead to it. The extent to which this is accomplished by matching is assessed by comparing dropouts to matched non-dropouts

on numerous background characteristics. As such, it is desirable to use a wide range of measures to compare the two groups.

I use the National Longitudinal Survey of Youth 1997 to address this question. Delinquency is measured in two ways: a six-item variety scale and a scale based on a graded-response model. Dropout is based on self-reports of educational attainment supplemented with official transcripts provided by high schools. Because of the breadth of topics covered in this survey, it is very well-suited to matching methods. The richness of these data allows comparisons on over 300 characteristics to assess whether the assumptions of matching methods are plausible.

I find that matching based on trajectory models is unable to achieve balance in pre-dropout characteristics between dropouts and non-dropouts. Propensity score matching successfully achieves balance, but dropout effects are indistinguishable from zero. I conclude that first-time dropout between the ages of 16 and 18 is not a turning point in offending trajectories. Implications for life course criminology and dropout research are discussed.

CAUSAL INFERENCE WITH GROUP-BASED TRAJECTORIES AND
PROPENSITY SCORE MATCHING: IS HIGH SCHOOL DROPOUT A
TURNING POINT?

By

Gary Allen Sweeten

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

Professor Shawn Bushway, Chair
Professor Denise Gottfredson
Professor John H. Laub
Professor Raymond Paternoster
Professor Seth Sanders

© Copyright by
Gary Allen Sweeten
2006

Acknowledgements

While I only have an N of one, I consider my graduate education at the University of Maryland, and in particular, my exposure to Shawn Bushway, as a turning point in my life. The counterfactual outcome is quite clear in my mind: I would now be a juvenile probation officer, or perhaps an analyst for a police department somewhere on the West coast. While this would not have been an entirely abhorrent path to take, I expect that a future in academia will be more fulfilling, much more interesting, and a better use of my talents. Shawn convinced me of the merits of the academic path, and as a mentor, employer, drill sergeant, advisor, role model, and friend, he has guided me down the first part. I would not be in my current position, and this document would not exist, were it not for his influence.

Every member of my dissertation committee and many other University of Maryland professors have influenced the way I approach criminological research. John Laub, in particular, has been a strong influence. He introduced me to the world of criminological theory, of examining and challenging theoretical assumptions, of careful word choice, and of revising revisions of revisions. I am also indebted to my statistics instructors: Laura Dugan, who introduced me to the world of regression; Doug Smith, who displayed a clarity of mind to which I constantly aspire; and Jeff Smith, whose insightfulness and humor made applied econometrics enjoyable.

Finally, I am deeply grateful to my wife, Erin, for her love, support, forbearance, and numerous hours of editing over the past five years.

Table of Contents

Acknowledgements	ii
List of tables	v
List of figures	vi
Chapter 1: Introduction	1
Statistical methods for identifying turning points	5
Interpreting matching models	8
Summary	10
Chapter 2: The Life Course Perspective and High School Dropout	12
The life course perspective	12
Turning points: Theoretical perspectives	15
Dropout and offending	17
Chapter 3: Data and Methods	21
Data	21
Sample selection	22
Measuring offending	24
Measuring dropout	27
Modeling strategies	33
Semi-parametric group-based trajectory models	36
Matching methods	40
Strategy 1: Balancing on posterior probabilities of group assignment	42
Strategy 2: Propensity score matching	49
Summary	55
Chapter 4: Results	58
Baseline comparisons	58
Delinquency trajectories through age 16	59
Tests of balance	61
Dropout effect estimates	66
Propensity score matching	70
Reductions in standardized bias	73
Conclusion	75
Chapter 5: Conclusion	78
Theoretical and research implications	79
High school dropout	79
Trajectory model matching and turning points	81
Propensity score matching	82
Item response theory	84

Policy implications	85
Future research	87
Appendix A. Variables used to compare dropouts and non-dropouts	108
Appendix B. Unadjusted and adjusted standardized bias (variety scores)	118
Appendix C. Unadjusted and adjusted standardized bias (IRT scores)	125
References	132

Tables

1. Characteristics of final sample and excluded observations	90
2. Mean delinquency variety and IRT scores	91
3. Baseline comparison of dropouts and non-dropouts on offending between the ages of 16 and 18 (T_1), and each of three waves thereafter (T_2 to T_4)	92
4. Characteristics of four groups in delinquency variety trajectories	93
5. Characteristics of four groups in delinquency IRT score trajectories	94
6. Within-trajectory balance on prior delinquency and selected high-bias background characteristics, classify-analyze method	95
7. Within-trajectory balance on prior delinquency and selected high-bias background characteristics, expected value method	96
8. Balance on prior delinquency and selected high-bias background characteristics, kernel matching with trajectory probabilities	97
9. Group-average treatment effects (GATE) and population-average treatment effects on the treated (PATT), variety score models	98
10. Group-average treatment effects (GATE) and population-average treatment effects on the treated (PATT), IRT models	99
11. Frequency of propensity scores by dropout status	100
12. Balance on prior delinquency and selected high-bias background Characteristics, propensity score method	101
13. Propensity score estimates of dropout effect (PATT), variety score models	102
14. Percent bias reduction from matching methods for most biased characteristics	103

Figures

1. Delinquency variety trajectories, ages 11-15.5	105
2. Delinquency IRT score trajectories	106
3. Distribution of propensity scores by dropout status	107

CHAPTER 1: INTRODUCTION

In my sophomore year of college, I discovered that one of my closest childhood friends had been sentenced to seven years in prison for bank robbery. He had, in fact, been my best friend during five formative years: from about age 8 to age 13. We shared houses, advanced through scouts together, played sports, fought each other, skipped school together, and generally encouraged each other's misbehavior. Together, we were becoming increasingly involved in delinquency. Had his family not moved to the other side of town, we would have gone to high school together, and most likely would have remained friends. I sometimes wonder: What caused our lives to diverge so dramatically? Did he experience some detrimental life event which I managed to avoid? Would my life resemble his had I experienced the same things?

These personal questions are specific cases of more general questions posed by the life course perspective, which provides a framework for studying lives over time. This perspective focuses on the interdependence of multiple life trajectories, such as educational attainment, offending, drug use, employment and family makeup. Life transitions are embedded within trajectories, and give the trajectories meaning (Elder, 1985). When a life transition causes an individual's trajectory to shift to another path, it is identified as a "turning point" (Rutter, 1987; Abbott, 1997).

In this study, I ask whether high school dropout is a turning point in trajectories of criminal propensity. High school dropouts, while representing less than 15 percent of 18 to 24 years olds in the United States (U. S. Department of Education, 2001), comprise over two-thirds of state prison inmates (The Sentencing Project, 2004). In addition, a high school dropout can expect to earn 20 percent less than a high school graduate (with

no college), and over 50 percent less than a college graduate over the course of a working career (U. S. Census Bureau, 2002). Not surprisingly, dropouts are more likely to receive public assistance as well (U.S. Department of Education, 2001). When considered in the aggregate, the 3.8 million high school dropouts aged 16- through 24 in the United States (U. S. Department of Education, 2001) represent lost productivity and a net drain on public funds. For certain youth, lack of a high school diploma may result in limited employment opportunities, increased criminal offending, and other negative outcomes. If these youths on the margins would be set on a different life trajectory with a diploma, then dropout may be considered a turning point. In this dissertation, I assess whether first-time dropout between the ages of 16 and 18 is a turning point by measuring differences between dropouts and matched non-dropouts between the ages of 18 and 22.

Many recent dropout studies suggest that high school dropout is best understood as a process rather than an event. Risk factors for high school dropout can be identified very early in the school career (Alexander, Entwisle, and Kabbani, 2001; Ensminger and Slusarick, 1992). Finn (1989) suggests that dropout may be the end result of a process of disengagement from school. Because of this process, dropouts and non-dropouts exhibit marked differences in numerous domains long before and long after the event of dropout. There is no question that the process of disengagement from school, which often ends in high school dropout, is an important topic of study. Policy efforts that effectively interrupt early disengagement from school can have numerous long-term positive outcomes (e.g., Schweinhart, Montie, Xiang, Barnett, Belfield, and Nores, 2005). However, high school dropout itself may have an additional effect on future behavior independent of the processes that lead to dropout. In essence, this study seeks to measure

the effect of dropout by treating it as an event, and controlling for the process of disengagement from school.

Not only are there strong selection processes affecting who drops out of high school, but there is some evidence that a bi-directional relationship exists between delinquency and educational attainment. A few studies claim to uncover a causal effect of dropout on subsequent crime (Thornberry, Moore, and Christenson, 1985; Jarjoura, 1993, 1996), but others find no aggregate effect (Krohn, Thornberry, Collins-Hall, and Lizotte, 1995; Sweeten, Bushway, and Paternoster, 2005). Looking at the other direction of causation, most studies find that justice system involvement is detrimental to education outcomes (Bernburg and Krohn, 2003; De Li, 1999; Sweeten, 2006). The “treatment” of dropout does not arrive at random. Rather, individuals self-select into dropout, as they do into offending. Individuals with high criminality may be more prone to drop out. In addition, the same individual characteristics may account for both dropout and offending. Thus estimates of the effect of dropout on offending may be contaminated by the effect of earlier offending and official reactions to said behavior on dropout itself. Any assessment of the effect of dropout on offending, in order to be valid, must account for endogeneity in the dropout-delinquency relationship. In this study, I seek not only to assess the effect of dropout on delinquency, but to determine whether dropout is a turning point in pathways of delinquency. The analysis must not only address issues of endogeneity, but must allow an assessment of whether dropout is a turning point.

While the concept of a turning point is fairly straightforward, methods for identifying one are not. Qualitative analysis uses self-identified turning points to identify classes of events which are recognized by individuals as turning points (Giordano,

Cernkovich, and Holland, 2003; Laub and Sampson, 2003; Rönkä et al., 2002).

Quantitative analysis of turning points begins with longitudinal data, but uses a number of different statistical techniques: fixed effect analysis (Paternoster, Bushway, Brame, and Apel, 2003; Sweeten, 2004), hierarchical growth curves (a type of fixed effect analysis) (Hoffmann and Cerbone, 1999; Laub and Sampson, 2003), semi parametric group-based trajectory models (Laub, Nagin, and Sampson, 1998; Nagin, Pagani, Tremblay, and Vitaro, 2003), standard regression models (Warr, 1998; Wright and Cullen, 2004), and survival analysis (Uggen, 2000). Ideally, the method used to analyze turning points should be consistent with the definition of a turning point. Abbott (1997) points out that individual life trajectories are relatively stable, but turning points cause individuals to “leap to a new steady trajectory” (92). Similarly, Rutter (1987) describes positive turning points which “may change the life course onto a more adaptive trajectory” (328). An event is a turning point if it causes a lasting change on a life trajectory. This simple definition requires careful consideration of counterfactual outcomes.

For each individual, one can imagine two alternate outcomes: offending under treatment (dropout), and offending without treatment (non-dropout). The term “treatment” is borrowed from program evaluation literature in which treatment is some form of intentional intervention. While dropout is endogenous to individual characteristics, it can still be thought of as a treatment. It is simply a treatment that does not arrive at random. The goal of this analysis is to measure the “treatment effect” of dropout, which is simply the difference between an individual’s outcome as a dropout and his or her outcome as a graduate. If this difference persists over time, dropout can be

considered a turning point in trajectories of offending propensity. In this case, dropout causes the individual to jump to a different trajectory. The fundamental problem of causal inference is that for each individual, only one of these two outcomes is observed, making direct estimation of the causal effect impossible. Either statistically approximated (using an experiment or other statistical technique), or constructed from personal judgment (in the case of self-identified turning points), one must compare the actual outcome to the unobserved counterfactual outcome (what would have happened in the absence of treatment).

STATISTICAL METHODS FOR IDENTIFYING TURNING POINTS

Prior turning point studies have used many different methods to approximate counterfactual outcomes. Qualitative studies allow respondents to assess whether counterfactual outcomes would differ from observed outcomes. Fixed effect analysis compares the individual to him- or herself before the potential turning point. While this is effective in eliminating bias due to stable between-individual differences, it is not an appropriate method for identifying turning points. Fixed effect analysis is very sensitive to the time period used and it estimates the effect of a potential turning point using only individuals who actually experience the event. Furthermore, fixed effect analysis does not allow assessment of the timing of an effect. Standard regression models estimate counterfactual outcomes by holding all observed characteristics constant, sometimes relying heavily on functional form assumptions to do so.

In this dissertation, I use matching methods to determine whether dropout between the ages of 16 and 18 is a turning point. Matching methods estimate causal

effects by determining which of the non-treated individuals can serve as counterfactuals for the treated sample. The first method I employ, introduced by Haviland and Nagin (2005), is matching based on pre-dropout trajectories of offending. A semi-parametric group-based trajectory model is estimated for those individuals who do not drop out prior to age 16. For each estimated trajectory and each individual, a probability of group assignment is estimated based on observed patterns of delinquency. These posterior probabilities of group assignment are then used in a variety of ways to match dropouts to non-dropouts and obtain estimates of the effect of dropout for up to four years after age 18. The second matching method I use—propensity score matching—has a relatively longer history (Rosenbaum and Rubin, 1983, 1985). In this method, a model predicting treatment is estimated using all relevant observed characteristics in order to obtain the estimated propensity to receive treatment (in this case, to drop out of high school) for each individual. This metric is then used to match dropouts to non-dropouts with similar propensity scores.

Both of these matching methods present considerable advantages over previous statistical methods for identifying turning points. First, they allow direct observation of whether dropouts and non-dropouts differ so much on pre-existing characteristics that meaningful comparisons cannot be made. In standard regression models, an estimate of dropout would be calculated regardless of the lack of comparability between dropouts and non-dropouts. With matching methods, it may be determined that no matches can be found for a certain portion of the treated sample. Second, these methods allow for an easy assessment of whether pre-existing differences are controlled conditional on matching. Background characteristics are compared between dropouts and matched non-

dropout counterparts to determine whether matching succeeds in creating a comparison sample that differs only on treatment status. Third, these methods provide a clear interpretation of the timing of the turning point. In this research, only characteristics prior to age 16 are used to estimate balancing scores. Dropout is then assessed between the ages of 16 and 18, and the effect of dropout is measured after age 18 for up to four years. Because of endogeneity concerns, youths who drop out prior to age 16 are eliminated from the sample. Thus, the estimated dropout parameters refer only to dropout between the ages of 16 and 18. The life course framework holds that timing of life events is a key determinant of their effect on life trajectories (Elder, 1998; Nagin et al., 2003). It is reasonable, therefore, to assess the effect of dropout over a small range of ages.

Matching based on pre-16 delinquency trajectories provides the additional benefit of estimating heterogeneous effects of dropout. Within each estimated trajectory group, dropouts and non-dropouts are compared in order to obtain a group-average treatment effect which represents the effect of dropout conditional on membership in the trajectory group. Life course research suggests that individual adaptations to potential turning points may depend on developmental history (Elder, 1985, 1998). Similarly, high school dropout research has focused on different “types” of dropout characterized by different individual characteristics or different contexts. Past research has classified dropouts based on stated reason for dropout (Jarjoura, 1993, 1996), different causal pathways to dropout (Battin-Pearson, Newcomb, Abbot, Hill, Catalano, and Hawkins, 2000), socio-economic status (Thornberry et al., 1985; Jarjoura, 1996) or a configuration of characteristics determined by cluster analysis (Cairns, Cairns, and Neckerman, 1989;

Janosz and LeBlanc, 2000). Furthermore, several dropout studies have uncovered different effects of dropout on offending depending on the stated reason for dropout (Jarjoura, 1993, 1996; Sweeten, 2004), socio-economic status (Thornberry et al., 1985; Jarjoura, 1996) and race (Thornberry et al., 1985; Voelkl, Welte, and Wieczorek, 1999). Estimation of the effect of dropout within groups defined by developmental histories of offending is a valuable exercise because it allows researchers to associate treatment effects with substantively interesting groups (Haviland and Nagin, 2005).

INTERPRETING MATCHING MODELS

Because dropout can be understood as the end result of a long process of disengagement from school, it is essential to assess the extent to which pre-existing differences between dropouts and non-dropouts are eliminated by matching. As such, interpretation of the results of this analysis will focus on three separate elements: 1) common support (i.e., are there non-dropouts who are comparable to dropouts?), 2) pre-dropout differences between dropouts and non-dropouts, and 3) post dropout differences between dropouts and non-dropouts.

This study relies on matching methods to compare dropouts to non-dropouts who appear similar on observed covariates. The first matching method uses patterns of delinquency prior to dropout, and the second uses a propensity score for dropout calculated from all observed characteristics believed to be associated with either dropout or delinquency. Lack of adequate matches for dropouts is evidence that the process of disengagement from school is indeed more important than dropout itself for a sub-group of dropouts. Matching methods require individuals with like characteristics to have

different treatment outcomes (dropout or non-dropout in this case). Lack of matches, in this context, shows that given a certain configuration of risk factors, or indicators of disengagement from school, dropout is inevitable.

On the other hand, if there is common support, meaning that for every dropout, there is at least one non-dropout who appears similar enough to serve as the individual counterfactual, then measurement issues become a great concern. With propensity score matching, common support would mean that there are individuals in the survey who exhibit a constellation of risk factors for dropout, yet do not drop out. The danger is that youths with high estimated propensities to drop out may falsely report completion of school, invalidating the results of the analysis. In order to guard against this possibility, self-reported dropout measures are augmented with official reports drawn from high school transcripts. This measurement issue is addressed in detail in Chapter 3.

Given common support, the dropout effect is measured by contrasting delinquency outcomes of high school dropouts to matched non-dropouts. In the same manner, prior delinquency and other background characteristics are compared between dropouts and matched non-dropouts. If prior delinquency and other background characteristics differ between the two groups even after matching, then it must be concluded that matching did not succeed in eliminating pre-dropout differences. In this case, post-dropout differences cannot be interpreted as the effect of the event of dropout, but must be interpreted as the effect of both dropout and other pre-existing differences, including, but not limited to, disengagement from school. If, on the other hand, there are little or no discernable pre-dropout differences between dropouts and non-dropouts after matching, then post-dropout differences between the two groups can be interpreted as the

effect of the dropout event, holding constant pre-dropout differences such as disengagement from school. If there is common support, balance on pre-existing covariates, and significant differences between dropouts and non-dropouts over the course of four years following dropout, then there is strong evidence that dropout is a turning point.

SUMMARY

This research provides a number of valuable contributions. First, it contributes to life course criminology by presenting a method to systematically assess long-term effects of life transitions in order to identify turning points. I contend that, barring experimental data, the methods I present for identifying turning points are the most appropriate and the most consistent with the definition of a turning point. Second, I extend the application of matching models based on group-based trajectories, first presented by Haviland and Nagin in 2005, to other evaluation problems, and I link the model with turning point studies. In Chapter 5, I suggest some guidelines for when to use this matching method rather than propensity score matching. I also contribute to the literature on the effect of dropout on delinquency by placing the problem within the life course context, and assessing whether the effect of dropout endures over time, and whether it differs by patterns of delinquency prior to dropout. Finally, I conduct all analyses using variety scores of delinquency and delinquency scales based on a graded response model (Samejima, 1969; Osgood, McMorris, and Potenza, 2002), an extension of item response theory (IRT). The graded response model is used in an attempt to create a scale of offending that reflects both seriousness and frequency of offending rather than just one or

the other. If the graded response model successfully generates a scale that more closely approximates true latent delinquency scores than variety scores, then balance on pre-existing characteristics related to latent delinquency should be greater using IRT scales.

In Chapter 2, I provide some background on the life course perspective with particular focus on turning points, and I review the literature on the effect of dropout on offending. Chapter 3 details the data that will be employed in this study, discusses measurement issues for dropout and delinquency, and describes in greater detail the modeling strategies that will be employed. Chapter 4 presents the results of the analysis, and Chapter 5 contains a discussion of the results.

CHAPTER 2: THE LIFE COURSE PERSPECTIVE AND HIGH SCHOOL DROPOUT

THE LIFE COURSE PERSPECTIVE

Application of the life course perspective to criminological theory and research has resulted in numerous insights into patterns of offending over the life course. The life course perspective has roots in longitudinal research conducted in the 1920s and 1930s, but the conceptual framework has gained more clarity in recent decades (Elder, 1985). Elder (1994) characterizes the life course as a series of interrelated trajectories and transitions. Trajectories are long-term behavioral patterns whereas transitions are short-term changes embedded in trajectories, giving them form and meaning. Transitions that result in altering a trajectory are considered turning points (Laub and Sampson, 1993; Sampson and Laub, 1993).

In addition to a focus on trajectories and transitions, four principles guide life course research. First, human lives are embedded in, and influenced by historical time and place. Consideration of historical time and place focuses attention on external validity of regional studies, and on variation in effects in national studies and over time. Second, the life course perspective focuses on timing of life transitions. That is, the effect of any particular transition is presumed to vary depending on the age at which the transition occurs. For example, there is a growing literature concerned with the effects of employment during adolescence, which is thought to differ from the effects of employment during adulthood (Staff and Uggen, 2003). Third, the life course perspective focuses on linked lives. It recognizes between-individual interdependence of

life trajectories. Finally, individuals are endowed with human agency, meaning that they take an active role in shaping their own lives within the constraints of opportunity and structure (Elder, 1998).

Several enduring issues have emerged from the debate surrounding life-course and developmental theories of crime. There is continuing debate over the extent to which changes in developmental trajectories are due to selection bias. This debate has been framed in a number of ways: ontogenetic vs. sociogenic models (Dannefer, 1984), static vs. dynamic explanations (Paternoster, Dean, Piquero, Mazzerolle, and Brame, 1997), or non-causal vs. causal life events (Gottfredson and Hirschi, 1990; Sampson and Laub, 1993). The key distinction between these two positions is their stance on the etiological significance of life events after some particular age identified by ontogenetic theories. Ontogenetic theories hold that crime and delinquency are determined by events or characteristics evident early in life; events or processes which occur after a certain time in life (usually early childhood) are considered non-causal. Sociogenic theories, on the other hand, contend that later life events can have a causal impact on offending. With respect to the relationship between high school dropout and subsequent offending, ontogenetic theories would propose that any apparent relationship between the two is spurious, both influenced by common factors. On the other hand, a sociogenic explanation would allow for a causal relationship between the two in either direction. Turning points, because they mark discontinuity in a life trajectory which cannot be explained by pre-existing factors, are consistent with sociogenic explanations.

Another ongoing debate in criminology concerns the extent of continuity and change in offending trajectories. This literature is fueled, in part, by Lee Robins' (1978)

observation that although serious adult offending is almost always preceded by juvenile offending, most juvenile delinquents do not become adult offenders. Consistent with this observation, a sizable body of literature has found evidence of continuity in offending across the life course, and a sizeable body of research has found evidence of considerable change in offending over the life course.

Several criminological theories speak to the issue of continuity and change in offending, with widely divergent claims. Gottfredson and Hirschi (1990) employ a population heterogeneity argument, claiming that all offending is explained by opportunities to offend and stable (after age 8) individual differences in self-control. In sharp contrast to Gottfredson and Hirschi, Sampson and Laub (1993) suggest that social bonds throughout the life course can divert criminal pathways towards desistance or desisting pathways towards more offending. Moffitt (1993, 1994) posits a typological theory of offending wherein “life-course persistent” offenders evince continuity and “adolescent limited” offenders, change.

These three prominent criminological theories differ markedly in their positions on turning points. Gottfredson and Hirschi (1990) do not address this issue, but presumably they would confine causal turning points to those that affect the latent trait of self-control before age eight. Sampson and Laub (1993; Laub and Sampson, 2003) give prominent place to turning points in their age-graded theory of informal social control, recognizing both positive and negative turning points, and positing mechanisms of change following turning points. Moffitt (1993, 1994) incorporates a reciprocal relationship into her typological theory of crime, suggesting that certain consequences of

antisocial behavior, called “snares,” can restrict opportunities for change, leading to continuity in antisocial behavior, particularly among life-course persistent offenders.

Gottfredson and Hirschi’s (1990) theory contends that any apparent casual relationship between dropout and offending is spurious as both are caused by low self-control. Laub and Sampson’s (2003; Sampson and Laub, 1993) age-graded theory suggests that the effect of dropout will depend on the extent to which social bonds are changed, and the way in which individual routine activities are altered. The most probable outcome is that both social bonds and structured routine activities would decrease, leading to a decrease in social control, and higher likelihood of offending. Moffitt’s (1993, 1994) typological theory characterizes high school dropout as a snare which serves to maintain continuity in offending, particularly among life-course persisters. For this group, the snare of dropout would not be considered a turning point because offending is continuous before and after dropout. However, it is possible that the snare of dropout could divert adolescent-limited offenders to more persistent offending although they are generally more resilient to the negative effects of snares. Moffitt’s theory suggests that dropout may cause continuity in offending among serious offenders, and potential increases in offending among less serious offenders.

TURNING POINTS: THEORETICAL PERSPECTIVES

A turning point results in a change in direction of a developmental trajectory (Abbott, 1997). For the most part, however, trajectories, are distinctive for their continuity, “their quality of enduring large amounts of minor variation without any appreciable change in overall direction” (Abbott, 1997:93). Whether or not a particular

life event becomes a turning point depends on the response of the individual (Elder, 1985; Sampson and Laub, 1993; Rutter and Rutter, 1993; Rutter, 1996). In fact, the same event can have positive outcomes, negative outcomes, or no effect depending on how the individual adapts to the event. Individual differences in responses to life events may depend on gender, social circumstances (Rutter, 1996), pre-existing differences in neuropsychological deficits (Moffitt, 1994), the co-occurrence of purposeful human agency (Laub and Sampson, 2003), or other pre-existing individual differences. The life course perspective highlights the importance of considering whether certain subgroups of people experience turning points differently.

Elder (1985) points out that the concept of stable trajectories punctuated by transitions calls attention to the question of duration. How long does it take to transition from one trajectory to another? Definitions have varied with respect to the duration of turning points (Elder, 1985). Most suggest that a turning point can be either an event or a process (Mandelbaum, 1973; Sampson and Laub, 1993, 1997); some claim that a turning point is a process which can take many years to complete (Hareven and Masaoka, 1988), while others recognize turning points as events with finite duration (Abbott, 1997). Turning points take a certain amount of time, and in order for them to be considered meaningful transitions, their duration must be short relative to the trajectories they punctuate. To some extent, the question of duration of turning points is one of scale. Criminological researchers can study individual and social processes from a phenomenological perspective (Katz, 1992), on a month-to-month basis (Horney, Osgood, and Marshall, 1995), or with respect to the entire life course (Laub and

Sampson, 2003). Each of these approaches may uncover turning points of varying duration, relative to the time-scale of their choosing.

Duration of the turning point is of particular concern for the study of dropout. Much of the literature on high school dropout describes a process of withdrawal or detachment from school (Alexander et al., 2001; Ensminger and Slusarick, 1992; Finn, 1989). In fact, some studies have found that prediction of high school dropout is as strong using early grade school measurements as it is using measures much more proximal to the dropout event (Alexander et al., 2001). While it is true that early childhood characteristics and family background characteristics are strong predictors of dropout, they are not perfect predictors. Not all high-risk youth drop out, and not all low-risk youth graduate. Although dropout may be strongly predicted by early risk factors, this does not require that dropout itself cannot have a causal impact on offending over and above early risk factors (Sampson and Laub, 1993).

DROPOUT AND OFFENDING

A causal relationship between dropout and offending can be understood from a number of different theoretical perspectives. Early dropout-delinquency studies posited that poor school performance was a source of frustration for low achieving students. This theoretical argument, dubbed the “frustration-self-esteem” model by Finn (1989), suggests that the lack of school success leads to feelings of low self-esteem and pressure to seek relief from the strain. Students could relieve the pressures of low school achievement through either delinquent adaptations or through dropping out. According

to this argument, dropping out of school will relieve strain, reducing pressure to offend, and leading to reduced offending (Cohen, 1955, Cloward and Ohlin, 1960, Finn 1989).

An alternate theoretical explanation for the link between dropout and delinquency, dubbed the “participation-identification” model by Finn (1989), and corresponding to control theory in criminology, suggests that delinquent behavior and higher risk of dropout are explained by low attachment to school. Students with high attachment to school are inhibited from offending because they wish to please their teachers and succeed in school. This explanation is consistent with bidirectional causation between dropout and offending. Strong conventional bonds inhibit anti-social conduct. Thus, behavior which is in violation of conventional norms would tend to weaken the social bond. Criminal activity and drug use can lead to problems in school such as absenteeism, a failure to concentrate, and a lack of concern with school personnel and school activities. Involvement in crime and problem behaviors like drinking and drug use could, therefore, be predicted to increase the risk of dropping out of high school. To the extent that dropout further reduces the conventional bond, more offending could result from the dropout event.

Other theoretical perspectives suggest that dropout and delinquency only appear to be related because they are both caused the same factors. Newcomb and Bentler (1988), in their explanation of adolescent drug use, sexual behavior and early entry into the labor force, argue that these acts are all the outcomes of what they call “precocious development”, driven by an inability to defer immediate gratification. They suggest that impulsivity is related to a wide range of adolescent outcomes because it develops early in childhood. Similarly, Jessor and colleagues (Jessor and Jessor 1977; Donovan and Jessor

1985; Donovan, Jessor, and Costa, 1988) hypothesize that the underlying cause of different but related adolescent problem behaviors is unconventionality, which develops early in childhood. Finally, as previously noted, the most prominent criminological theory of this type, Gottfredson and Hirschi's (1990) General Theory of Crime, contends that delinquency, crime, and analogous behaviors are expressions of an underlying, stable trait of low self-control.

Empirical support is mixed for these competing explanations for the link between dropout and delinquency. Each explanation receives some support. Many early studies find that dropout decreases delinquency. Some studies find no effect of dropout on delinquency. Others conclude that the effect of dropout is moderated by another life domain such as employment or family characteristics. In contrast, numerous studies conclude that there is a direct effect of dropout on delinquency.

Consistent with strain theory, several early studies of dropout found that youths who dropped out of high school tended to commit less crime after dropout than before (Elliott 1966; Elliott and Voss 1974; Mukherjee 1971), although the weak analytic models employed likely were driven by the age-crime curve rather than the effect of dropout itself.

Numerous studies have found that dropout puts youth at a greater risk for delinquency or drug use (Polk, Adler, Bazemore, Blake, Cordray, Coventry, Galvin, and Temple, 1981; Voelkl et al., 1999; Crum, Buchulz, Helzer, and Anthony, 1992; Crum, Ensminger, Ro, and McCord, 1993). Others have suggested that only dropout for certain reasons is criminogenic (Jarjoura 1993, 1996). And several studies conclude that post-

dropout outcomes are mediated by other life domains such as employment (Farrington, Gallagher, Morley, Ledger, and West, 1986; Thornberry et al., 1985).

Finally, another set of studies conclude that post-dropout differences in offending are explained by pre-dropout differences. According to these studies, dropout has no causal effect on offending (Bachman and O'Malley, 1978; LeBlanc, Vallieres, and McDuff, 1993; Drapela, 2005). This is also consistent with research that suggests that dropout is the result of a process of disengagement from school which is evident quite early in childhood (Ensminger & Slusarick, 1992; Finn, 1989; Alexander, Entwisle, and Horsey, 1997).

Because support can be found for each explanation of the link between dropout and delinquency, continued efforts at sorting out causality between dropout and delinquency are warranted. This study is designed to determine whether dropout has a causal effect on delinquency and adult offending, whether this effect persists over time, and whether the effect differs depending on prior patterns of delinquency. In addition, because the effect is estimated using matching models that identify potential counterfactuals for dropouts among non-dropouts, if a lasting causal effect is found, then it can be concluded that dropout is a turning point.

CHAPTER 3: DATA AND METHODS

DATA

The analysis will utilize the National Longitudinal Survey of Youth – 1997 cohort (NLSY97). The NLSY97 is a panel survey sponsored by the U.S. Bureau of Labor Statistics. 8984 youths between the ages of 12 and 16 in 1996 were surveyed in the first wave of data collection. Because of oversampling, each participant was assigned a sampling weight, which is used in all analyses in order to ensure that inferences to the national population are not biased by the oversampling. In the first wave, 8984 youths aged 12 to 17 were interviewed. Youths are interviewed every year, with seven waves of data released to date.

In the early 1990s, the Bureau of Labor Statistics changed from paper and pencil surveying (PAPI) to computer assisted personal interviewing (CAPI) and self-administered questionnaires (SAQ) for sensitive items. This led to a decrease in surveyor-induced measurement error, and a slight increase in response rates to sensitive questions in the NLSY97 relative to the NLSY79 (Zagorsky and Gardecki, 1998).

The NLSY97 is particularly well-suited for a propensity score matching analysis because of the breadth of topics covered. Propensity-score matching, which will be described in detail later in this section, relies on observed characteristics to control for differences between dropouts and non-dropouts. As such, unobserved characteristics related to either dropout or delinquency will bias the results. In propensity-score matching all available observed characteristics are utilized to match dropouts to non-dropouts. In both propensity-score matching and matching by delinquency trajectories, observed characteristics of dropouts are compared to those of matched non-dropouts to

determine the success of the matching. If dropouts are statistically indistinguishable from matched non-dropouts, and if no salient characteristics are unobserved, then the two groups are “balanced,” and the only difference between the two groups is dropout itself.

Observed characteristics include gender, race, ethnicity, urbanity, region of the country, numerous socioeconomic-status indicators, household characteristics, numerous family process variables, parental education, health indicators, delinquency, substance use, several peer characteristics, victimization, academic achievement, school attachment, and aptitude, religious affiliation, time use, neighborhood characteristics, local labor market conditions, and employment. A detailed list of every observed background characteristic can be found in Appendix A. Appendix B details which variables are included in propensity score matching, and indicates whether dropouts and non-dropouts are statistically indistinguishable (balanced) on each of the characteristics for each of the matching methods.

SAMPLE SELECTION

Not all of the 8984 individuals in the NLSY97 are included in the analysis. In order to match individuals on prior trajectories of delinquency, at least three pre-16 observations of delinquency are required. This age requirement limits the sample to 3470 individuals. I further require that at least one of the pre-16 observations occur between the ages of 15 and 16. This requirement, which is imposed in order to improve the quality of the delinquency trajectory matches, reduces the sample size to 3387. In order to determine the effect of dropout, there must be some post-16 observations as well. This requirement reduces the sample size to 3177. Furthermore, because I am measuring the

effect of *first-time* dropout between the ages of 16 and 18, all youths who drop out prior to age 16 are eliminated from the sample, even if they had returned to school after dropping out. This brings the sample size down to 2990, which is used for both delinquency trajectory matching and propensity-score matching. The bulk of the sample attrition is due to age restrictions, which do not bias my estimates. However, those individuals who are eliminated from the sample due to early dropout differ significantly from those in the sample on a number of factors (see Table 1). Still, these differences do not bias my estimate of first-time dropout between the ages of 16 and 18 because earlier dropouts are not relevant for the effect being estimated. As shown in Table 1, those who are dropped from the sample because of missing observations tend to exhibit higher levels of risk factors associated with dropout. However, demographic factors do not significantly differ between the groups. Attrition due to non-observation could bias my results because the high school dropouts who remain in the survey may be more stable and exhibit less serious delinquency than those high school dropouts who leave the survey. This would bias the effect estimates downward.

Two measures of offending, described in the following section, are used in this study. The latent delinquency measure requires more information than the variety score measure, and because of this, the sample size for the analysis using latent delinquency scores is 1633. The reasons for this smaller sample size are discussed in the following section.

MEASURING OFFENDING

Measuring offending is a particularly difficult task for a number of reasons. A primary concern is that offending frequency and offending seriousness are difficult to capture in a single scale. For example, when raw offending counts are summed over a number of different kinds of offenses, the resultant scale typically reflects less serious, more frequent offenses. On the other hand, if a simple dichotomous offending variable is constructed, both seriousness and frequency are ignored.

Hindelang, Hirschi and Weis (1981) address various issues of measurement of delinquency in their book *Measuring Delinquency*. Hindelang and colleagues recommend using an “ever variety” scale of offending, which reports the total number of types of delinquent behavior an individual has ever participated in. So, for example, if six different types of delinquent behavior are measured in a survey, a variety scale is constructed, ranging from zero to six, indicating how many of those six offenses the individual committed. They choose this method because it performs the best in terms of validity and reliability. In addition, it seems to best approximate an individual measure of latent offending propensity.

While variety, frequency, and prevalence scales of offending have been the most commonly employed in studies of delinquency and offending, there is a long history of alternate ways to scale offending which attempt to take into account both seriousness and frequency, better representing a latent offending propensity. While numerous early efforts focused on developing a single scale of offending (Guttman, 1950, Nye and Short, 1957, Sellin and Wolfgang, 1964), the criminal careers approach of the 1980s rejected these efforts in favor of etiological studies of every dimension of offending (Blumstein,

Cohen, Roth, and Visher, 1986). The criminal career debate gained volume with the addition of Gottfredson and Hirschi's (1990) General Theory of Crime in which they posited that diverse deviant behaviors are manifestations of the single underlying trait of self-control. These two perspectives entail entirely different methods of measurement, analysis and theory-building.

Osgood and Rowe (1994) responded to the criminal career debate by testing the hypothesis that different dimensions of the criminal career can be explained by different factors. In their first article in a series, they modeled the frequency of offending with an underlying criminal trait (Rowe, Osgood, and Nicewander, 1990). Their latent trait model, which related the observed frequency of offending to the most likely value of the latent trait, proved a good fit for both the observed frequency and prevalence of offending, suggesting that different etiological stories are not needed for each dimension of offending. More recently, Osgood and colleagues published two articles which demonstrated the efficacy of item response theory for creating a coherent unidimensional scale of offending from multi-item categorical offending questions (Osgood, McMorris, and Potenza, 2002; Osgood, Finken, and McMorris, 2002).

Item response theory (IRT) begins with the assumption that some latent dimension, θ , accounts for the observed pattern of responses to multiple items. It is further assumed that θ is distributed normally in the population (Osgood, McMorris, and Potenza, 2002). In addition, IRT suggests that the relationship between the underlying trait and the probability of a positive response to a particular question can be modeled with an item characteristic curve (ICC). Where there are multiple graded responses to a question, such as Likert scales or frequency categories, item characteristic

curves are graphed for each possible response. That is, the probability of each response category is graphed with respect to the underlying trait. IRT analysis proceeds in two steps. First, response patterns are analyzed to score questions on two dimensions: the strength of their relationship to θ , and the level of θ at which the likelihood of an affirmative response to the question passes fifty percent. In cases of graded responses, such as frequencies of offending reported in four different categories (e.g. 0, 1, 2-3, 4+), thresholds of θ for each categorical response are estimated. This step is analogous to factor analysis combined with seriousness scaling. The seriousness of the various items, reflected in thresholds for θ , is estimated by maximum likelihood, which determines the most likely seriousness thresholds given the observed data. In the second step, individuals are assigned values on the θ dimension, again using maximum likelihood estimation, according the level of θ which maximizes the probability of the observed response patterns. If the assumptions of the model are met, this scale reflects individual latent propensity to offend.

In the NLSY97, subjects are asked about participation in six kinds of offending: intentional destruction of property, theft of items under \$50, theft of items greater than \$50 (including autos), other property crimes, attacking someone with intent to seriously hurt them, and selling illegal drugs. For this study, the extent of delinquency involvement is represented by two different variables: a variety score indicating how many of the six items the individual reports (ranging from 0 to 6), and a latent delinquency score estimated with a graded response model (ranging from 0 to 4.2).¹ If

¹ Although the delinquency score from the graded response model takes into account frequency of offending within each category, it is very highly correlated with the variety score measure. The correlation between these two scales is .93. Despite this high correlation, results could differ if one of the measures exhibits more temporal stability.

the variety score best represents the latent trait of offending propensity, then there ought to be little difference between results using the different measures. Table 1 shows the mean variety and IRT scales for three observations prior to age 16 and one post-16 observation.

It should be noted that up to five pre-16 observations of delinquency variety and up to four pre-16 observations of latent delinquency are observed. More observations of delinquency variety are measured because survey questions allowed an estimate of delinquency variety over a year prior to the first wave of data collection. Respondents were asked if they had ever participated in each of the six delinquent activities. If so, they were further questioned on the age at which they first participated in such behavior. If the onset age for the behavior was more than 1.5 years younger than the current age, then it was assumed that the youth participated in that behavior in “wave 0.” This allows one additional wave of data to be constructed for the variety measure. But because the graded response model requires information on frequency of offending, it cannot be estimated in “wave 0.” For this reason, the sample size for the analyses using IRT scores is smaller than for those using variety scores.

MEASURING DROPOUT

One might assume that the measurement of high school dropout would be much more straightforward than that of delinquency. On the contrary, measuring high school dropout is a non-trivial task leading some researchers to lament that the true dropout rate in the United States is in fact unknown (Morrow, 1986; Rumberger, 1987). In fact, even the facts on United States dropout trends are not agreed upon by dropout researchers.

The long-accepted story espoused by the National Center for Education Statistics states that dropout rates decreased sharply from about 1920 to 1970 and have held fairly constant since then at about 15 percent of the population (U.S. Department of Education, 2001; Kaufman, Alt, and Chapman, 2004). This story, based largely on the Current Population Survey of the U.S. Census Bureau, and reports from individual states with widely varying standards of recordkeeping, has recently been challenged by several researchers. A number of researchers have pointed out that nearly a third of high school freshmen do not graduate with their class four years later and that this proportion has been increasing over the past few decades (Barton, 2005; Greene, 2002; Swanson and Chaplin, 2003). They suggest that census estimates are biased against finding dropouts. They also assert that states employ a number of methods to report lower rates of dropout. For example, some report only the percent of youth who drop out between two grades rather than over the course of four years of high school. The dropout problem has received increased attention recently as the extent of the problem has become better understood (Thornburgh, 2006). The Bush administration has made education reform a major component of domestic policy, resulting in high-stakes testing due to the No Child Left Behind act, with as yet unknown consequences for dropout rates. The Gates Foundation, in addition to commissioning an original dropout study (Bridgeland, DiIulio, and Morison, 2006) has committed over \$1 billion to education reform in the United States. Measuring dropout accurately is a pressing issue which must be addressed before the problem can be effectively addressed.

Rumberger (1987) identified six factors which influence the calculation of dropout rates: choice of cohort, initial membership in the cohort, the definition of

dropout, time for determining dropout status, source of information, and the level of determination. Widely disparate dropout rates can be traced to differences in these factors. While these issues were raised in the context of calculating aggregate dropout rates, they are also important to consider in individual-level analyses such as this. Being clear about the population of interest and how dropout is measured is necessary in order to assess the internal and external validity of the study.

Choice of cohort can make quite a difference in aggregate dropout rates, and it makes a difference when calculating dropout effects as well. The two most common choices of cohorts in dropout studies are birth cohorts and class cohorts. The cohort of interest in this study is that of 16-year-old students who have never dropped out of high school. Because of the sample design of the NLSY97, this group is drawn from a fairly narrow birth cohort, but because early dropouts are excluded, the estimates are not representative of any birth cohort. Rather, the dropout estimates can be generalized to high school students who have not dropped out by age 16.

Rumberger notes that the most critical aspect of dropout measurement is the definition of dropout. Because matching strategies are employed in this study, it is very important to measure dropout correctly. Matching strategies estimate the effect of dropout by comparing dropouts to non-dropouts who appear similar on background characteristics. If there is error in the measurement of dropout, then matched observations may in fact experience the same “treatment:” either both are dropouts or both are non-dropouts.

In most dropout studies based on large surveys such as the NLSY97, dropout is self-reported, and is essentially a residual status. If a youth is not currently enrolled in

school and has not yet completed high school, then he is considered a high school dropout. On the other hand, if he reports having completed 12th grade, then he is counted as a high school graduate. This is the baseline measure of dropout in this study. Those individuals who earn a G.E.D. instead of a high school diploma are counted as dropouts as well because their outcomes tend to resemble those of dropouts more than those of graduates (Murnane, 1999; Tyler, Murnane, and Willett, 2003).² This measure of dropout is made at every wave of data collection. For this study, all interviews which occur between the ages of 16 and 18 are aggregated to produce the dropout measure. If the individual is counted as a dropout for any of interviews between these ages, then he is considered a dropout.

There are two primary concerns with a self-reported dropout measure. First, individuals may incorrectly report their years of schooling. More likely, because of social desirability, respondents may over-report grades completed. In this case, individuals would be counted as high school graduates who are in fact dropouts. Second, even if an individual correctly reports having completed 12th grade, it may not be the case that a high school diploma was received. School districts vary in their requirements for high school graduation, and it is certainly possible to complete 12th grade without completing the requirements for graduation. In this case, again, individuals would be falsely counted as high school graduates.

² The dropout literature concerned with labor market outcomes treats GED holders as a special kind of dropout, rather than a kind of graduate (Murnane, 1999; Tyler et al., 2003) because the educational credentials of a GED are not equivalent to those of a high school diploma (Rumberger, 1987). In prior analyses using these data, estimates were compared between models where GED earners were counted as dropouts vs. where GED earners were counted as graduates. When they were counted as graduates, estimates were somewhat smaller but the results did not substantively change.

These two kinds of measurement error would lead to smaller estimates of the dropout effect because true dropouts would be compared to false graduates. In order to guard against the under-reporting of dropout, I merge the self-reported dropout measure with official reports of dropout from high school transcripts. As of wave 7 of the NLSY97, high school transcripts were requested from the last school of record for every individual who was over 18 and no longer enrolled in high school (according to self-reports). High school transcripts were obtained and coded for 6232 members of the sample by wave 7. Of these, 6100 indicate when and why the individual left the school. If the school indicated that the youth left between the ages of 16 and 18 because of dropout, expulsion or to acquire a GED, then they are counted as a dropout in this study. Merging self-reported dropout with official records of dropout increases the age 16 to 18 dropout rate from 9.8% to 12.3%.

The advantage of merging self-report with official records of dropout is that under-reporting of dropout is greatly reduced. However, this may introduce a different kind of measurement error from official records. In cases where the NLSY is not aware of the most recent school of record, the researchers may draw an out-of-date transcript which falsely indicates that the youth is a dropout. However, this kind of measurement error would derive from an individual *under*-reporting his or her own education. The NLSY relies on individuals to identify which schools they have attended. Once the youth reaches 18 years of age, and reports having left school, then a transcript is requested from the last school the youth reported attending. If this turns out not to be the most recently-attended school, then it means that the youth attended a different school without informing the NLSY administrators. The main concern for self-reported dropout is that

individuals will upgrade their education credentials rather than downgrade. This kind of over-reporting mechanism would result in inflated reports of completed grades, but more accurate reports of latest school attended, except in extreme cases where a youth falsely reports attending a particular school. Combining these two dropout measures produces the most accurate measure of dropout possible using the NLSY97.

Because this study assesses the effect of first-time dropout between the ages of 16 and 18, it is worthwhile to consider the proportion of dropouts who leave school either before age 16 or after age 18 in order to assess the extent to which the results can be generalized to dropouts in general. Event dropout rates based on the Current Population Survey can provide some guidance on this issue. The 2001 CPS found that 53.9 percent of the dropouts between the ages of 15 and 24 dropped out between the ages of 16 and 18.9. The CPS asks only if the individual dropped out in the previous 12 months, so it cannot provide exact estimates of the age at which individuals dropped out (Kaufman et al. 2004). The National Longitudinal Survey of Youth (1997 cohort), which is used in this study, can provide a much more accurate indication of when individuals dropped out of school and the proportion of dropouts who leave school between the ages of 16 and 18. Over the course of the first 7 waves of data collection, 2575 of the 8984 youth, or 28.7 percent, either self-report dropping out of school or are reported as dropped out on the collected high school transcript. Of these 2575 dropouts, 12.2 percent drop out for the first time before age 16. These early dropouts, who exhibit numerous risk factors for both dropout and other problem behaviors, are eliminated from this study. About 45 percent of the dropouts leave school for the first time between the ages of 16 and 18. These dropouts are the focus of this study. It is important to point out that around 43

percent of self-reported or transcript-reported first-time dropout occurs after age 18 in this survey. The parameter estimates derived from this study refer to the effect of dropout between the ages of 16 and 18. Dropouts between these ages are compared to non-dropouts between the same ages and contrasting outcomes are followed up to four years after age 18. Because this is not a controlled experiment, the comparison sample may in fact experience dropout after age 18, and the dropouts may return to school. It must be kept in mind that the estimates in this analysis refer specifically to dropout between the ages of 16 and 18. Earlier dropouts are eliminated from the sample, and later dropouts may be used as comparison observations.

MODELING STRATEGIES

The traditional strategy for testing the relationship between dropout and offending is a standard regression model, controlling for observed variables that may affect either dropout or offending outcomes. The ability of this “selection on observables” strategy (Heckman and Hotz, 1989) to identify the relationship between arrest and education outcomes rests on the assumption that there are no unobserved differences between dropouts and non-dropouts which explain different offending outcomes. This problem of unobserved heterogeneity suggests that model estimates are biased unless the covariates entered into the model fully account for selection processes.

A more sophisticated strategy is the use of fixed or random effects models with panel data, which allows the researcher to control for unobserved stable between-individual differences. Under certain conditions, these methods allow researchers to estimate the effects of observed time-varying characteristics free from the bias of stable

individual differences (Brame, Bushway, and Paternoster, 1999). However, these methods are not able to control for unobserved time-varying characteristics, which, under assumption, must be uncorrelated with observed static and time-varying predictors. Hence, fixed and random effects models cannot fully rule out the counter-arguments of ontogenetic theories because they are subject to omitted variable bias.

Previous work with these types of models has suggested that dropout is not a turning point event (Sweeten, 2004; Sweeten et al., 2005). However, this kind of modeling strategy has some serious shortcomings for turning points research which warrants revisiting the question with more appropriate analytical tools. In the context of turning points research, the fundamental problem for fixed effects analysis is its identification of causal effects based only on within-individual variation. Because the individual is his or her own counterfactual in this method, it cannot shed any light on the question of overlap in the traits of dropouts and non-dropouts. Are dropouts fundamentally different from non-dropouts (Eckstein and Wolpin, 1999) as some research suggests? This type of analysis does not identify the turning point effect as defined: the difference between treated and untreated trajectories of offending following treatment. In this study, “treatment” refers to high school dropout. Thus, what is needed is a treatment group that drops out of high school, and a non-treatment group, as alike as possible to the dropout group, that remains in high school. Propensity score matching directly addresses the issue of comparability of dropouts and non-dropouts, estimating the effect of dropout by comparing dropouts and non-dropouts with like characteristics.

Fixed effects analysis further assumes that the causal effect of dropout is the same for all people. Group-based methods of analysis relax this assumption, allowing for the

possibility that prior developmental patterns of behavior interact with turning point events, resulting in different outcomes for different groups. The primary benefits of the group-based trajectory method in this analysis are twofold: 1) grouping individuals in such a way so as to identify counterfactual individuals who appear as alike as possible to dropouts, 2) allowing the researcher to discern when there are no non-dropouts who can serve as counterfactuals for the high school dropout. Also, the extent of non-overlap in characteristics between dropouts and non-dropouts illustrates to what degree standard regression models depend on functional form to identify dropout effects.

Estimation of the effect of dropout on crime will proceed under two distinct but related strategies. First, I will estimate the effect of first-time dropout on offending using group-based trajectory modeling. This method allows estimation of both trajectory-specific effects of dropout, and population average effects. It uses posterior probabilities of group assignment to create balance between dropouts and non-dropouts on observable pre-dropout characteristics, and on unobservable characteristics which are correlated with observed offending patterns. The second method, propensity score matching, follows the same logical argument, but matches individuals on their probability of dropout based on observable characteristics. With propensity score matching, all pre-dropout observable characteristics related to dropout or offending, including posterior probabilities of group assignment from the first method, are used to match dropouts to non-dropouts. The first method is a much simpler version of this, where only one type of pre-dropout characteristic is used to match individuals. Both methods attempt to estimate dropout effects non-parametrically, by comparing like individuals, and both methods allow the

researcher to assess the extent to which dropouts and non-dropouts resemble each other on observable characteristics.

SEMI-PARAMETRIC GROUP-BASED TRAJECTORY MODELS

Semi-parametric group-based trajectory models begin with the assumption that patterns of behavior over time can best be approximated with a certain number of groups characterized by polynomial growth curves (Nagin, 1999; Nagin and Tremblay, 1999; Nagin et al., 2003). Using repeated observations over time, this method reports latent patterns of development over time for a set number of groups (specified by the researcher). It also reports the estimated proportion of the population which follows each developmental pattern, and, most importantly for my purposes, it reports a posterior probability of group membership for each individual and each group. Because this method models a latent developmental pathway, it is preferable to use a measure of offending which corresponds to the latent trait of interest. To the extent that IRT and variety scales represent individual latent offending propensity, and frequency scales more closely account for petty offenses, trajectory models of variety and IRT scales will better balance individuals on observable characteristics known to influence offending propensities.

In its most basic version, this technique models some dependent variable over age (or time) with a polynomial function of the following form (Nagin et al., 2003:349):

$$y_{it}^j = \beta_0^j + \beta_1^j age_{it} + \beta_2^j age_{it}^2 + \varepsilon$$

where y_{it}^j is the level of the dependent variable for some individual i at time t given membership in group j . Each group j is defined by the parameters β_0 , β_1 , and β_2 .

Polynomials of higher order can be specified as well. One potential weakness of this approach is that it requires the researcher to specify the number of groups. The researcher arrives at a final solution for the number of groups guided by improved fit, reliability of the estimated groups, and whether additional groups add to the explanatory power of the model by revealing distinctive groups of nontrivial size. It is important to note that “groups” identified by this model are not intended to represent observable distinct groups in the population, but rather support points in a continuous distribution (Nagin, 2004) (i.e. it is recognized that there may be considerable variation around the estimated average growth curves). Posterior probabilities of group assignment can be used to assess the extent to which the group-based model reliably reflects heterogeneity in developmental patterns.

Semi-parametric group-based modeling should not be blindly applied. There are a number of tools available to the researcher which can shed light on the adequacy of the model. The main problem is choosing an optimal number of groups. Normally, one begins with a one group model, then proceeds with two groups, three groups, et cetera, until adding additional groups does not improve the fit of the model. One way to assess improvement in model fit is the Bayesian Information Criterion:

$$\text{BIC} = \log(L) - .5k * \log(N),$$

where N is the sample size (number of persons, not person-waves), k is the number of estimated parameters, and L is the model’s maximum likelihood estimate (Nagin, 2005). Because adding groups will always result in a “better” fit in terms of the likelihood function, the second term serves as a discount for the additional parameters required for more complicated models. When comparing several models with different

parameterizations, one can utilize the BIC statistic associated with each of the models to estimate the probability (p_j) that each of the estimated models is the “correct” one:

$$p_j = \frac{e^{BIC_j - BIC_{MAX}}}{\sum_j e^{BIC_j - BIC_{MAX}}} \quad (\text{Nagin, 2005:70}).$$

While this is a useful tool for model selection, one cannot rely solely on this diagnostic. There are two statistics which allow assessments of the reliability of the models. First, for each of the estimated groups, one can calculate the average posterior probability of group assignment ($AvePP_j$). For each individual, and each group, the probability of group membership is estimated. By definition, these probabilities sum to 1 for each individual. To calculate $AvePP_j$, one first classifies individuals into groups according to their largest posterior probability, then the average posterior probability is estimated for each group. In a perfectly-fitting model, $AvePP_j$ would equal 1 for each group. This standard is never achieved in real-world situations, and there is no recognized test statistic for determining how far $AvePP_j$ can be from 1 before one rejects the model. Nagin’s (2005) rule of thumb is that the minimum is .70. One can use this statistic when comparing models with different numbers of groups. If the minimum $AvePP_j$ in a four-group model is .92, the minimum in a five-group model is .82, and the minimum in a six group model is .60, then one can conclude that the six-group solution does not reliably identify groups whereas the four- and five-group models do.

Another useful statistic for assessing model reliability is the odds of correct classification (OCC), which is essentially an odds ratio which reports the increase in odds of “correctly” classifying an individual into a group based on posterior probabilities relative to the estimated proportion of the sample which belongs to the group. This

compares two scenarios for classification. In the first, individuals are assigned to groups based on their highest estimated posterior probability of group assignment. The odds that this is a correct classification is represented by the ratio of $AvePP_j$ to $(1 - AvePP_j)$. In the second scheme, individuals are randomly assigned to groups according to the estimated proportion of the sample belonging to each group ($\hat{\pi}_j$). The odds of correctly classifying an individual into group j is $\hat{\pi}_j$ divided by $(1 - \hat{\pi}_j)$. The odds of correct classification using posterior probabilities relative to the proportion of the sample belonging to each group is:

$$OCC = \frac{AvePP_j / (1 - AvePP_j)}{\hat{\pi}_j / (1 - \hat{\pi}_j)} \quad (\text{Nagin, 2005:88})$$

In the case of perfect assignment using posterior probabilities, the OCC would equal infinity. However, if the model did no better or worse than chance, then the OCC would be less than or equal to 1. As with $AvePP_j$, there is no test statistic to determine when OCC is sufficiently greater than 1 to indicate a good model fit. However, it is useful to use this statistic to compare across models to see if reliability of group assignment substantially deteriorates when one adds an additional group to the model. Nagin's (2005) rule of thumb for the OCC is that it ought to be greater than or equal to 5 for each group.

Finally, one must assess model stability. In estimating complicated trajectory models, one provides starting points for the algorithm. These starting points are often based on solutions from simpler models. If different starting points lead to different final

solutions, this is problematic. One can assess the stability of the model by entering different starting values and noting whether the final estimates are consistent.

In addition to the BIC, AvePP, OCC, and model stability, one must rely on individual judgment to choose an appropriate number of groups using semi-parametric group-based modeling. The primary way of exercising individual judgment is by ocular assessment of the trajectories identified by the model. First, are the patterns parallel? If so, a non-group-based growth curve model, which estimates the average growth curve, and deviation from that curve, may be more appropriate (Raudenbush, 2001). In this case, individuals are different in levels, but not in overall patterns of change. Second, does the model identify groups of substantive interest? If a particular group is estimated to make up .5% of the total population, this may be of little substantive interest for the researcher. Even if the pattern of development is of interest, very small groups are difficult to reliably identify. Also, for the current research, very small groups make it difficult to assess whether dropouts and non-dropouts are balanced on pre-dropout characteristics because of low statistical power. Another way in which adding an additional group may be of little substantive interest is if the model simply splits a previous group into two parallel groups.

MATCHING METHODS

In order to establish a causal effect, endogeneity must be dealt with. This is of particular concern for estimating the effect of high school dropout on offending because a considerable amount of evidence suggests that delinquency and other problem behaviors increase the probability of dropout (Alexander et al., 1997, 2001; Ensminger and

Slusarick, 1992; Mensch and Kandel, 1988; Williams and Wynder, 1993). Thus, dropout may appear to have a causal effect on offending due to continuity in offending among individuals who drop out, and the real explanation is that they were involved in offending both before and after dropout. Group-based trajectory models can provide a solution for this problem, and can show whether dropout has different effects depending on criminality trajectories prior to dropout.

The methods for estimating causal effects in this study rely heavily on the work of Rosenbaum and Rubin (1983, 1985) focused on matching estimators. They proposed that in order to reproduce the conditions of an experiment, one must explicitly balance pre-treatment observable characteristics between treated and untreated groups. In the case of an experiment, treated and untreated groups are, by definition, balanced on all characteristics, observed and unobserved. However, random assignment into dropout or non-dropout status is impossible. Therefore, non-experimental methods must be used in order to determine the causal effect of dropout.

The first estimation strategy used in this study attempts to balance characteristics of dropouts and non-dropouts based on posterior probabilities of criminality trajectory group membership. Patterns of offending prior to dropout are used to “match” dropouts to non-dropouts, comparing individuals who have similar estimated developmental paths of criminality before the dropout age. The second estimation strategy extends the first by incorporating posterior probabilities of group membership into a propensity score matching framework which matches individuals on their estimated probability of high school dropout.

Haviland and Nagin (2005) point out three benefits of balancing individuals over prior developmental patterns of the outcome of interest. First, this addresses the problem of selection bias. Individuals with high levels of offending are prone to drop out, but comparing dropouts to non-dropouts with the same pattern of offending prior to the age of dropout accounts for this problem. Second, this method allows one to balance dropouts and non-dropouts on time-stable characteristics that affect offending. Third, this method allows one to balance on time-varying characteristics which affect patterns of offending both before and after dropout. There are numerous competing theoretical explanations for patterns of offending, and patterns of criminality through adolescence. The advantage of this model is that it compares people who are alike on both developmental patterns of criminality *and on those individual characteristics or social conditions which produce the trajectories*. A more detailed description of the first estimation strategy follows. The discussion below follows closely that of Haviland and Nagin (2005) as they developed this estimation strategy.

Strategy 1: Balancing on posterior probabilities of group assignment

Let D_{it} be a treatment status indicator which equals 1 if the individual i is a dropout at time t , and equals 0 otherwise. The outcome of interest is Y_{it} , the level of offending at time $t=T$. We can also consider outcomes at any time period after T , as many as are observed. Each individual has two potential outcomes:

$Y_{it} | D=0$ outcome without treatment (non-dropout), and

$Y_{it} | D=1$ outcome with treatment (dropout),

only one of which is observed. The individual's unobserved outcome must be inferred from other individuals. In order to compare like individuals, we estimate a group-based

trajectory model for all individuals in the sample for whom $D=0$ from $t=1$ to $t=T-1$. That is, only non-dropouts up to the time period of interest are included in the trajectory model. The trajectory model is estimated using a vector of lagged outcome variables from $t=1$ to $t=T-1$. This trajectory model estimates posterior probabilities of group assignment, π_{ij} , for each individual i and each group j , based on the model parameters, and the individual vectors of lagged outcomes. These posterior probabilities are then used, in a number of different ways, to group individuals. Two key tasks follow from this estimation and grouping. First, I determine whether there is within-group balance on both lagged outcomes and other time-stable and time-varying characteristics prior to time T between treated and untreated individuals at time T . Second, I proceed to estimate group-specific first-time treatment effects, and a population average first-time treatment effect based on a weighted average of the group-specific effects. It is important to note that these are “first-time” treatment effects. That is because individuals who were treated prior to time T are excluded from the analysis. In this particular study, time T refers to ages 16 to 18.

The group-specific average treatment effects (GATE) are estimated as follows:

$\hat{GATE}_j = E[Y_T^1 - Y_T^0 \mid G = j]$, where Y_T^1 equals an individual’s outcome at time T under treatment, and Y_T^0 is an individual’s outcome at time T without treatment. For all individuals, only one of these two outcomes is observed. The unobserved outcome, or counterfactual, must be estimated from other individuals. The population average treatment effect (PATE) is estimated with an average of the group-specific treatment

effects, weighted by the estimated proportion of the sample belonging to each trajectory

$$\hat{PATE} = \sum \hat{GATE}_j \hat{\pi}_j .$$

Population-average treatment on the treated (PATT) is calculated by weighting each of the group-average effects by the estimated proportion of the treated that belong to each group. This parameter is intended to reflect the effect of dropout on individuals who actually drop out. Weighting GATEs by the proportion of the untreated in each group yields the population-average treatment on the untreated (PATU). I contend that the PATT is the most interesting of the three parameters because policy efforts are directed at preventing dropout among at-risk populations. There is no reason to assess what would happen to graduates if they were induced to dropout (PATU), and because there are many more graduates than dropouts, the PATE, which is a weighted average of the PATU and the PATT, primarily reflects the effect of dropout on the untreated. For these reasons, I report the PATT.

Two assumptions are required for the estimated effects to be valid. First, treatment at time T (dropout status in this case) must be strongly ignorable given the prior developmental history of the outcome variable. What this means is that there is no selection bias once developmental history of offending is controlled for. This allows comparison of untreated individuals to treated individuals with similar patterns of offending prior to treatment (dropout), in order to estimate treatment effects. The second assumption is that posterior probabilities of group assignment serve as balancing scores for lagged outcomes. This means that there is no difference in offending histories among dropouts and non-dropouts once posterior probabilities of group assignment are controlled for. Three methods for grouping individuals based on posterior probabilities of

group assignment are discussed below. Each method slightly alters the second assumption.

The first, and simplest, method for grouping individuals based on posterior probabilities of group assignment is the classify-analyze method. In this method, individuals are classified into particular groups according to their maximum posterior probability of group assignment. This method requires a more stringent balancing assumption: estimated group membership, not the probability of group membership, is a balancing score. An individual's estimated group membership can be represented with j dummy variables, $(\hat{\delta}_{ij})$, which equal 1 if the posterior probability of membership in group j is the maximum group membership probability for that individual, and 0 otherwise. Only one of these dummy variables equals 1 for each individual, indicating which is the individual's estimated group membership. Given estimated membership in a particular group, there must be no differences in past patterns of offending between dropouts and non-dropouts. If this assumption holds, group-average treatment effects can be estimated by comparing treated to untreated groups within each trajectory. A simple two-sample t-test can be used to test whether trajectory-specific treatment effects are distinguishable from zero.

Similarly, the assumption of estimated group membership as balancing score is tested by using t-tests on within-trajectory differences in offending between dropouts and non-dropouts. Evidence for strong ignorability is derived from comparisons of non-treatment characteristics, both stable and varying, in the time periods before the age at which dropout is assessed. If, within trajectories, prior stable or time-varying characteristics differ, evidence for strong ignorability is weakened. Evidence of such

differences may indicate that the estimated effect of dropout on offending is confounded by the impact of the unbalanced characteristic. However, this is only the case if the impact of the unbalanced characteristic on offending is different before dropout than after dropout.

The second method for comparing dropouts to non-dropouts by trajectory group is the “expected value method.” In this method, rather than assigning individuals to the group to which they most likely belong, individuals are weighted according to their estimated posterior probabilities of belonging to each group. Thus, when estimating group-specific treatment effects, all individuals in the sample are considered, utilizing group assignment probabilities as weights. The balancing assumption in this case remains as originally stated: treatment is independent of prior levels of offending given posterior probabilities of group assignment. Tests for balance and for strong ignorability proceed in this method by weighting each individual according to posterior probabilities. Variance estimates for group-average treatment effects and for balancing tests must be adjusted using weighted means formulas. Variance estimates for the population-average treatment effect are obtained using bootstrapping.

The expected value method will obtain identical treatment effect estimates to the classify-analyze method if all group assignment estimates are either 0 or 1. To the extent that group assignment estimates differ from 0 and 1, the treatment effects using these two methods will diverge. The expected value method better accounts for uncertainty in group assignment by taking into account all information about estimated group assignment.

The third method for grouping individuals based on posterior probabilities of group assignment is kernel matching. Using this method, the full set of posterior probabilities is used simultaneously to create counterfactuals. The ignorability assumption is relaxed further in this case: treatment is assumed to be independent of prior levels of offending given the vector of posterior probabilities of group assignment. In this method, counterfactuals are estimated for each treated individual by weighting untreated individuals according to the distance between their vectors of posterior probabilities of group assignment.

For example, if a three group model is estimated, three posterior probabilities of group assignment are estimated for each person. The first two of these can be used to create a scatter plot (the third probability is unnecessary because it is determined by the first two). Untreated cases which are closest to the treated cases on the scatter plot are used as counterfactuals, to estimate treatment effects. There are a variety of weighting schemes (kernels) one can use to create these estimates. The most prudent course is to use several different methods to determine whether treatment effect estimates are sensitive to the kernel used. When more groups are estimated, this example is extended by measuring distances between vectors of posterior probabilities in $(j-1)$ -dimensional space.

The population average treatment effect is estimated by averaging the individual-level treatment effects. Group average treatment effects are obtained by averaging estimated treatment effects over all individuals assigned to particular groups (classify and analyze method) or by weighted averaging based on posterior probability of group assignment (expected value method). Standard errors must be obtained using

bootstrapping. Haviland and Nagin (2005:10) provide a formula for the group-average treatment effect using Mahalanobis distance, an Epinechnikov kernel, and the classify/analyze method for creating groups. In the formula below, K_{ik} is the weight provided by the kernel evaluated at d_{ik}/h where d_{ik} is the distance between two individuals' vector of trajectory probabilities, and h is a bandwidth chosen by the researcher. A dummy variable ($\hat{\delta}_{ij}$) indicates group membership according to the highest posterior probability of group assignment, and n_j is the estimated number of individuals in the sample who belong to the particular group:

$$\hat{GATE}_j = \left(\frac{1}{n_j} \right) \left(\sum_{i=1}^n \hat{\delta}_{ij} D_{iT} \left\{ y_i - \frac{\sum_{k=1}^n (1 - D_{kT}) K_{ik} y_k}{\sum_{k=1}^n (1 - D_{kT}) K_{ik}} \right\} + \sum_{i=1}^n \hat{\delta}_{ij} (1 - D_{iT}) \left\{ \frac{\sum_{k=1}^n D_{kT} K_{ik} y_k}{\sum_{k=1}^n D_{kT} K_{ik}} - y_i \right\} \right)$$

The formula inside the first brackets represents individual estimated treatment effects for the treated individuals, for whom D_{iT} equals 1 (i.e. dropouts). The formula inside the second brackets is individual estimated treatment effects for the untreated individuals (i.e. what would have been the outcomes for graduates had they dropped out). For certain policy questions, it may be desirable to report the average treatment effect on the treated, or perhaps the group-average treatment effect on the treated. Particularly for the question of the effect of dropout on offending, policymakers may be more concerned with how much crime is caused by dropout rather than how much crime is prevented by non-dropout. The above formula can be modified to obtain the group-average treatment effect on the treated by retaining the first formula inside the large brackets and dividing by the number estimated to receive treatment within each group.

This method can also be extended to measure the lasting impact of dropout by estimating treatment effects in subsequent years. In my own prior work on this topic, I have found that the effect of different kinds of dropout (as determined by reasons stated for dropping out) decay over time, and practically disappear after a few years (Sweeten, 2004; Sweeten et al., 2005). The same pattern may or may not be present when grouping individuals according to past developmental history of offending. The current data comprise seven waves (spanning seven and a half years) of information, allowing assessment of up to four post-dropout years if three years are used to estimate patterns of offending behavior prior to dropout. Estimating the effect of dropout over multiple years will indicate whether dropout may be considered a turning point in trajectories of offending. A one year effect of dropout on offending does not mean that dropout is a turning point in life course offending. Rather, the effect of dropout must be long-lasting; it must divert individuals off of their previous developmental pathway of offending (Abbott, 1997; Rutter, 1996).

Strategy 2: Propensity score matching

The second strategy is a more general form of the first. Propensity score matching, developed by Rosenbaum and Rubin (1983, 1985) also seeks to identify the counterfactual outcome using observed individual characteristics. A key assumption, called the conditional independence assumption (CIA) of matching is that treatment status is random conditional on some set of observed characteristics. Rosenbaum and Rubin (1983) showed that if the conditional independence assumption holds, then one can restate it as such: treatment status is random conditional on one's probability of treatment as determined by observable characteristics. Thus, individuals are matched

based on their “propensity” for treatment. This extension of matching was extremely useful as it reduced the dimensionality of matching to just one, which allowed the application of propensity score matching to expand to many different contexts.

The CIA of propensity score matching is a relaxed version of the assumption in the previous method that treatment status is random given prior developmental history of offending. In the current application, prior developmental history of offending is included in propensity scores using posterior probabilities of group assignment. In addition, all observable characteristics which are thought to affect either dropout or offending are also included in the model. If the techniques in the previous section are not able to convincingly achieve balance in prior developmental histories of offending or in other characteristics, propensity score matching may be a way to extend the method. Propensity score matching allows retention of developmental history of offending in the model predicting treatment, but also allows inclusion of many other factors, which makes the CIA more plausible.

Propensity score matching is similar to standard regression techniques in that it is assumed in both cases that selection into treatment is random conditional on observed characteristics. However, propensity score matching differs considerably from regression techniques in two ways. First, it does not rely on a linear functional form to estimate treatment effects. Although propensity scores are estimated using either a logit or probit model, once these are obtained, individuals are matched non-parametrically. The second key difference between propensity score matching and regression models is that propensity score matching highlights the issue of common support. It shows the researcher, in a practical way, how many of the untreated (non-dropout) individuals

actually resemble the treated individuals (dropouts) on observed characteristics.

Regression techniques, on the other hand, obscure this issue, and can, in some situations, extrapolate treatment effect estimates based solely on functional form when treated and untreated groups are actually not comparable at all. In contexts where treated and untreated groups are not comparable, propensity score matching easily allows the researcher to recognize the problem and either abandon or supplement the data.

Multivariate analyses, in particular, complicate the matter, easily obscuring reliance on functional-form assumptions. However, no matter how many variables are taken into account in propensity score matching, it will still produce a single scale representing one's propensity to drop out (or some other treatment of interest), allowing for easy assessment of the extent to which non-treated cases can serve as counterfactual estimates for the treated cases. In some applications, only a small proportion of the untreated population is useful for estimating treatment effects.

As previously mentioned, propensity scores are estimated using a logit or probit model where the dependent variable is treatment status, and the independent variables comprise a set thought to influence either treatment or the outcome of interest. In the current application, I will use posterior probabilities of group assignment to control for prior history of offending, and I will include an array of variables identified both by theory and by prior research that influence either dropout or offending. The logit model will be used to assign each individual a propensity score, which is an estimated probability of dropping out.

Once propensity scores have been estimated, there are a number of ways in which one can test whether or not balance has been achieved. That is, whether conditioning on

the propensity score makes treatment appear random. One method is to divide the sample into equal-sized bins based on the propensity score level, and estimate differences in both propensity scores and other independent variables between treated and untreated cases within bins (Dehejia and Wahba, 1999, 2002). If treated and untreated cases are not balanced, the initial propensity score model is modified to include more variables, or to include interaction terms or squared terms.

Another way to assess balance is standardized bias. This method, first described by Rosenbaum and Rubin (1985:36) begins with a measure of unadjusted bias, which is the difference between the treated and untreated divided by an equally weighted combination of the standard error within the two groups, multiplied by 100. If this statistic exceeds 20, the characteristic is considered unbalanced. This is a useful measure to compare balance across multiple matching methods. As such, it is used to assess population-average balance in each of the three trajectory based matching methods, and for propensity score matching.

There are also a number of options for matching treated to untreated cases using propensity scores. The first consideration is the issue of support. This concerns the range of propensity scores for the treated individuals versus the range for the untreated. In some cases, these will not entirely overlap, and so the researcher must decide how to treat those cases which fall outside the range. Of particular concern are treated individuals for whom there are no comparable untreated cases. While the lack of common support is valuable information, simply omitting these individuals may alter the nature of the estimated treatment effect, particularly if a large portion are excluded. Certain matching methods described below can lessen this problem by matching slightly outside the

boundaries of common support, but some cases may still be excluded depending on the extent of non-overlap.

The simplest matching method is nearest neighbor matching, in which the untreated case with the closest propensity score to a treated case is used as a comparison (Smith and Todd, 2005). There are several variants to the method: matching can be done with or without replacement, and individuals can be matched to one or several of their nearest neighbors. Matching without replacement means that once a non-treated case has been matched to a treated case, it is removed from the candidates for matching. This potentially leads to poor matches when the density of propensity scores is quite different for the treated and untreated groups. Matching with replacement allows untreated cases to serve as the counterfactual for multiple treated cases. This allows for better matches, but reduces the number of untreated cases used to create the treatment effect estimate, which increases the variance of the estimate (Smith and Todd, 2005). Matching with replacement can also be implemented using multiple nearest neighbors, the mean of which serves as the counterfactual. Nearest neighbor matching typically does not address the issues of common support, although common support restrictions can be applied.

A second matching method is caliper matching, which is similar to nearest neighbor matching, but imposes a restriction on the distance between counterfactuals, thus eliminating poor matches, but potentially throwing out treated cases where no sufficient match can be found. This can be a useful strategy for imposing common support, but in practice, there is no guideline for what size caliper to use.

Another common method groups cases into intervals based on propensity score levels, and estimates average treatment effects within each interval by comparing all

treated cases to all untreated cases. The population average treatment effect on the treated is derived by weighting the interval-level treatment effects by the portion of the treated population within each bin (Smith and Todd, 2005).

Similar to the method previously described for matching on a vector of posterior probabilities of group membership, one can also employ kernel matching with propensity scores. In fact, kernel matching is simpler in this context as there is only one dimension used to match cases. Kernel matching is a method for weighting counterfactual cases according to their distance from treated cases. In fact, all matching methods may be characterized as weighting functions, but kernel matching allows for finer distinctions in weighting than other methods. The researcher must decide which kernel function and bandwidth to use. When there are many cases at the boundaries of the propensity score distribution, it may be useful to generalize kernel matching to include a linear term; this is called local linear matching. Its main advantage over kernel matching is that it yields more accurate estimates at boundary points in the distribution of propensity scores and it deals with different data densities better (Smith and Todd, 2005).

Each of these methods for matching may be useful depending on the data. If, for example, there are numerous treated and untreated cases throughout the propensity score distribution, then nearest neighbor matching without replacement may be the best option. However, if the distributions are quite different, then kernel matching may be preferred. If the distributions are different, and there are many cases at the boundaries, then local linear matching may be the best method. In this paper, I will implement a number of these matching strategies in order to determine the sensitivity of the estimates to the

matching method. If the estimates differ considerably, the distribution of the propensity scores should explain the differences and point to which method is the most appropriate.

SUMMARY

The methodological innovations proposed in this study are aimed at remedying several issues which plague the study of the effect of high school dropout on crime, and on etiological criminology studies in general. First, there is a fundamental measurement problem for criminal behavior. Contemporary strategies for measuring delinquency and crime are unsystematic and various. In this study, delinquency and crime is measured in two ways: a variety score and latent scores derived from a graded response model. The advantage of latent scores from a graded response model is that they are based on a theoretical model of measurement which models the observed pattern of responses to the unidimensional underlying trait proposed to account for the responses. The resultant measure of crime accounts for both seriousness and frequency of offending while trajectories of this construct capture developmental patterns of criminal propensity, and in so doing, balance individuals on both offending and characteristics which lead to offending.. Treatment effects derived from this measure can be directly related back to patterns of offending behavior based on the estimated relationship between the response items and the underlying trait. If, in fact, estimates using variety score measures of crime are as informative as those using propensity scores, this may serve to affirm the continued use of variety scores, and may suggest that they closely approximate latent offending propensities.

Standard regression models for estimating the effect of some individual characteristic or life event on crime are subject to a number of criticisms which this study tries to avoid. One common criticism is that there is selection on unobservables, or at least on certain variables not included in the model. The strategies employed in this study contend that there is no selection on unobservables under certain conditions. Endogeneity issues are minimized by requiring that individuals cannot have dropped out prior to the age of interest, and by modeling lagged outcomes of offending behavior using semi-parametric trajectory modeling. Furthermore, balancing tests can lend some credibility to the claim of observable selection into treatment. Standard regression models are less often criticized for their reliance on functional form to estimate treatment effects. The methods employed in this study shed light on the issue of common support and indicate the extent to which regression models rely on functional form to yield estimates. The methods employed here are also very explicit about the construction of a counterfactual, and allow utilization of different matching techniques to suit the particular distribution of propensity scores.

Finally, several researchers have proposed that the effect of dropout on offending may actually depend on different individual characteristics or reasons for dropping out (Jarjoura, 1993, 1996). Furthermore, life course research suggests that turning point events can have very different effects on different individuals depending on a number of factors including prior developmental history of offending. This study explicitly estimates the effect of high school dropout on offending among groups with different developmental patterns of criminality prior to dropout. Finally, this study points to a method for determining whether a certain life event is a turning point. The turning point

event is characterized as a treatment, and outcomes among matched groups are assessed over a number of years to determine whether treatment effects are stable, decaying, or increasing over time. If for at least one of the estimated groups, the effect of dropout on offending is statistically distinguishable from zero, and persists over time, then there is evidence that it is a turning point.

CHAPTER 4: RESULTS

I begin my analysis with baseline comparisons of dropouts and non-dropouts on post-16 offending, unadjusted for observed covariates. Then I present estimated trajectory models of delinquency variety and delinquency IRT scores. In the following section, within-trajectory, and population-average balance is assessed for three types of matching: classify-analyze, expected value, and kernel matching with posterior probabilities of group assignment. I compare the extent of balance between IRT-based trajectory models and variety score-based trajectory models in order to determine if the latent trait trajectories provide more effective balancing scores. I then present trajectory-adjusted treatment effect estimates of dropout on delinquency using the three matching methods for both variety scores and IRT scores. These results are contrasted to propensity score matching estimates. Finally, I discuss the bias reduction using each of the matching methods.

BASELINE COMPARISONS

In agreement with a large body of prior research, it is clear from Table 2 that dropouts are more delinquent than non dropouts. Further, these differences persist over time. Even three to four years after dropping out of high school, dropouts exhibit around twice as much delinquency as non-dropouts. The unadjusted population-average treatment effects (U-PATE) based on these comparisons remain statistically significant four years after dropout even though they decrease in magnitude every year. Because

these estimates are unadjusted, any of a number of pre-dropout differences could explain observed post-dropout differences.

As discussed in the Chapter 2, much of the current literature on dropout sees it as the end result of a long process of disengagement from school. There are large observable differences between dropouts and non-dropouts both before and after dropout. The question is whether adjusting for pre-dropout differences eliminates post-dropout differences, meaning that the dropout “event” has no criminogenic effect over and above the process that led to it.

DELINQUENCY TRAJECTORIES THROUGH AGE 16

Group-based semi-parametric trajectories of delinquency variety are estimated through age 16 on up to 5 waves of data for the 2990 youths who have at least three pre-16 observations of delinquency variety, at least one post-18 observation, and who had not dropped out of school prior to age 16. Through consideration of BIC statistics, indicators of group stability and group size, a four-group model was chosen. The minimum average within-group posterior probability in this model is .761, and the minimum odds of correct classification (OCC) is 9.69. According to Nagin’s (2005) rules of thumb, these statistics indicate that a four-group model is a good fit for the data. A five-group model fits the data somewhat better according to comparison of the BIC statistics, but is not used because it isolates a group comprising only 4% of the sample. This small group proved to be unstable in bootstrap replications of the trajectory models, which are necessary for estimating standard errors of population-average effects of dropout. Preliminary

investigation into results obtained from the five-group solution indicate that the general conclusions of the analysis would be the same.

A graphic representation of the four-group trajectory model used for matching is shown in figure 1. The graph plots estimated delinquency variety scores for each of the four groups from ages 11 to 15.5. I do not plot up to age 16 because ages are centered in the reporting period, and because all observations are prior to age 16, very few reporting periods are centered higher than 15.5. The graph goes down to age 11 because of the constructed “wave 0” delinquency variety measure. The largest estimated group, representing 42 percent of the sample, reports zero delinquency in every wave. The second largest estimated group exhibits low levels of delinquency early, with less delinquency as they approach age 16. Two estimated groups increase in delinquency over this time period, one initiating offending around age 14, and the other offending consistently throughout the time period.

Additional characteristics of these four estimated groups are show in Table 3. Dropout clearly does not arrive at random with respect to these four groups. The 16-18 dropout rate within the most delinquent group is nearly four times as large as the dropout rate among the non-offending group (27% vs. 7%). The four-group trajectory model differentiates between youths on several dropout risk factors. The most delinquent group receives worse grades, has lower math scores, is less likely to have a computer in the home, is five times more likely to smoke, and 27 percent less likely to live with both biological parents than the non-offending group. The four-group trajectory model identifies theoretically interesting groups for the study of dropout. The most delinquent group appears to be more disengaged from school than the least delinquent group. In

order for estimates of within-group dropout effect estimates to be valid, the trajectories must serve to balance prior offending and other characteristics. Balance will be examined in the following section

A four-group model is used for trajectories of latent delinquency scores (IRT models) as well. The four-group model, with a minimum within-group avePP of .789 and minimum OCC of 11.54, is clearly the best-fitting model for this smaller sample of 1633. Fewer cases are available for the IRT trajectory model because a “wave 0” can not be estimated. A five-group model is inestimable; regardless of starting values, the model fails to reach convergence. The picture, shown in Figure 2, is fairly similar to that of the variety score model. The largest estimated group, 52 percent of the sample, reports zero delinquency. The remaining three groups, although distributed differently in the population, roughly correspond to the decreasing, increasing, and highly delinquent groups estimated using variety scores. The dropout rates within the four groups do not differ as much as the groups based on variety scores. In the IRT trajectory model, the dropout rate of the most serious group is 2.5 times higher than the lowest delinquency group. The other risk factors, shown in Table 4, do not differ as much between groups either. Nevertheless, the IRT model does estimate groups that differ considerably on pre-16 development of delinquency, as does the variety score model.

TESTS OF BALANCE

Trajectory models are used to group individuals according to patterns of development in delinquent behavior. For each person and each group, a posterior probability of group assignment is calculated. This is an estimate of the probability that the person belongs to each of the four groups given his or her observed pattern of

delinquent behavior. These posterior probabilities are then used in several different ways to group individuals and compare dropouts to non-dropouts within groups. The focus of this section is examining balance in pre-dropout characteristics between dropouts and non-dropouts under each of the grouping schemes. If dropouts and non-dropouts are statistically indistinguishable on pre-dropout characteristics conditional on group assignment, then post-dropout comparisons capture the effect of the dropout event.

There are a number of ways to assess balance in this context. The first is to conduct simple t-tests within trajectories. Second, dropouts and non-dropouts can be compared within the entire sample, adjusting for group assignment. Because of uncertainty in group assignment, the correct way to obtain standard errors for this type of comparison is to measure variability in the adjusted difference between dropouts and non-dropouts in multiple bootstrap iterations. For each bootstrap sample, the trajectory model is re-estimated using starting values from the original four-group trajectory model. Then within-group differences are weighted according to the proportion of dropouts estimated to belong to each group in order to obtain both the population-average treatment on the treated effect (PATT) and population-average differences in background characteristics.

The few studies that have used semi-parametric group model matching methods have assessed balance on relatively few background characteristics. Haviland and Nagin (2005) checked balance on just nine background characteristics. Haviland, Nagin and Rosenbaum (2006), in their study of the effect of gang membership on teen violence, checked balance on 15 covariates and 7 missing value indicators. Apel and colleagues' (2006) study of the effect of adolescent employment on antisocial behavior assesses

balance on 111 covariates using the same dataset as this study. There is no standard for how many comparisons is enough, but in order for effect estimates from matching to be valid, there must not be significant differences in either observed or unobserved pre-treatment characteristics. As such, the more characteristics for which balance can be demonstrated, the stronger the evidence for the validity of the matching estimates. In this study, standardized and adjusted bias is estimated for 314 covariates for four different matching methods. Many of these covariates are different measures of the same construct, but a very wide range of topics are covered, and it would be difficult to make a case that any important characteristic is omitted that is not highly correlated with at least one of the observed characteristics. Tests of balance using the standardized bias measure are presented in appendices B (variety scores) and C (IRT scores) for all 314 covariates and all four matching methods. Within-group tests are presented in the body of this text for certain characteristics that displayed the most unadjusted standardized bias. There is very strong evidence that pre-dropout characteristics differ quite a bit between dropouts and non-dropouts. Of the 314 characteristics assessed in Appendix B, only 183 (58%) are balanced before adjusting for group membership.

With the classify-analyze grouping method, individuals are classified according to the maximum posterior probability of group assignment. Individuals are placed into four groups according to the most likely group membership. Within-group and population-average comparisons on pre-dropout characteristics are presented in Table 5.

Comparisons are presented for three waves of prior delinquency, and select high-bias covariates. Full population-average comparisons are presented in Appendix B. The lack of balance evident in these comparisons is quite striking. Grouping dropouts and non-

dropouts according to the most likely group membership does very little to balance dropouts and non-dropouts even for past delinquent behavior. Within trajectories, and in the entire sample adjusting for group membership, dropouts are more delinquent, have worse middle school grades and worse math aptitude, are less likely to have a computer, more likely to smoke and less likely to live with both biological parents. In fact, only 7 of the 37 comparisons in Table 5 are *not* statistically significantly different.

Dropouts and non-dropouts within the non-offending group are balanced by definition on prior delinquency as none of the group reports any. However, even within the non-offending group, there are large differences between dropouts and non-dropouts on other characteristics. Standardized post-matching bias for this method is presented in Appendix B for all 314 pre-dropout characteristics. According to Rosenbaum and Rubin's (1985) standard for judging balance, only 222 of 314 (71%) of the covariates are balanced. While this is an improvement over the unadjusted comparisons, there remain large differences for numerous pre-dropout characteristics including numerous socio-economic, school, peer, family and delinquency measures. It should be pointed out, contrary to bootstrapping tests of balance, the standardized bias method of assessing balance leads to the conclusion that balance is achieved on pre-dropout offending. However, because there are so many unbalanced characteristics, there is little reason to believe that effect estimates for dropout are unbiased by prior differences between dropouts and non-dropouts.

The classify-analyze matching method is a relatively coarse matching method, grouping the entire sample into four bins. The expected value method better captures the probabilistic nature of the trajectory model by weighting all observations according to the

posterior probability of group assignment for each group. It is possible that better accounting for the probabilistic nature of group assignment will better balance individuals on pre-dropout characteristics. T-tests of within-trajectory balance, and bootstrap tests of group-adjusted balance for the entire sample are presented in Table 6. Again, there are very large within-group and group adjusted differences between dropouts and non-dropouts. The expected-value method does not improve balance compared to the classify-analyze method. In fact, according to the standardized bias method of assessing balance, it performs slightly worse, balancing dropouts and non-dropouts on only 69% of 314 covariates. Because of lack of balance, there is no reason to believe that the expected value estimates of the effect of dropout on delinquency are any freer of bias than the classify-analyze estimates. Both would appear to be potentially biased by pre-existing differences.

The third matching method weights individuals based on their vector of posterior probabilities of group assignment. The mahalanobis distance between observations is used with a gaussian kernel and a bandwidth of .04.³ In this case, because the vector of posterior probabilities is used, it is not possible to compare dropouts and non-dropouts within groups. Only population-average comparisons can be made. These comparisons are presented in Table 7. While the estimated differences in pre-dropout offending are smaller than those from the previous two matching methods, a significant difference in offending prior to age 16 is observed. Also, dropouts and non-dropouts exhibit large adjusted differences in pre-dropout characteristics. This method performs just slightly better than the previous two according to the standardized bias method of assessing bias,

³ Both the balance estimates and effect estimates based on this method were largely insensitive to different types of kernels and different bandwidths.

presented in Appendix B. In this case, 228 of 314 covariates are balanced, six more than in the classify-analyze method. Still, 27% of the pre-dropout characteristics remain unbalanced. There is little evidence that this or any of the matching methods using group-based trajectory models are able to achieve balance on pre-existing characteristics of dropouts and non-dropouts. Because of this, there little reason to believe that estimates of the effect of dropout using these methods are unbiased.

Parallel analyses are performed with the IRT trajectory analyses. Within-group tests of balance are not presented because the large disparity in sample size and statistical power could affect statistically observable differences. For this reason, only the standardized bias method of testing balance is used with IRT models as this method is insensitive to differences in statistical power. These tests of balance are presented in Appendix C. The results very closely parallel those using variety score trajectories. Each of the three matching methods falls short in balancing pre-dropout covariates among dropouts and non-dropouts. The classify-analyze method performs the best in terms of matched variables, and it leaves 26% of the covariates unbalanced. There is no evidence, therefore, that there is any advantage to using IRT scores rather than variety scores in matching dropouts and non-dropouts. Further, there is little reason to believe that effect estimates based on IRT trajectories are unbiased by pre-existing characteristics. This reservation should be kept in mind when assessing dropout effect estimates in the following section.

DROPOUT EFFECT ESTIMATES

Despite the fact that none of the trajectory-based matching methods is able to achieve balance, it is still useful to estimate the effect of dropout using each method.

Comparing the magnitude of these estimates to unadjusted estimates and propensity score estimates can give a sense of both the sources and the magnitude of bias. Group-average treatment effects (GATE) and population-average treatment on the treated (PATT) for each of the three matching methods are presented in Table 8. T_1 refers to the period between age 16 and 18 during which dropout itself occurs. The differences between dropouts and non-dropouts for this time period are concurrent with dropout. In fact, offending may occur prior to dropout during this time period. For this reason, the main focus is on wave T_2 , the first post-18 observation of offending. In this case, the timing of events is straightforward. Dropout occurs prior to this measure of offending. Because of lack of balance between dropouts and non-dropouts, there is little reason to believe that these estimates are unbiased.

The effects of dropout appear to differ by both group and time-period. Offending prior to age 16 is balanced by construction in group 1, despite imbalance on other factors. For this group, both matching methods yielded identical effect sizes. The post-dropout effect of dropout is .13 within this group and actually increases in the second post-18 observation. However, by the third post-18 observation, there are no discernable differences between dropouts and non-dropouts in offending. Thus, for youths who exhibit no offending behavior prior to age 16, dropout appears to have a short-term criminogenic effect. Lack of balance could challenge this conclusion if the pre-existing differences in background characteristics account for post-dropout differences in offending.

Within group 2, the teen-onset group, post-dropout differences are inconsistent from year to year. According to the classify-analyze method, there are no discernable

post-dropout differences in offending between dropouts and non-dropouts in group 2. However, using the expected-value method, dropouts offend more than non-dropouts in waves T₂ and T₄, but not T₁ and T₃. There is no theoretical explanation for such an inconsistent effect over time. The most probable explanation is that the sample for which these estimates are made differs by wave. A T₄ observation is made for only about half of the sample. In the expected value method, it is estimated that this effect is based on the difference between 17 dropouts and 96 non-dropouts. For this select group, the differences between dropouts and non-dropouts are actually quite consistent over the three post-dropout observations, increasing from .18 in wave T₂ to .25 in wave T₄. Although this group is younger than the remaining sample in wave 1 of the NLSY97, they are aligned on age in this analysis, and there is no reason to expect the effect of dropout to differ among cohorts who differ by only a year in age.

Among Group 3, the group that exhibits low levels of decreasing offending leading up to age 16, the effect of dropout is somewhat larger: .25 in wave T₂ according to the classify-analyze method, and .29 according to the expected value method. Recall, dropouts and non-dropouts appear to be balanced on offending just prior to age 16, but are not balanced in waves T₂ and T₃. In addition, as with all groups, there are large differences in other pre-dropout characteristics that can explain post-dropout differences. Using the expected value method, differences between dropouts and non-dropouts persist even three years after dropout. However, using the classify-analyze method, only differences in T₁ and T₂ are statistically significant.

For the most delinquent group, I have the least confidence in post-dropout estimates. Dropouts had much higher delinquency variety scores in wave T₁. In fact,

differences between dropouts and non-dropouts are larger in this pre-dropout period than in any of the post-dropout periods. This would seem to suggest that elevated levels of delinquency lead to dropout rather than dropout leading to increased delinquency.

Within this group, the differences between dropouts and non-dropouts actually decreased after dropout occurred.

Although this is an interesting parameter, there is little reason to believe that PATT estimates presented in table 8 are unbiased because pre-dropout differences in offending rival post-dropout differences in magnitude, suggesting that there is little effect of dropout whatsoever. According to the classify-analyze matching method, pre-dropout differences in offending are .16 in both of the two waves prior to age 16. Post-16 and post-18 differences using this method are .20 and .22 respectively. All of these differences are statistically significant, but there is no reason to attribute the whole of the post-dropout differences to dropout itself. In fact, the difference in offending between dropouts and non-dropouts increases by only .04 post-dropout. While the effect of dropout differs somewhat using different matching methods, the “bump” in offending due to dropout is very comparable in size. In the expected value method, the pre-dropout difference between dropouts and non-dropouts is .21 and the post dropout effect is .24, an increase of .03. In the kernel matching method, which exhibited the most balance of the three methods, the pre-dropout difference between dropouts and non-dropouts is .09 and the post-dropout difference is .14, an increase of .05. For all three methods, differences between dropouts and non-dropouts two and three years after dropout are actually smaller in magnitude than the pre-dropout differences. Because pre-dropout differences in offending are comparable in size to post-dropout differences, there is little reason to

believe that the dropout event itself is criminogenic. If it is criminogenic, the effect is quite small, and endures for only one year, as there are no discernable differences between dropouts and non-dropouts in offending two and three years after dropout despite significant pre-dropout differences. According to this analysis, first-time dropout between the age of 16 and 18 is not a turning point in trajectories of offending.

Post-dropout differences using IRT scores are presented in Table 9. Here, there are significant positive post-18 dropout effects only for groups 1 and 4. For group 2 in wave T₃, there is actually a statistically significant *negative* effect of dropout. That is, the dropouts commit less crime than non-dropouts. However, none of the post-18 PATT estimates are statistically significant despite statistically significant pre-dropout differences. Because of the lack of pre-dropout balance between dropouts and non-dropouts, all of the effect estimates in this section are suspect. Effects of the process of disengagement which leads to dropout are not eliminated by matching on trajectories of offending. In the next section, propensity score matching is employed in an attempt to better control for these large pre-existing differences between dropouts and non-dropouts.

PROPENSITY SCORE MATCHING

Propensity score matching begins with a probit model predicting dropout. Every observable pre-dropout characteristic which is conceivably related to either dropout or offending is included in this prediction model, the purpose of which is to generate a scale indicating the probability that each individual will drop out based on observed covariates. Dropouts are then matched to non-dropouts who exhibit roughly the same propensity

score. This analysis is conducted on the same sample used for variety score trajectory models to ensure that any differences are not due to sample selection.

Variables included in the probit model are indicated in Appendix B with asterisks. In cases where valid observations are not obtained for the entire sample, missing observations are assigned a value of zero, and a missing value indicator is included. Where missing indicators are identical for multiple variables, only one of the indicators is retained. Balance is assessed only on valid observations.

The distribution of propensity scores by dropout status is shown in Table 10 and Figure 3. While there is common support, there are very few non-dropouts with high propensity scores. The vast majority of non-dropouts are not useful counterfactuals for dropouts while a very small minority of non-dropouts are heavily weighted in effect estimates based on propensity scores.

Balance on pre-dropout characteristics conditional on propensity scores is presented in Table 11. Clearly, propensity score matching performs much better on tests of balance than matching based on trajectory models. Although the magnitude of pre-dropout differences between dropouts and non-dropouts remains similar to that of kernel matching using posterior probabilities of group assignment, they are statistically indistinguishable from zero. Furthermore, the other covariates exhibit much better balance. Assessing matching using the standardized bias method, presented in Appendix B, yields even more striking differences. The best-performing matching method based on trajectory models was only able to achieve balance for 73% of the 314 covariates. Propensity score matching achieves balance on all but four of the 314 covariates (99%). This is better balance than one would expect in a randomized experiment. Propensity

score matching comes closest to approximating an experiment using observational data. Conditional on an individual's propensity score, dropout appears to arrive at random. Therefore, I am much more confident that dropout effect estimates from propensity score matching are unbiased by unobserved differences between dropouts and non-dropouts. Estimates from propensity score matching are the most appropriate for answering the question of whether dropout is a turning point in pathways of offending.

Treatment effects obtained from propensity score models are presented in Table 12. None of the treatment effects are statistically distinguishable from zero. A 95% confidence interval is presented for each estimate, showing that the data is consistent with a wide range of possible treatment effects. Recall, dropouts have .10 high variety scores than non-dropouts in time T_1 . If we subtract this difference from the estimates in Table 12, the estimated effect of dropout in time T_2 is .08, which is larger than the estimates obtained using trajectory model matching, but still quite small and not statistically significant. It is notable that the matching method that provides the best balance in pre-dropout differences is the only method for which no significant post-dropout differences are obtained. This is strong evidence that first-time dropout between the ages of 16 and 18 is not a turning point in offending trajectories.

Propensity score estimates using IRT scales parallel the variety score estimates. The pre-dropout difference in offending is .10, but not statistically significant. Post-dropout differences are statistically indistinguishable from zero, and similar in magnitude to pre-dropout differences. Thus, there is no evidence of a criminogenic effect of dropout on offending using IRT delinquency scores. This method provides the best balance of all the matching methods using IRT delinquency scores, balancing 93% of 311 observed

covariates, compared to 74% for the best matching method using trajectory analysis. In the next section, I more closely compare the four matching methods in terms of reductions in standardized bias.

REDUCTIONS IN STANDARDIZED BIAS

A useful metric of balance is reduction in absolute standardized bias, which is calculated by first determining unadjusted bias, then calculating adjusted bias and the percent decrease in the absolute value of each. Reductions in standardized bias by method are presented in Table 13 for the covariates that are the most biased in unadjusted comparisons. These variables have unadjusted bias estimates above 40 in magnitude. The variables are grouped according to type. The largest group is related to school achievement and aptitude. The most biased of all 314 covariates is middle school grades, with an unadjusted bias score of 87. The three matching methods based on trajectory analysis are largely ineffective in reducing bias in middle school grades. Nor do they substantially reduce bias in any of the school-related variables. Dropouts have poorer grades, less aptitude in numerous ASVAB sub-tests, are more likely to be suspended, less likely to be academically gifted and spend less time doing homework. Matching based on developmental history of delinquency does very little to balance these school-related characteristics. On the other hand, propensity score matching performs remarkably well. Bias is reduced by over 90 percent for most of the school-related variables, and the smallest bias reduction using propensity score matching is 78 percent, which is over twice as large as the largest bias reduction in this category using trajectory-based matching methods.

The second-largest category of biased variables is related to socio-economic status. Dropouts are much less likely to have a computer in the home. This characteristic is not strictly one of socio-economic status, but also of family educational support because acquisition of a computer is within reach of families even with modest means. Dropouts tend to come from poorer families with less educated parents. This is demonstrated by numerous large unadjusted differences between dropouts and non-dropouts on these characteristics. Matching dropouts to non-dropouts based on developmental history of delinquent behavior performs dismally in reducing this bias. The greatest reduction in bias for any of these factors using the classify-analyze matching method is 7 percent. For kernel matching with posterior probabilities of group assignment, the greatest reduction is 21 percent. Propensity score matching performs dramatically better, reducing standardized bias by over 90 percent for all but two of these variables.

Dropouts are also more delinquent than non-dropouts prior to dropping out of high school. They are more likely to smoke cigarettes, use marijuana, and attack others, and they have higher delinquency variety scores in general. Matching individuals on developmental history of delinquency is a fairly effective way to reduce bias among delinquency characteristics. The three trajectory model matching methods perform much better in this category than in the previous two. Kernel matching produces the greatest decreases in standardized bias of the three methods, with at least 50 percent reductions for the high-bias delinquency items. Even so, bias reduction in these variables using propensity score matching rivals that of kernel matching. Although this is the category in

which trajectory matching performs the best, it does not perform significantly better than propensity score matching in reducing standardized bias.

Dropouts also have more antisocial peers than non-dropouts. Peers of future dropouts use drugs more, are more likely to be gang-involved, cut classes more, smoke more, and are less likely to attend church. Reductions in bias for these factors using trajectory matching methods are modest at best. The kernel matching method again performs the best of the three methods. However, once again, propensity score matching is the most effective in reducing standardized bias among these factors. The same pattern holds among high-bias family characteristics.

Comparisons of bias reduction have yielded some interesting conclusions. First, among the three different matching methods which utilize posterior probabilities of group assignment, the best method for reducing standardized bias is kernel matching. Second, matching on patterns of delinquency are only effective in reducing bias among delinquency items, or those items, such as antisocial peers, which are highly correlated with delinquency. For items that are not highly correlated with delinquency, matching on posterior probabilities of group assignment does little to reduce bias. Finally, in this context, propensity score matching performs much better than the other three matching methods.

CONCLUSION

There is no evidence from these analyses that the event of first-time dropout between the ages of 16 and 18 is turning point. There are numerous striking differences between dropouts and non-dropouts prior to dropping out of high school. Although

dropouts are more delinquent than non-dropouts, matching only on patterns of delinquency does not adequately balance the two groups on other characteristics such as school achievement, aptitude, socio-economic status, and peer characteristics. The failure of trajectory-based matching models to achieve balance does not suggest that dropouts are fundamentally different from non-dropouts. This is evident from unadjusted comparisons alone. Rather, it suggests that developmental patterns of delinquency are only a small part in the selection process leading to dropout.

Despite the lack of balance achieved by trajectory matching methods, comparisons of post-dropout differences in offending to pre-dropout differences can give some guidance on the magnitude of the dropout effect. This is a type of difference-in-difference estimate. For group-adjusted population-average estimates, dropouts are more delinquent than non-dropouts *before* dropping out of school, and the differences between the groups do not markedly increase following dropout. This suggests that post-dropout differences in offending can be attributed to pre-dropout differences that are evident long before dropout occurs. This is consistent with the idea that dropout is the result of a long process of disengagement from school, and that the dropout event itself does not have an added criminogenic effect over and above the process that led to it.

The difference in balance achieved by propensity score matching in comparison to trajectory model matching is striking. Using the standardize bias metric, trajectory score matching, at best, balances only 73% of the observable characteristics. Propensity score matching balances 99% of these characteristics, which is better than one would expect even from a random assignment experiment. As such, dropout effects derived from propensity score matching are the only parameters that I trust not to be biased by

pre-existing differences. It is striking, therefore, that propensity score estimates, in contrast to trajectory matching estimates, are indistinguishable from zero. This further confirms the conclusion that first-time dropout between ages 16 and 18 is not a turning point. When pre-existing differences are eliminated through propensity-score matching, post-dropout differences disappear.

Propensity score matching confirms that dropouts are very different from non-dropouts. Three-fourths of non-dropouts have propensity scores below .1 whereas propensity scores for dropouts are fairly evenly distributed across the full range.

Although there is common support, a small minority of the non-dropouts comprise the bulk of the comparison sample for dropouts. While this is not a surprising result, it is an interesting finding with implications for dropout studies. Studies which equally weight all individuals in estimating the effect of dropout are estimating the population-average treatment effect, which primarily reflects the estimated effect of dropout for individuals who have a remote chance of dropping out. In this study, on the other hand, I estimate treatment on the treated, the effect of dropout on dropouts. Because this group has so many other deficits, there is no added effect of dropout itself. An alternate way to think of this result is that graduation does not reduce delinquency among dropout-prone youth. Controlling for propensity to dropout, there is no discernable difference between dropouts and non-dropouts on delinquency.

CHAPTER 5: CONCLUSION

The Gates Foundation study on high school dropout, published in March 2006, warns of the many negative outcomes that await dropouts: “Dropouts are much more likely than their peers who graduate to be unemployed, living in poverty, receiving public assistance, in prison, on death row, unhealthy, divorced, and single parents with children who drop out from high school themselves” (Bridgeland et al., 2006:i). These unadjusted differences between dropouts and non-dropouts are well known. This study asks whether dropout *causes* such negative outcomes, criminal offending in particular, or whether, in fact, those individuals who drop out would have the same outcomes regardless of whether they receive a diploma. The suitable policy response to dropout as a causal event is quite different from the response to dropout as a non-causal marker. This study also seeks to determine whether such negative outcomes are long-lasting, diverting an individual onto a worse life trajectory. If so, then dropout may be considered a turning point.

Assessing the causality of dropout as an event is particularly challenging because of the strong selection process leading up to dropout. Evidence of disengagement from school among future dropouts is present as early as first grade (Alexander et al., 2001; Ensminger and Slusarick, 1992). Determining whether dropout is causal requires that the preceding selection process be held constant so that similar individuals may be compared. In order for the effect of the dropout event to be identified, there must be some degree of random variation in dropout conditional on disengagement from school. I address this task using matching models which allow a direct assessment of the comparability of dropouts and their matched comparisons.

This research brings together theoretical perspectives and research methods from several different sources in an effort to provide new insights on several different research fronts. I link Haviland and Nagin's (2005) trajectory model matching method and Rosenbaum and Rubin's (1983,1985) propensity score matching method to the identification of turning points, a central concept of the life course perspective. I contribute to the literature on high school dropout by applying a new method to the old question of the dropout-delinquency link. I also place the question in the context of the life course perspective. Finally, I contribute to the much slower-advancing literature on the measurement of criminal behavior by employing a graded response model to generate a scale of offending that incorporates both seriousness and frequency. All analyses are conducted with these generated scales and variety scores in order to determine if scales derived from graded response models better reflect latent offending propensity, yielding better balance on pre-dropout characteristics. For the remainder of this chapter I discuss the theoretical, research, and policy implications of this research, and suggest directions for future studies.

THEORETICAL AND RESEARCH IMPLICATIONS

High school dropout

If high school dropout indeed causes a change in offending trajectories, it can be understood under two general criminological perspectives. Finn's (1989) frustration-self-esteem hypothesis, based on strain theory, suggests that dropout causes reduced offending through alleviating stress caused by failure in the school setting. On the other hand, his participation-identification hypothesis suggests that dropout exacerbates

offending by severing social bonds. Both of these hypotheses suggest that dropout is a causal event and prior research can be cited in support of both.

Finn also suggested a third option: “a youngster’s leaving school before graduation may be just one more event, albeit a conspicuous event, in a chain that may have begun years before” (1989:118). Under this scenario, high school dropout is not causal, but simply an indication of the process of disengagement that preceded it. Several other theories make the same prediction, but locate the causes of both dropout and delinquency earlier in the life course (Gottfredson and Hirschi, 1990; Newcomb and Bentler, 1988; Jessor and Jessor, 1977; Donovan and Jessor, 1985; Donovan et al., 1988).

Very large pre-dropout differences between dropouts and non-dropouts are observed, and the success of each matching method in this analysis rests on the extent to which bias between dropouts and non-dropouts is eliminated. For the one method that successfully eliminates pre-existing differences between these two groups, no discernable post-dropout differences in offending are detected. Neither strain theory nor control theory is supported by these results. Rather, these results support the conclusion that the dropout event does not have a causal effect on offending. Ultimately, because dropouts and non-dropouts are balanced on so many characteristics, this analysis reduces the large differences between the two groups to the single act of receiving a diploma. Given equivalent characteristics but for a diploma, there are no discernable post-dropout differences in offending.

Because a large number of pre-dropout characteristics are used to create propensity scores, it is not possible to pinpoint which factors are most instrumental in creating balance between dropouts and non-dropouts. As modeled, propensity score

matching does not illuminate the process of disengagement from school. It would be fruitful to develop more parsimonious age-graded matching models in order to better understand which pre-dropout differences are the most effective in creating balance.

It may also be fruitful to conduct the analysis within certain high-dropout populations such as residents of inner-cities, and minorities. Life course theory suggests that the effect of life events may vary with social context and this is particularly likely when rates of dropout differ dramatically in different populations. In inner-city areas where dropout rates reach up to 50 percent, the benefits that accrue to high school graduates may dramatically differ from dropouts. On the other hand, because dropout is so prevalent in such areas, the relative disadvantage of lacking a diploma may be diminished. In this study, no significant effect of first-time dropout between the ages of 16 and 18 is observed *on average* but there may be certain contexts, certain populations, or certain types of dropout for which there is a discernable effect. In addition, while there may be no average effect on offending, dropout may cause other negative outcomes such as substance use, decreased employment opportunities, or poor family outcomes. The analyses employed here can serve as a template to examine any of these other outcomes.

Trajectory model matching and turning points

I suggest in this dissertation that Haviland and Nagin's (2005) matching model based on trajectories is an ideal method for identifying turning points. Although this endeavor was not successful in identifying the effect of dropout on offending using pre-dropout trajectories of offending, I would suggest that under certain conditions, it remains the best method for identifying turning points.

In order for trajectory model matching to be successful, the following conditions must be met. First, the developmental process of selection into the turning point event must be well-understood. Second, this process must be measurable over a certain period of time preceding the turning point. Third, theoretically, there should be the expectation that the effect of the turning point differs based on different patterns of development. If not, there is no motivation to estimate group-average treatment effects. Finally, there must be a certain age-range of interest for which the turning point is particularly salient.

These conditions reveal why trajectory model matching does not achieve balance in this study. Selection into dropout is fairly well understood. It is suggested that dropout is the end result of a process of disengagement from school. The reason trajectories of delinquency do not balance dropouts and non-dropouts is because delinquency is just one part of the selection mechanism. There are two logical extensions of this analysis. First, one could match dropouts to non-dropouts using delinquency trajectory membership, and other strong risk factors for dropout such as middle school grades. Second, one could model trajectories of disengagement from school in order to assess the effect of dropout on post-dropout outcomes. If disengagement from school is the primary explanation for high school dropout, then matching individuals based on patterns of disengagement from school ought to create balance on the characteristics that differ between dropouts and non-dropouts. This kind of analysis is not possible in the current dataset because measures of school attachment are not repeated for every wave.

Propensity score matching

The extent of balance achieved by propensity score matching in this study is striking. Standardized bias in the strongest predictors of dropout decreases dramatically,

and 99 percent of the 314 variables are balanced according to the standardized bias statistic. One disadvantage of this analysis is that it does not allow discernment of which variables produce pre-dropout balance. A more parsimonious propensity score model which produced the same degree of balance could shed light on the most important pre-dropout differences. Another disadvantage of standard propensity score matching is that it does not allow estimation of group-average treatment effects. Nevertheless, this analysis confirms that when appropriate counterfactuals for dropouts are identified among non-dropouts, post-dropout differences are indistinguishable from zero.

These results could be challenged based on the possibility that unobserved characteristics distinguish between dropouts and non-dropouts, but identification of such characteristics would likely lead to even smaller estimates rather than larger ones. Furthermore, because variables from so many different domains are included, it is difficult to suggest a construct that is not represented among the observed variables, or at least highly correlated with one or more of the observed variables.

Another challenge to the results could come from validity of the dropout measure. Because the distribution of propensity scores among non-dropouts is so thin for high propensity levels, it is important that dropout is accurately measured among these individuals. To address this concern, I augmented self-reported dropout measures with official dropout measures from high school transcripts. As previously noted, transcripts were available for a majority, but not all of the sample. Around 70 percent of the sample has valid transcripts. Using this measure changes the prevalence of age 16 to 18 dropout in my sample to 12 percent from 9 percent. Of the 91 percent of the sample who are counted as non-dropouts according to self-reports, about 3 percent are identified as

dropouts by the most recent school of record. If indeed 3 percent of the apparent non-dropouts are in fact dropouts, then estimates using solely self-reported dropout would be biased downward because falsely measured non-dropouts would likely be matched to correctly-measured dropouts. In fact, my dropout effect estimates increase by 10 to 20 percent when using the combined measure. If the inaccuracy of self-reported education is the same among those for whom no transcript is available, then actual effect sizes could be even larger. On the other hand, if official reports of dropout are inaccurate, then false dropouts are compared to non-dropouts, which would bias the estimates downward. It is impossible to arbitrate between these two measures without a different source of information.

Item response theory

All analyses are conducted with variety scales of offending and an estimated latent offending scale based on a graded response model. It was hypothesized that if greater balance was achieved using IRT scores then it would suggest that variety scales do not accurately reflect latent criminality. However, if the balance is about the same using the two methods, then variety scores are indeed the best simple measure of latent criminality available. Because balance was not markedly better using IRT scores I confirm that variety scores are the best alternative to simple dichotomous or frequency offending scales because they more closely approximate latent criminality rather than petty offending. This confirms the assertions of Hindelang and colleagues (1981). A variety score measure of offending is the most efficient summary score of offending.

Contrary to Osgood and colleagues' (2002) assertions, IRT scales provide no advantage in this analysis over variety scores. In fact in this analysis, IRT scores held

several disadvantages. First, because the graded response model required information on frequency of offending and variety scales require only information on whether each offense was committed, pre-wave 1 summative scales were produced for variety scores, but not IRT scores. Second, although effect sizes were roughly comparable in the two analyses, interpretation of the IRT results is more difficult because the estimated scale is not easily translated to actual offense counts.

POLICY IMPLICATIONS

The dropout problem has recently garnered increased attention in the United States. In the first half of 2006 a major report on dropout was released by the Gates Foundation (Bridgeland et al., 2006), followed by a cover story in Time magazine (Thornburgh, 2006). In April of 2006, the Gates Foundation awarded the Chicago Public Schools system \$21 million to combat high school dropout (Mendell, 2006). The Gates Foundation report attempted to determine the causes of dropout by asking dropouts themselves. In turn, it suggests several strategies for reducing dropout based on the reports of high school dropouts.

This analysis suggests that the consequences of first-time dropout between the ages of 16 and 18 are determined not by dropout itself but by pre-16 differences between dropouts and non-dropouts, some of which are evident very early in the life course. While further manipulation of the propensity score model may allow for a better understanding of the selection process that eventually leads to high school dropout, the current analysis has implications for the policy recommendations of the Gates Foundation report.

The Gates report suggested a number of changes to schools themselves that may result in less dropout: improved teaching and curriculum, improved access to supports for struggling students, and building a school climate that fosters academics. The main focus of the \$21 million grant to the Chicago Public School system is improving curriculum in 14 high schools. This analysis would suggest that efforts targeted so late in the process face an uphill battle. Many of the strongest differences between dropouts and non-dropouts in this study reside in pre-high school characteristics such as middle school grades and academic aptitude. Also, there were large pre-dropout differences in non-school characteristics such as socio-economic status indicators and family characteristics. Future dropouts in this study had lower assessments of the school environment and were less attached to school, but the biggest differences were in other characteristics. This would suggest that attempts to reverse the process of disengagement that target high school curriculum may arrive too late in the process of disengagement to reverse prior problems.

The Gates report also suggested that an early warning system should be put into place that would flag certain students who exhibit risk factors for dropout early in their school careers. To the extent that appropriate flags for school disengagement are identified early, this would appear to be a promising first step in combating dropout. Of course, once the at-risk student is identified, effective interventions must be applied in order to reverse the process of disengagement.

Another avenue of education reform which has been considered by several state senates in the early part of 2006, and which is recommended by the Gates Foundation, is to increase compulsory school age requirements. Absent other changes, I am skeptical of

the effectiveness of this kind of strategy to ameliorate post-dropout deficits. This research suggests that given a certain constellation of risk factors for dropout, it makes no difference whether a diploma is received or not – the same level of delinquency will be observed. If students are simply required to remain in the high school, but no efforts are made to ameliorate other problems which predict dropout, I would expect no improvements in post-high school offending regardless of an improved dropout rate.

This research suggests that pre-existing dropout differences determine post-dropout differences. As such, policy interventions that ameliorate pre dropout risk factors ought to decrease offending regardless of whether they actually reduce the rate of dropout. Dropout prevention programs can be beneficial on a number of levels if they address the conditions that lead to dropout. This research cannot finely distinguish between which of the pre-16 differences account for large post-dropout differences, nor can it speak to non-offending post-dropout differences. Future research employing the same methods can address these issues.

FUTURE RESEARCH

This research suggests a number of fruitful avenues for future research. First, trajectory model matching can be applied in numerous ways. To date, all applications of trajectory model matching have used trajectories of violence or delinquency. The model is very flexible and can be applied to numerous types of problems. Ideally, in order to assess the effect of dropout using this method, trajectories of disengagement from school will be estimated. If disengagement from school is in fact the primary component of the selection process, then this method ought to better balance dropouts and non-dropouts.

Trajectories of delinquency ought to be used to measure potential turning points that are strongly related to developmental patterns of delinquency such as gang membership, arrest, or violent victimization.

Alternately, dropouts and non-dropouts can be matched on pre-dropout trajectories of offending and a handful of strong dropout predictors such as middle school grades, parental education, and whether the student has ever been retained in school. If a more parsimonious matching model is able to create adequate pre-dropout balance, it has the added advantage of illuminating which of the pre-dropout differences are the most important determinants of post-dropout differences. Because this analysis points to the importance of the process of disengagement from school, analyses which illuminate this process are particularly desirable.

When possible, future research concerning turning points in the life course should employ trajectory model matching to identify counterfactuals among the untreated population, and evaluate patterns of development after the turning point event in order to assess whether the effect persists over time, indicating that a shift in trajectory has indeed occurred. This method is the most consistent with the definition of a turning point. Individuals are followed on a particular trajectory until a particular age, and then patterns of change are observed among individuals who experience the turning point and matched counterfactuals in order to determine if a lasting change occurs. If so, the event is considered a turning point. Methods for identifying turning points that rely solely on within-individual variation do not provide enough information to determine whether the analyzed events are turning points.

In this analysis, propensity score matching was used to identify a comparison sample of non-dropouts that, when combined with dropouts, produced a sample in which dropout was essentially a random process. This very small random component of dropout was not found to be a casual predictor of offending after age 18. I conclude that dropout itself has no added causal impact on offending, but that pre-dropout differences explain post-dropout differences in offending. As such, it is most beneficial to direct attention to turning points earlier in the life course to pinpoint the sources of disengagement from school that eventually lead to dropout so that the process of disengagement from school can be interrupted, leading to less dropout and less offending.

Table 1. Characteristics of final sample and excluded observations (unweighted, Wave 1, unless otherwise noted)

Variables	Final Sample (N=2990)	Missing Observations (N=293)	Pre-16 dropout (N=187)
Age	13.4 (.6)	13.2 (.6)	13.2 (.6)
Male	.51	.58	.52
Hispanic	.21	.22	.27
Black	.25	.30	.28
Delinquency T ₁ *	.48 (.97)	.55 (1.11)	.75 (1.33)
Delinquency T ₂ *	.47 (.92)	.54 (1.06)	.78 (1.34)
Delinquency T ₃	.39 (.74)	.40 (.77)	.44 (.77)
Middle School Grades*	5.7 (1.7)	5.4 (1.6)	4.9 (1.8)
ASVAB: Math Knowledge*	-.08 (.86)	-.22 (.75)	-.43 (.75)
Computer in home*	.53	.45	.37
Smokes*	.26	.24	.34
Biological parents*	.53	.44	.42

* Statistically significant ($p < .05$) between-group variation using either an ANOVA or chi-square test.

Table 2. Mean delinquency variety and IRT scores

	N	T ₋₃	T ₋₂	T ₋₁	T ₁
Variety	2990	.39 (.74)	.47 (.95)	.49 (.99)	.40 (.81)
IRT	1633	.31 (.60)	.37 (.65)	.33 (.64)	.31 (.56)

*T₋₁ refers to the last pre-16 observation, T₋₂ refers to the second to last pre-16 observation, etc., and T₁ is the first post-16 observation.

Table 3. Baseline comparison of dropouts and non-dropouts on offending between the ages of 16 and 18 (T₁), and each of three waves thereafter (T₂ to T₄), (Number of valid observations in parentheses).

	T ₁	T ₂	T ₃	T ₄
Delinquency variety				
Dropouts	.74 (367)	.69 (367)	.58 (356)	.48 (178)
Non-dropouts	.36 (2623)	.32 (2623)	.31 (2573)	.23 (1244)
U-PATE	.38**	.38**	.27**	.24**
Delinquency IRT				
Dropouts	.52 (198)	.42 (192)	.30 (176)	n/a
Non-dropouts	.28 (1435)	.25 (1406)	.19 (1326)	n/a
U-PATE	.24**	.18**	.12*	n/a

*p<.05, **p<.01

Table 4. Characteristics of four groups in delinquency variety trajectories (wave 1 unless otherwise noted, N=2990)

Variable	Group 1: Abstainers	Group 2: Teen onset	Group 3: Decreasing	Group 4: Increasing
Proportion of population	.42	.13	.35	.10
Expected delinquency at 12	.00	.00	.52	1.50
Expected delinquency at 13	.00	.00	.40	1.96
Expected delinquency at 14	.00	.49	.26	2.21
Expected delinquency at 15	.00	1.04	.15	2.22
Middle School Grades	6.22	5.49	5.59	4.90
ASVAB: math knowledge	.13	-.03	-.02	-.07
Computer in house	.62	.61	.56	.52
Smokes	.12	.24	.40	.67
Both biological parents	.62	.53	.50	.45
Dropout at 16-18	.07	.12	.14	.27

Table 5. Characteristics of four groups in delinquency IRT score trajectories (wave 1 unless otherwise noted, N=1633)

Variable	Group 1: Abstainers	Group 2: Pre-teen	Group 3: Teen Onset	Group 4: High
proportion of population	.52	.16	.16	.16
Expected delinquency at 12	.00	.53	.00	.82
Expected delinquency at 13	.00	.75	.02	1.09
Expected delinquency at 14	.00	.24	.63	1.19
Expected delinquency at 15	.00	.00	.87	1.09
Middle School Grades	6.26	5.40	5.41	5.15
ASVAB: math knowledge	.13	-.09	.04	-.06
Computer in house	.63	.54	.59	.53
Smokes	.11	.35	.26	.52
Both biological parents	.62	.47	.49	.47
Dropout at 16-18	.08	.13	.13	.20

Table 6. Within-trajectory balance on prior delinquency and selected high-bias background characteristics, classify-analyze method (N=2990)

Variable	Group	Dropout at 16-18		Difference	P-value
		Yes	No		
Delinquency variety T ₁	All	.94	.78	.16	.01
	Abstain	.00	.00	.00	-
	Teen onset	1.43	1.46	-.03	.81
	Decreasing	.35	.30	.04	.46
	Increasing	3.03	2.27	.76	<.01
Delinquency variety T ₂	All	.93	.77	.16	.01
	Abstain	.00	.00	.00	-
	Teen onset	.66	.49	.17	.41
	Decreasing	.82	.60	.22	<.01
	Increasing	2.62	2.31	.30	.12
Delinquency variety T ₃	All	.73	.69	.04	.36
	Abstain	.00	.00	.00	-
	Teen onset	.00	.02	-.02	.02
	Decreasing	.98	.79	.19	.01
	Increasing	1.47	1.59	-.12	.43
Middle School Grades	All	4.49	5.82	-1.33	<.01
	Abstain	5.01	6.31	-1.30	<.01
	Teen onset	4.30	5.65	-1.35	<.01
	Decreasing	4.51	5.76	-1.25	<.01
	Increasing	3.81	5.30	-1.49	<.01
ASVAB: math knowledge	All	-.41	.07	-.48	<.01
	Abstain	-.39	.17	-.55	<.01
	Teen onset	-.56	.05	-.61	<.01
	Decreasing	-.45	.05	-.50	<.01
	Increasing	-.29	.00	-.29	<.01
Computer in house	All	.34	.62	-.28	<.01
	Abstain	.32	.64	-.33	<.01
	Teen onset	.37	.64	-.27	<.01
	Decreasing	.36	.60	-.24	<.01
	Increasing	.31	.60	-.28	<.01
Youth smokes	All	.47	.33	.14	<.01
	Abstain	.24	.11	.13	<.01
	Teen onset	.43	.22	.22	<.01
	Decreasing	.55	.38	.18	<.01
	Increasing	.68	.66	.01	.84
Both biological parents	All	.33	.56	-.23	<.01
	Abstain	.34	.64	-.30	<.01
	Teen onset	.23	.58	-.34	<.01
	Decreasing	.33	.53	-.20	<.01
	Increasing	.36	.49	-.13	.06

* Within trajectory groups, *p*-values are based on simple two-sample, two-sided t-tests, assuming unequal variance. Variance for the aggregate difference is based on 500 bootstrap iterations.

Table 7. Within-trajectory balance on prior delinquency and selected high-bias background characteristics, expected value method (N=2990)

Variable	Group	Dropout at 16-18		Difference	P-value
		Yes	No		
Delinquency variety T ₁	All	.94	.73	.21	<.01
	Abstain	.00	.00	.00	1.00
	Teen onset	1.19	1.09	.10	.08
	Decreasing	.41	.34	.07	.08
	Increasing	2.81	2.06	.76	<.01
Delinquency variety T ₂	All	.93	.72	.21	<.01
	Abstain	.00	.00	.00	1.00
	Teen onset	.61	.46	.15	.03
	Decreasing	.78	.51	.26	<.01
	Increasing	2.47	2.09	.38	<.01
Delinquency variety T ₃	All	.73	.66	.07	.10
	Abstain	.00	.00	.00	1.00
	Teen onset	.01	.02	-.01	.36
	Decreasing	.88	.67	.20	<.01
	Increasing	1.45	1.48	-.03	.63
Middle School Grades	All	4.49	5.85	-1.35	<.01
	Abstain	5.01	6.32	-1.31	<.01
	Teen onset	4.45	5.78	-1.33	<.01
	Decreasing	4.49	5.82	-1.33	<.01
	Increasing	3.90	5.33	-1.43	<.01
ASVAB: math knowledge	All	-.41	.08	-.49	<.01
	Abstain	-.38	.17	-.55	<.01
	Teen onset	-.58	.07	-.65	<.01
	Decreasing	-.42	.06	-.49	<.01
	Increasing	-.33	-.01	-.32	<.01
Computer in house	All	.34	.62	-.28	<.01
	Abstain	.32	.64	-.33	<.01
	Teen onset	.35	.63	-.28	<.01
	Decreasing	.34	.61	-.27	<.01
	Increasing	.34	.60	-.26	<.01
Youth smokes	All	.47	.32	.15	<.01
	Abstain	.24	.11	.13	<.01
	Teen onset	.42	.19	.23	<.01
	Decreasing	.52	.35	.16	<.01
	Increasing	.69	.60	.09	<.01
Both biological parents	All	.33	.57	-.24	<.01
	Abstain	.34	.64	-.30	<.01
	Teen onset	.25	.59	-.34	<.01
	Decreasing	.32	.55	-.23	<.01
	Increasing	.36	.49	-.13	<.01

* Within trajectory groups, *p*-values are based on simple two-sample, two-sided t-tests, assuming unequal variance. Variance for the aggregate difference is based on 500 bootstrap iterations.

Table 8. Balance on prior delinquency and selected high-bias background characteristics, kernel matching with trajectory probabilities (N=2990)

Variable	Group	Dropout at 16-18		Difference	P-value
		Yes	No		
Delinquency variety T ₁	All	.94	.85	.09	.02
Delinquency variety T ₂	All	.93	.90	.03	.51
Delinquency variety T ₃	All	.73	.78	-.03	.45
Middle School Grades	All	4.49	5.70	-1.21	<.01
ASVAB:math knowledge	All	-.41	.08	-.49	<.01
Computer in house	All	.34	.61	-.27	<.01
Youth smokes	All	.47	.35	.08	<.01
Both biological parents	All	.33	.55	-.22	<.01

* Standard errors for these estimates were obtained through 500 bootstrap replications of the matching model, not the trajectory model. Therefore, these standard errors do not take into account uncertainty in trajectory probabilities. Haviland and Nagin (2005) estimate that this uncertainty inflates standard errors by up to 17%.

Table 9. Group-average treatment effects (GATE) and population-average treatment effects on the treated (PATT), variety score models

	T ₁	T ₂	T ₃	T ₄
Group 1 ATE, Classify-Analyze	.16*	.13	.19*	.04
Group 1 ATE, Expected Value	.16*	.13*	.19*	.04
Group 2 ATE, Classify-Analyze	.00	.08	-.02	.19
Group 2 ATE, Expected Value	.08	.16*	.04	.25*
Group 3 ATE, Classify-Analyze	.17*	.29*	.13	.23
Group 3 ATE, Expected Value	.16*	.25*	.13*	.21*
Group 4 ATE, Classify-Analyze	.43*	.27	.18	.03
Group 4 ATE, Expected Value	.48*	.36*	.23*	.10
PATT, Classify-Analyze	.20*	.22*	.14	.13
PATT, Expected Value	.22*	.24*	.15*	.15
PATT, Kernel	.15*	.14*	.08	.08

*p<.05; Standard errors for GATE were calculated using simple two-sided t-tests. Standard errors for PATT calculated from 500 bootstrap iterations of trajectory models.

Table 10. Group-average treatment effects (GATE) and population-average treatment effects on the treated (PATT), IRT models

	T ₁	T ₂	T ₃
Group 1 ATE, Classify-Analyze	.09	.13*	.19*
Group 1 ATE, Expected Value	.09*	.13*	.19*
Group 2 ATE, Classify-Analyze	.17	-.09	-.12
Group 2 ATE, Expected Value	.15*	-.04	-.08*
Group 3 ATE, Classify-Analyze	.19	.07	-.08
Group 3 ATE, Expected Value	.18*	.06	.01
Group 4 ATE, Classify-Analyze	.24	.25	.14
Group 4 ATE, Expected Value	.26*	.21*	.06
PATT, Classify-Analyze	.16*	.10	.07
PATT, Expected Value	.16*	.11	.07
PATT, Kernel	.11*	.06	.07

*p<.05; Standard errors for GATE were calculated using simple two-sided t-tests.

Standard errors for PATT calculated from 500 bootstrap iterations of trajectory models.

Table 11. Frequency of propensity scores by dropout status

	0-.09	.1-.19	.2-.29	.3-.39	.4-.49	.5-.59	.6-.69	.7-.79	.8-.89	.9-1
Non-dropout	1966	298	132	82	61	35	21	16	9	3
Dropout	70	53	36	44	40	24	27	29	33	21

Table 12. Balance on prior delinquency and selected high-bias background characteristics, propensity score method (N=2990)

Variable	Group	Dropout at 16-18		Difference	P-value
		Yes	No		
Delinquency variety T ₁	All	.94	.83	.11	.28
Delinquency variety T ₂	All	.93	.83	.10	.23
Delinquency variety T ₃	All	.73	.71	.02	.80
Middle School Grades	All	4.49	4.74	-.25	.04
ASVAB:math knowledge	All	-.41	-.37	-.04	.52
Computer in house	All	.34	.34	.00	.91
Youth smokes	All	.47	.43	.04	.30
Both biological parents	All	.33	.42	-.09	.02

* Standard errors were obtained using 500 bootstrap replications.

Table 13. Propensity score estimates of dropout effect (PATT), variety score models

	T ₁	T ₂	T ₃	T ₄
PATT, Propensity Score Matching	.13	.18	.07	.12
Lower Limit (95% confidence level)	-.06	-.03	-.07	-.06
Upper Limit (95% confidence level)	.32	.38	.22	.29

* Standard errors obtained from 500 bootstrap replications.

Table 14. Percent bias reduction from matching methods for most biased characteristics (CA=classify-analyze, EV=expected value, K=kernel, PS=propensity score)

Variable	Unadjusted Values		Bias	% bias reduction by method			
	Dropout Yes	Dropout No		CA	EV	K	PS
school achievement/aptitude							
middle school grades	4.49	5.99	-87	12%	10%	20%	84%
ASVAB: math knowledge	-0.41	0.11	-64	7%	6%	5%	93%
suspended	0.39	0.15	57	16%	14%	12%	85%
gifted	0.06	0.25	-56	11%	9%	10%	88%
ASVAB: paragraph comprehension	-0.35	0.10	-53	8%	7%	-1%	88%
ASVAB: arithmetic reasoning	-0.31	0.10	-46	5%	5%	-6%	92%
ASVAB: assembly of objects	-0.29	0.08	-45	8%	7%	9%	99%
ASVAB: coding speed	-0.29	0.08	-44	8%	7%	11%	95%
# weekdays/week does homework	2.93	3.62	-43	13%	12%	35%	96%
ASVAB: mechanical comprehension	-0.29	0.09	-43	7%	6%	-1%	100%
ASVAB: general science	-0.27	0.08	-42	2%	3%	-10%	99%
ASVAB: word knowledge	-0.29	0.08	-42	4%	3%	-8%	99%
ASVAB: numerical operations	-0.26	0.08	-40	8%	7%	3%	78%
socio-economic status							
computer in house	0.34	0.63	-61	3%	3%	8%	99%
parent received government aid	0.69	0.42	58	7%	6%	11%	98%
# of assets	2.17	3.06	-55	0%	0%	1%	96%
gross household income	33930	55040	-55	1%	2%	10%	90%
outside of house: nice	0.42	0.67	-52	3%	3%	7%	95%
inside of house: nice	0.44	0.68	-51	5%	5%	13%	87%
parent has pension / retirement account	0.35	0.59	-50	0%	0%	7%	96%
mother is dropout	0.33	0.13	49	4%	3%	1%	71%
mother went to college	0.26	0.48	-48	-1%	-1%	10%	93%
received food stamps / WIC	0.56	0.33	47	7%	6%	13%	93%
lives in house	0.59	0.80	-47	4%	4%	16%	98%
family received medicaid	0.39	0.19	47	6%	5%	8%	91%
outside of house: poor	0.18	0.05	43	0%	0%	21%	98%
father went to college	0.20	0.39	-43	4%	3%	13%	98%
family received afdc	0.38	0.20	42	7%	6%	13%	99%
delinquency							
youth smokes	0.47	0.24	49	40%	35%	65%	84%
trajectory 4 probability	0.22	0.08	47	84%	71%	98%	81%
delinquency count 2 waves prior to age 16	0.93	0.41	47	69%	60%	94%	80%
used marijuana	0.21	0.06	45	27%	24%	52%	89%
attacked others (wave 1)	0.23	0.08	45	38%	33%	71%	83%
delinquency count in wave prior to age 16	0.94	0.43	44	68%	59%	82%	78%
trajectory 1 probability	0.26	0.44	-43	100%	86%	100%	72%
committed property offense (wave 1)	0.30	0.14	40	66%	58%	97%	48%
peers							
prosocial peer scale	23.00	25.92	-53	17%	16%	41%	97%

antisocial peer scale	1.53	0.82	49	13%	12%	36%	95%
peers use drugs	2.92	3.44	-47	20%	18%	42%	76%
peers in gang	0.32	0.13	48	19%	17%	36%	88%
peers cut classes	2.78	3.27	-44	14%	13%	27%	95%
peers smoke	2.59	3.09	-42	18%	17%	53%	93%
peers attend church	1.57	2.01	-40	7%	6%	18%	89%
family							
both biological parents	0.33	0.59	-56	12%	11%	18%	68%
father knows friend's parents	1.23	1.80	-46	21%	18%	40%	100%
mother was teenage mom	0.26	0.11	40	1%	1%	4%	99%

Figure 1. Delinquency variety trajectories, ages 11-15.5 (N=2990)

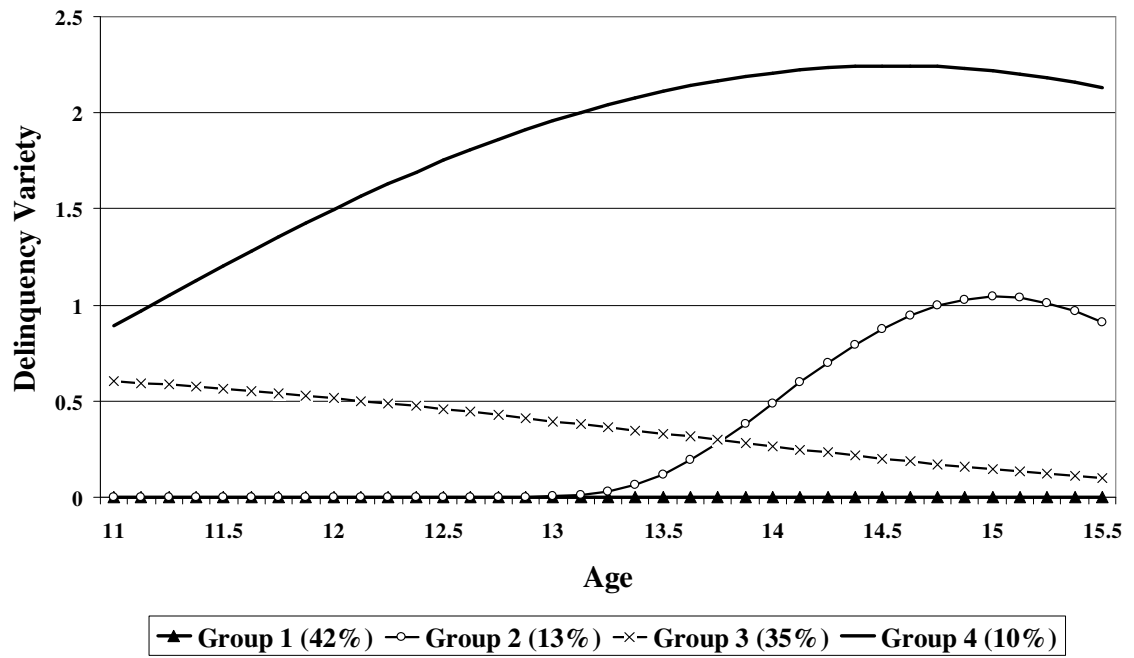


Figure 2. Delinquency IRT score trajectories (N=1633)

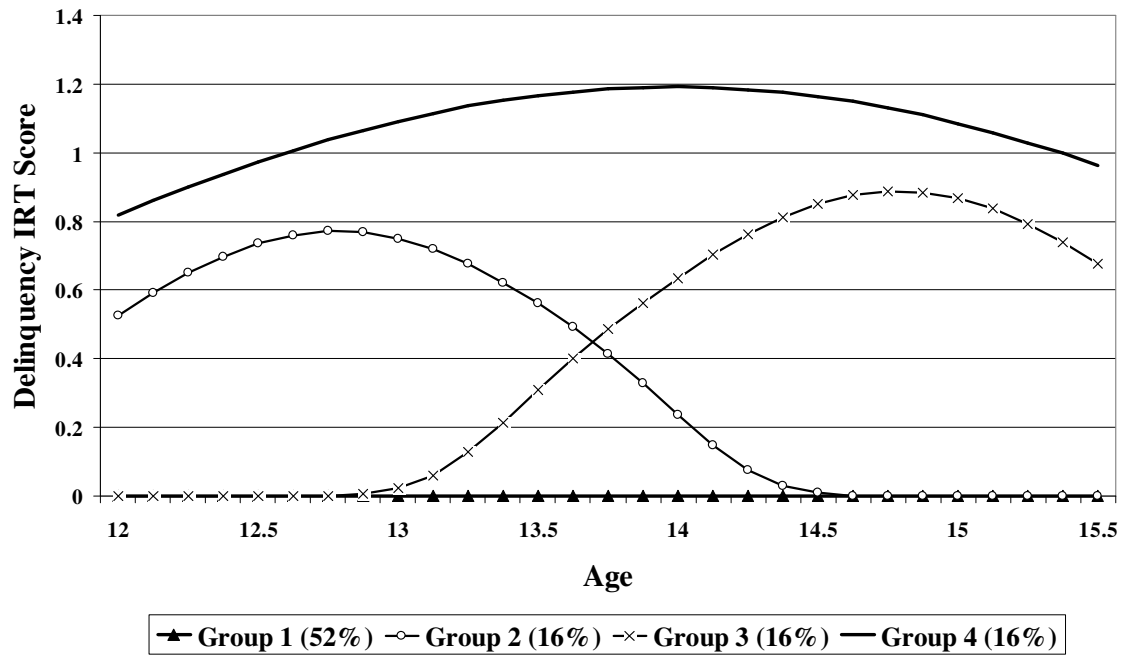
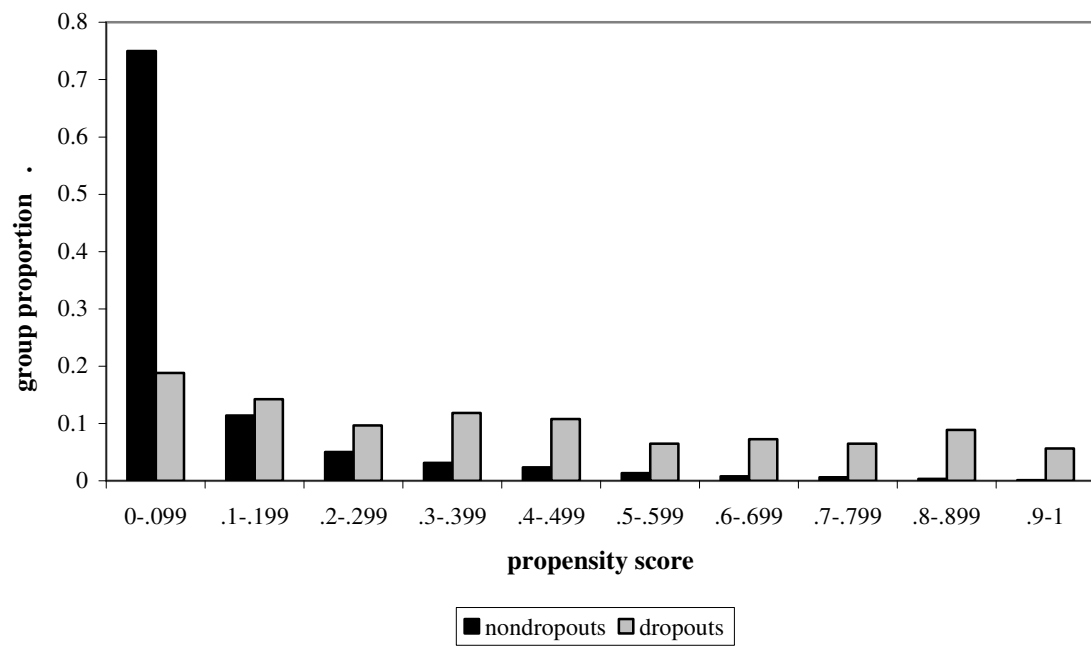


Figure 3. Distribution of propensity scores by dropout status



Appendix A. Variables used to compare dropouts and non-dropouts. (Note: Variables refer to wave 1 unless indicated otherwise.)

Variable Name	N	Mean (S.D.)	Min	Max	Description
<u>Variables from trajectory model</u>					
dcpre4	1552	.33 (.63)	0	5	delinquency count 4 waves before age 16
dcpre3	2990	.39 (.74)	0	6	delinquency count 3 waves before age 16
dcpre2	2990	.47 (.95)	0	6	delinquency count 2 waves before age 16
dcpre1	2990	.49 (.99)	0	6	delinquency count 1 wave before age 16
dcpos1	2990	.40 (.81)	0	6	average delinquency count between ages 16-18
dcpos2	2990	.36 (.87)	0	6	delinquency count first wave after age 18
dcpos3	2929	.34 (.86)	0	6	delinquency count 2 waves after age 18
dcpos4	1422	.26 (.73)	0	6	delinquency count t2 waves after age 18
dpre3	1633	.31 (.60)	0	2.75	delinquency IRT score 3 waves before age 16
dpre2	1633	.37 (.65)	0	2.86	delinquency IRT score 2 waves before age 16
dpre1	1633	.33 (.64)	0	3.41	delinquency IRT score 1 wave before age 16
dpos1	1633	.31 (.56)	0	4.18	delinquency IRT score between ages 16-18
dpos2	1598	.27 (.56)	0	2.92	delinquency IRT score first wave after age 18
dpos3	1502	.20 (.49)	0	3.77	delinquency IRT score 2 waves after age 18
dgposc1	2990	.12	0	1	dropout between ages 16 and 18
dgposc2	2896	.13	0	1	dropout first wave after age 18
dgposc3	2792	.15	0	1	dropout second wave after age 18
dgposc4	1359	.15	0	1	dropout third wave after age 18
grp1prb	2990	.42 (.44)	0	.95	probability of membership in trajectory group 1 (variety score model)
grp2prb	2990	.13 (.24)	0	.98	probability of membership in trajectory group 2 (variety score model)
grp3prb	2990	.35 (.38)	0	1	probability of membership in trajectory group 3 (variety score model)
grp4prb	2990	.10 (.25)	0	1	probability of membership in trajectory group 4 variety score model)
grp1prb	1633	.52 (.46)	0	.96	probability of membership in trajectory group 1 (IRT model)
grp2prb	1633	.16 (.29)	0	.99	probability of membership in trajectory group 2 (IRT model)
grp3prb	1633	.16 (.29)	0	.97	probability of membership in trajectory group 3 (IRT model)
grp4prb	1633	.16 (.31)	0	1	probability of membership in trajectory group 4 IRT model)
<u>Demographic Variables</u>					
male	2990	.51	0	1	sex
hisp	2990	.12	0	1	hispanic ethnicity
white	2990	.73	0	1	race: white

Variable Name	N	Mean (S.D.)	Min	Max	Description
black	2990	.15	0	1	race: black
other	2990	.12	0	1	race: other
age1	2990	13.4 (.6)	12.2	15.3	age at first interview
ur_rur0	2990	.20	0	1	urban residence at age 12
ur_rur1	2990	.27	0	1	urban residence at wave 1
ur_urb0	2990	.58	0	1	rural residence at age 12
ur_urb1	2990	.68	0	1	rural residence at wave 1
ur_unk0	2990	.22	0	1	unknown residence at age 12
ur_unk1	2990	.04	0	1	unknown residence at wave 1
msacty0	2990	.23	0	1	central city MSA at age 12
msacty1	2990	.26	0	1	central city MSA at wave 1
msancc0	2990	.42	0	1	not central city MSA at age 12
msancc1	2990	.53	0	1	not central city MSA at wave 1
msanot0	2990	.14	0	1	not in MSA at age 12
msanot1	2990	.20	0	1	not in MSA at wave 1
msaunk0	2990	.21	0	1	MSA unknown at age 12 (all known at wave 1)
dwelh	2990	.78	0	1	dwelling place: house
dwell	2990	.15	0	1	dwelling place: apartment
dwell	2990	.08	0	1	dwelling place: other
moves0	2685	.26 (.65)	0	9	moves from age 12 to interview date
mobility0	2685	.22 (.71)	0	15	moves/year from age 12 to interview date
northel	2990	.18	0	1	northeast region
northc1	2990	.26	0	1	north central region
south1	2990	.34	0	1	south region
west1	2990	.21	0	1	west region
<u>Household characteristics</u>					
hhbb1	2990	.56	0	1	lives with biological parents
hhbs1	2990	.14	0	1	lives with one biological, one stepparent
hhbn1	2990	.26	0	1	lives with one biological parent (single)
hhoo1	2990	.04	0	1	other living arrangement
adopted	2990	.01	0	1	adopted
indep1	2989	.01	0	1	lives independently
deadp	2990	.04	0	1	at least one deceased biological parent
fborn	2705	.15	0	1	at least one foreign born parent
hhsz1	2990	4.47 (1.33)	1	10	size of household
mom19	2990	.13	0	1	mother was teenager at birth of first child
headst	2671	.14	0	1	attended head start
noins1	2990	.09	0	1	no health insurance
<u>Family process variables</u>					
cfpar1	2990	.57	0	1	confides first: in parents
cfrel1	2990	.11	0	1	confides first: in other relative
cfnd1	2990	.24	0	1	confides first: in friend
cfoth1	2990	.04	0	1	confides first: in other person
cfnon1	2990	.04	0	1	confides first: in nobody
cfmis1	2990	.00	0	1	confides first: missing
momhigh	2886	3.33 (1.07)	0	4	thinks highly of mother, 0: strongly disagree, 4: strongly agree

Variable Name	N	Mean (S.D.)	Min	Max	Description
momwant	2887	2.67 (1.09)	0	4	wants to be like mother
momenjy	2887	3.22 (.92)	0	4	enjoys mother's company
mattach	2886	9.22 (2.51)	0	12	sum of previous three variables, attachment to mother
dadhigh	2248	3.28 (1.00)	0	4	thinks highly of father, 0: strongly disagree, 4: strongly agree
dadwant	2246	2.68 (1.17)	0	4	wants to be like father
dadenjy	2249	3.18 (1.00)	0	4	enjoys father's company
dattach	2246	9.14 (2.74)	0	12	sum of previous three variables, attachment to father
msupprai	2887	3.14 (.92)	0	4	mother's support: praises for doing well, 0=never, 4=always
msupcrit	2887	2.97 (1.02)	0	4	criticizes ideas, 0=always, 4=never
msuphelp	2888	3.16 (.94)	0	4	helps with important stuff, 0=never, 4=always
msupblam	2888	3.56 (.82)	0	4	blames r for problems, 0=always, 4=never
msupplan	2887	3.48 (.86)	0	4	makes & cancels plans, 0=always, 4=never
msupscale	2885	16.32 (2.97)	0	20	sum of previous five
msup	2888	1.75 (.49)	0	2	how supportive is mother? 0=not very, 1=somewhat, 2=very
mstrict	2877	.56	0	1	is mother strict?
dsupprai	2248	3.00 (1.04)	0	4	father's support: praises for doing well, 0=never, 4=always
dsupcrit	2245	2.98 (1.10)	0	4	criticizes ideas, 0=always, 4=never
dsuphelp	2248	2.91 (1.10)	0	4	helps with important stuff, 0=never, 4=always
dsupblam	2248	3.62 (.80)	0	4	blames r for problems, 0=always, 4=never
dsupplan	2246	3.47 (.91)	0	4	makes & cancels plans, 0=always, 4=never
dsupscale	2241	15.98 (3.30)	1	20	sum of previous five
dsup	2246	1.66 (.56)	0	2	how supportive is father? 0=not very, 1=somewhat, 2=very
dstrict	2241	.61	0	1	is father strict?
mknofr	2888	2.50 (1.01)	0	4	mom knows about close friends, 0=nothing, 1=a little, 2=some things, 3=most things, 4=everything
mknofp	2887	2.13 (1.15)	0	4	friends' parents
mknoww	2885	3.04 (1.10)	0	4	who with when not at home
mknosl	2884	7.66 (2.58)	0	12	social life in general (sum of previous 3)
mknots	2884	2.97 (1.01)	0	4	mom knows about teachers and people at school
dknofr	2249	1.95 (1.17)	0	4	dad knows about close friends, 0=nothing, 1=a little, 2=some things, 3=most things, 4=everything
dknofp	2248	1.75 (1.24)	0	4	friends' parents

Variable Name	N	Mean (S.D.)	Min	Max	Description
dknoww	2246	2.53 (1.31)	0	4	who with when not at home
dknosl	2245	6.23 (3.13)	0	12	social life in general (sum of previous 3) dad knows about teachers and people at school
dknots	2247	2.43 (1.24)	0	4	count of number of items youth sets limits independently or jointly with parents (out of 3 categories)
auton	2814	1.86 (.91)	0	3	days/month youth breaks curfew limits, set to 30 if youth sets own limits
blcurf	2812	2.05 (5.89)	0	30	youth sets curfew limits
mblcurf	2812	.03	0	1	youth sets curfew limits
blmovi	2819	11.58 (14.11)	0	30	days/month youth breaks movie/tv limits
mblmovi	2819	.35	0	1	youth sets movie/tv limits
blpeer	2817	16.30 (14.71)	0	30	days/month youth breaks peer limits
mblpeer	2817	.52	0	1	youth sets peer limits
blimits	2807	29.90 (24.61)	0	90	sum of blcuf, blmovi, blpeer
perm	1061	.04 (.20)	0	2	permissive parent scale: parent responds permissively to limit breaking, missing for youths who set own limits
induct	1061	2.71 (.66)	0	3	inductive parent scale: discuss or take away privilege when limit broken
mfrcomp	1835	2.99 (.98)	0	4	mother-father relationship, mom compromises with dad, 0=never, 4=always
mfrscrm	1834	2.84 (1.09)	0	4	mom screams when angry at dad, 0=always, 4=never
mfrlove	1833	3.27 (1.00)	0	4	expresses love/affection, 0=never, 4=always
mfrcrit	1834	3.28 (.92)	0	4	insults/criticizes, 0=always, 4=never
mfrenc	1839	3.11 (1.04)	0	4	encourages what's important, 0=never, 4=always
mfrblme	1837	3.54 (.79)	0	4	blames dad for problems, 0=always, 4=never
mfrrel	1826	19.04 (3.86)	0	24	sum of previous 6 items, 0=least supportive, 24=most
fmrcomp	1836	2.95 (1.03)	0	4	father-mother relationship, dad compromises with mom, 0=never, 4=always
fmrscrm	1833	3.09 (1.07)	0	4	dad screams when angry at mom, 0=always, 4=never
fmrlove	1831	3.23 (1.02)	0	4	expresses love/affection, 0=never, 4=always
fmrrel	1835	3.41 (.87)	0	4	insults/criticizes, 0=always, 4=never
fmrrenc	1836	3.07 (1.06)	0	4	encourages what's important, 0=never, 4=always
fmrblme	1833	3.59 (.79)	0	4	blames mom for problems, 0=always, 4=never
fmrrel	1826	19.35 (4.17)	0	24	sum of previous 6 items, 0=least supportive, 24=most

Variable Name	N	Mean (S.D.)	Min	Max	Description
allwnce96	2983	.59	0	1	family gave youth an allowance in 1996
allwncei96	2924	149 (272)	0	5000	income from allowance in 1996
<u>Indicators of socio-economic status</u>					
heat1	2987	.97	0	1	heat/electricity in house
computer1	2987	.59	0	1	computer in house
study1	2986	.90	0	1	quiet study area in house
webster1	2985	.96	0	1	dictionary in house
inside of house, according to interviewer:					
innice1	2990	.65	0	1	nice
infair1	2990	.25	0	1	fair
inpoor1	2990	.05	0	1	poor
inunk1	2990	.04	0	1	unknown
outside of house according to interviewer:					
outnice1	2990	.64	0	1	nice
outfair1	2990	.28	0	1	fair
outpoor1	2990	.06	0	1	poor
outunk1	2990	.01	0	1	unknown
med_do	2990	.15	0	1	mom is dropout
med_hs	2990	.33	0	1	mom graduated high school
med_co	2990	.46	0	1	mom went to college
med_un	2990	.06	0	1	mom's education unknown
fed_do	2990	.14	0	1	father is dropout
fed_hs	2990	.34	0	1	father graduated high school
fed_co	2990	.37	0	1	father went to college
fed_un	2990	.15	0	1	father's education is unknown
pov000	2990	.13	0	1	family income under poverty level
pov100	2990	.15	0	1	1-2 times poverty level
pov200	2990	.16	0	1	2-3 times poverty level
pov300	2990	.13	0	1	3-4 times poverty level
pov400	2990	.20	0	1	4+ times poverty level
povunk	2990	.22	0	1	income unknown
income1	2275	52550 (44771)	0	246474	household income
govaid	2700	.45	0	1	did parent receive any govt aid since age 18 or since oldest child was born?
afdc	2990	.22	0	1	afdc received
mcaid	2990	.21	0	1	medicaid received
ssi	2990	.04	0	1	supplementary security income received
fdaid	2990	.36	0	1	food stamps, WIC received
afdc5y	2987	.34 (1.11)	0	5	years out of last 5 has received afdc
mcaid5y	2982	.43 (1.27)	0	5	years out of last 5 has received medicaid
ssi5y	2990	.10 (.66)	0	5	years out of last 5 has received SSI
fdaid5y	2985	.48 (1.29)	0	5	years out of last 5 has received food stamps, WIC
assetbp	2700	.12	0	1	parents (or youth if independent) own business/partnership
assetop	2695	.13	0	1	own other property
assettu	2677	.10	0	1	own pre-paid tuition account

Variable Name	N	Mean (S.D.)	Min	Max	Description
assetpr	2665	.56	0	1	pension/retirement account
assetcs	2676	.68	0	1	checking/savings account
assetbb	2670	.18	0	1	bonds/bills/cds
assetsm	2677	.17	0	1	stocks/mutual funds
assetve	2702	.90	0	1	vehicles
assetot	2670	.13	0	1	other assets
assets	2690	2.95 (1.72)	0	9	how many of 9 above types of assets are owned?
massets	2990	.09	0	1	missing assets variable
<u>Health Indicators</u>					
ghealth1	2988	1.92 (.87)	1	5	describe general health, 1=excellent, 5=poor
weighti1	2990	.55	0	1	describes weight as ideal
weight1	2880	117.5 (28.4)	50	300	reported weight in pounds
height1	2890	62.91 (3.81)	48	75	reported height in inches
bmi1	2814	20.81 (4.16)	10	50	body mass index as calculated from weight & height
weightu1	2990	.17	0	1	describes weight as under, or very under
weighto1	2990	.28	0	1	describes weight as over, or very over
puber1	2953	.68	0	1	puberty started: "underway" for boys, had first period for girls
conds	2688	.44 (1.15)	0	11	scale of severity/variety of physical/emotional conditions
slscale	2688	.24 (.60)	0	6	scale of severity/variety of sensory limitations
ccscale	2688	.15 (.48)	0	4	scale of severity/variety of chronic conditions
<u>Delinquency</u>					
arr1	2989	.05 (.40)	0	9	number of previous arrests
parr1	2989	.03	0	1	ever arrested
propt0	2990	.06	0	1	property offenses prior to wave 1
theft0	2990	.01	0	1	thefts prior to wave 1
oprop0	2990	.02	0	1	other property crimes prior to wave 1
attak0	2990	.01	0	1	attacked others prior to wave 1
sell0	2990	.01	0	1	sold drugs prior to wave 1
c1freq1	2990	.70 (3.46)	0	99	# of property offenses wave 1
c2freq1	2990	.08	0	1	petty theft wave 1 (0 or 1)
c3freq1	2990	.20 (2.65)	0	99	theft frequency wave 1
c4freq1	2990	.23 (2.68)	0	99	other property crimes, wave 1
c5freq1	2990	.41 (2.33)	0	79	attacking others, wave 1
c6freq1	2990	.17 (2.60)	0	99	selling drugs, wave 1
c1prev1	2689	.16	0	1	prevalence of property offenses, wave 1
c2prev1	2601	.09	0	1	petty theft, wave 1
c3prev1	2942	.03	0	1	theft, wave 1
c4prev1	2909	.03	0	1	other property crimes, wave 1
c5prev1	2858	.10	0	1	attacking others, wave 1
c6prev1	2971	.02	0	1	selling drugs, wave 1
<u>Victimization</u>					

Variable Name	N	Mean (S.D.)	Min	Max	Description
burgl1	2961	.14	0	1	house burgled before age 12
sawshot1	2964	.08	0	1	saw someone shot before age 12
bullied1	2965	.19	0	1	was bullied before age 12
<u>Peer Indicators</u>					
					proportion of peers who go to church, 0: <10%, 1: ~25%, 2: ~50%, 3: ~75%, 4: >90%
prchrch	2919	1.96 (1.14)	0	4	>90%
prsmoke	2943	3.03 (1.14)	0	4	peers smoke: 0: >90%, 4: <10%
prdrunk	2918	3.57 (.86)	0	4	peers get drunk: 0: >90%, 4: <10%
prsport	2982	2.85 (.99)	0	4	peers play sports: 0 <10%, 4: >90%
prgangs	2939	3.64 (.77)	0	4	peers in gangs: 0: >90%, 4: <10%
prcollg	2944	2.77 (1.06)	0	4	peers college-bound: 0 <10%, 4: >90%
					peers do volunteer work: 0 <10%, 4: >90%
prvolun	2948	1.16 (1.09)	0	4	>90%
prdrugs	2904	3.37 (1.00)	0	4	peers do illegal drugs: 0: >90%, 4: <10%
prtruan	2971	3.21 (1.02)	0	4	peers cut classes: 0: >90%, 4: <10%
propeer	2782	25.57 (5.05)	1	36	prosocial teen peers scale count of number of antisocial items 50% or more of peers participate in
antipeer	2990	.90 (1.31)	0	5	or more of peers participate in
gangn1	2962	.38	0	1	gangs in neighborhood
gangp1	2976	.15	0	1	friends/siblings in gang
cgang1	2985	.03	0	1	youth ever in gang
<u>School Behavior and Attachment</u>					
sbnthret	2985	1.00 (4.66)	0	99	times threatened at school
sb_thret	2990	.23	0	1	ever threatened at school
sbnstole	2985	.43 (1.17)	0	20	times stolen from at school
sb_stole	2990	.23	0	1	ever stolen from at school
sbnfight	2988	.36 (1.28)	0	40	times in fight at school
sb_fight	2990	.18	0	1	ever in fight in school
sbnlate	2977	1.08 (4.41)	0	99	times unexcused tardy
sb_2late	2990	.17	0	1	2+ unexcused tardies
sbnabsent	2921	3.65 (4.69)	0	90	times unexcused absent
sb_2wkab	2990	.08	0	1	2 weeks or more unexcused absence
					school attachment, teachers are good,
satgood	2989	2.12 (.61)	0	3	0=strongly disagree, 3=strongly agree
satintr	2987	2.23 (.64)	0	3	teachers are interested in students student disruptions get in the way of calss, 0=strongly agree, 3=strongly disagree
sadisru	2989	1.20 (.83)	0	3	grades distribute fairly, 0=strongly dis, 3=strongly agree
safairg	2984	2.18 (.70)	0	3	3=strongly agree there's a lot of cheating, 0=strongly agree, 3=strongly disagree
sacheat	2969	1.52 (.89)	0	3	3=strongly disagree discipline is fair, 0=strongly dis, 3=strongly agree
safaird	2980	1.96 (.75)	0	3	3=strongly agree
sassafe	2989	2.24 (.72)	0	3	school is safe
schattach	2970	8.51 (1.96)	0	12	satgood+satintr+safairg+safaird, school attachment scale

Variable Name	N	Mean (S.D.)	Min	Max	Description
schenviro	2968	4.95 (1.65)	0	9	sadisru+sacheat+sassafe, school environment scale
sus1	2990	.18	0	1	ever suspended
cretain1	2990	.09	0	1	ever retained
cskip1	2625	.01	0	1	ever skipped a grade
t_col1	2990	.01	0	1	college track (high schoolers only)
t_voc1	2990	.00	0	1	vocational track
gifted	1927	.23	0	1	takes gifted classes (>9th grade at wave 1, or from transcript)
msgrd	2962	5.82 (1.74)	1	8	middle school grades, 1=mostly below D's 8=mostly A's
hga1	2990	7.14 (.80)	0	10	highest grade attended at wave 1
hgc1	2990	6.32 (.89)	0	9	highest grade completed
r_gs	2990	.04 (.86)	2.77	3.64	ASVAB age-standardized general science score ~N(0,1)
r_ar	2990	.05 (.90)	2.97	2.72	arithmetic reasoning
r_wk	2990	.04 (.91)	2.85	3.28	word knowledge
r_pc	2990	.05 (.89)	2.33	2.54	paragraph comprehension
r_no	2990	.04 (.86)	3.72	3.76	numerical operations
r_cs	2990	.04 (.83)	4.47	2.98	coding speed
r_ai	2990	.02 (.84)	2.44	4.98	auto information
r_si	2990	.01 (.84)	3.01	4.16	shop information
r_mk	2990	.05 (.84)	2.57	2.96	math knowledge
r_mc	2990	.04 (.90)	2.84	3.35	mechanical comprehension
r_ei	2990	.04 (.83)	2.52	4.47	electronics information
r_ao	2990	.04 (.88)	2.11	2.68	assembly objects
m_asvab	2990	.16	0	1	missing ASVAB assessment
<u>Adult behaviors: sex & substance use</u>					
ysex1	2990	.20 (.90)	0	10	years sexually active
alcho1	2984	.27	0	1	ever drank alcohol
smoke1	2982	.27	0	1	every smoked cigarettes
marij1	2984	.08	0	1	ever used marijuana
cocan2	2963	.04	0	1	ever used cocaine (as of wave 2)
ac30dy01	2989	.30 (1.82)	0	30	days of past 30 used alcohol
ac30rt01	2989	.29 (2.23)	0	80	drinks/day in past month
ac30sw01	2990	.07 (.84)	0	23	days drank before/after school/work last 30 days
ac30bn01	2990	.12 (1.19)	0	30	days drank 5+ drinks of past 30
cg30dy01	2989	.97 (4.45)	0	30	days smoked of past 30
cg30rt01	2990	.28 (1.54)	0	30	cigarettes/day in past 30 days
mj30dy01	2990	.17 (1.55)	0	30	days smoked marijuana of past 30
mj30sw01	2990	.06 (1.01)	0	30	days used marijuana before/after school/work last 30 days
ckyrti02	2966	3.71 (35.01)	0	500	times used cocaine between waves 1 & 2
ck30sw02	2968	.10 (1.36)	0	50	days used coke before/after school/work of last 30 days
<u>Religion</u>					
relcath	2990	.28	0	1	religion: catholic

Variable Name	N	Mean (S.D.)	Min	Max	Description
relbapt	2990	.19	0	1	religion: baptist
relprot	2990	.36	0	1	religion: other protestant
relothr	2990	.05	0	1	religion: other
relnone	2990	.13	0	1	religion: none/unknown
prelcath	2705	.28	0	1	parent's religion: catholic
prelbapt	2705	.19	0	1	parent's religion: baptist
prelprot	2705	.40	0	1	parent's religion: protestant
prelothr	2705	.04	0	1	parent's religion: other
prelnone	2705	.09	0	1	parent's religion: none/unknown
prupcath	2705	.34	0	1	parent's upbringing: catholic
prupbapt	2705	.21	0	1	parent's upbringing: baptist
prupprot	2705	.38	0	1	parent's upbringing: protestant
prupothr	2705	.04	0	1	parent's upbringing: other
prupnone	2705	.03	0	1	parent's upbringing: none/unknown
<u>Time Use</u>					
hw	2968	.91	0	1	spends time doing homework days during week does homework (out of 5)
hwfreq	2966	3.54 (1.56)	0	5	minutes during week does homework
hwmweek	2964	69.2 (75.1)	0	2400	minutes/weekend does homework
hwmwknd	2968	51.7 (143.9)	0	5400	minutes/weekend does homework
ls	2971	.30	0	1	spends time taking lessons outside school
lsfreq	2970	.86 (1.59)	0	5	days during week takes lessons
lsmweek	2942	23.5 (49.0)	0	600	minutes during week takes lessons
lsmwknd	2969	11.7 (57.9)	0	960	minutes/weekend takes lessons
tv	2971	.97	0	1	spends time watching tv
tvfreq	2967	4.34 (1.32)	0	5	days/week watches tv
tvmweek	2945	152.4 (134.9)	0	2160	minutes during week watches tv
tvmwknd	2966	365.7 (329.6)	0	2880	minutes/weekend watches tv
rd	2971	.65	0	1	reads for pleasures
rdfreq	2967	2.02 (1.93)	0	5	days/week reads for pleasure
rdmweek	2925	43.4 (111.8)	0	2700	minutes during week reads for pleasures
rdmwknd	2970	49.6 (107.5)	0	1800	minutes/weekend reads for pleasure
dinner1	2962	5.33 (2.11)	0	7	days/week eats dinner with family
housework1	2963	5.66 (1.79)	0	7	days/week housework "gets done"
funfam1	2962	2.75 (2.02)	0	7	days/week does fun things with family
relig1	2968	1.57 (1.98)	0	7	days/week does religious things with family
<u>Miscellaneous</u>					
shots1	2971	.41 (1.15)	0	7	days/week hears shots in neighborhood
hours96	2989	6.5 (64.9)	0	1880	hours worked in formal employment in 1996
hours97	2987	32.4 (166.5)	0	3305	hours worked in formal employment in 1997
urtlo1	2990	.13	0	1	unemployment rate in MSA (or rest of state if not in MSA): <3%
urtme1	2990	.58	0	1	unemployment rate in MSA: 3-5.9%
urthi1	2990	.21	0	1	unemployment rate in MSA: 6-8.9%
urtvh1	2990	.07	0	1	unemployment rate in MSA: >9%

Variable Name	N	Mean (S.D.)	Min	Max	Description
piithf96	2988	.02	0	1	got income from theft in 1996
piithf97	2966	.02	0	1	income from theft 1997
piiprp96	2990	.02	0	1	income from property crimes 1996
piiprp97	2968	.02	0	1	income from property crimes 1997
piidrg96	2989	.01	0	1	income from drugs 1996
piidrg97	2967	.03	0	1	income from drugs 1997
pii96	2990	.04	0	1	any illegal income 1996
pii97	2968	.06	0	1	any illegal income 1997

Appendix B. Unadjusted and adjusted standardized bias (variety score models)

Variable	Unadjusted Bias	classify-analyze	expected value	trajectory kernel	propensity score
dcpre4	36	17	19	5	18
dcpre3	38	5	9	-4	2
dcpre2	47	15	19	3	9
dcpre1	44	14	18	8	9
grp1prb*	-43	0	-6	0	-12
grp2prb*	0	-1	0	0	0
grp3prb*	11	-5	-4	-1	6
grp4prb	47	8	14	1	9
male*	9	3	4	2	2
hisp*	4	3	3	6	-10
white	-2	-2	-2	-12	12
black*	12	12	12	17	-22
other*	-10	-10	-10	-5	5
age1*	8	9	7	0	-5
ur_rur0*	5	7	7	1	9
ur_rur1*	4	7	7	1	11
ur_urb0*	-4	-7	-7	-1	-10
ur_urb1*	-7	-10	-9	-3	-11
ur_unk0	1	1	1	0	4
ur_unk1	6	7	7	4	1
msacty0*	11	10	10	16	-1
msacty1*	11	9	9	17	1
msancc0*	-10	-10	-10	-15	-1
msancc1*	-13	-13	-13	-20	-3
msanot0*	-1	-1	-1	1	-2
msanot1	4	6	5	6	3
msaunk0	2	2	3	1	4
dwellh*	-47	-45	-45	-39	-1
dwell*	24	21	22	25	-2
dwell	36	37	37	25	4
moves0	27	27	27	21	12
mobility0*	16	16	16	11	5
northel*	-8	-7	-7	-3	9
northc1*	-14	-12	-12	-12	-3
south1*	27	26	26	23	1
west1	-10	-11	-12	-12	-7
hhbb1	-56	-49	-50	-46	-18
hhbs1*	18	14	15	8	3
hhbn1*	39	35	36	34	14
hhoo1*	10	9	9	15	5
adopted*	-8	-10	-9	-11	-2
indep1*	-6	-4	-4	-2	-5
deadp*	1	-3	-2	-1	-6
fborn*	-4	-4	-4	0	-5
hhsz1*	6	9	9	16	-9
mom19*	40	39	39	38	0
headst*	28	28	28	29	-10

Variable	Unadjusted Bias	classify-analyze	expected value	trajectory kernel	propensity score
noins1*	30	30	30	24	-9
cfpar1	-5	5	4	6	1
cfrel1*	12	11	10	9	-2
cfrnd1*	-4	-12	-11	-15	-5
cfoth1*	-8	-10	-9	-3	8
cfnon1*	8	3	4	1	4
cfmis1*	1	3	3	6	2
momhigh*	-7	0	-1	4	7
momwant*	-15	-4	-6	-2	-7
momenjy*	6	14	14	14	6
mattach	-8	4	2	6	2
dadhigh*	-15	-7	-8	-4	-2
dadwant*	-23	-16	-17	-11	-6
dadenjy*	-10	-4	-5	-3	0
dattach	-19	-11	-12	-7	-3
msupprai*	-20	-11	-12	-5	5
msupcrit*	-10	-2	-3	-2	11
msuphelp*	-19	-13	-13	-4	10
msupblam*	-12	-4	-5	-1	0
msupplan*	-23	-16	-17	-17	7
msupscale	-26	-14	-16	-8	11
msup*	-19	-11	-12	-3	13
mstrict*	-8	-6	-7	-7	-3
dsupprai*	-22	-15	-16	-10	1
dsupcrit*	-13	-8	-8	-4	6
dsuphelp*	-16	-9	-10	-6	-1
dsupblam*	-15	-8	-9	-1	-5
dsupplan*	-34	-30	-30	-25	-9
dsupscale	-31	-21	-22	-14	-2
dsup*	-19	-12	-13	-2	6
dstrict*	13	12	12	8	-4
mknofr*	-12	-4	-5	2	-2
mknofp*	-17	-10	-11	-5	-1
mknoww*	-24	-10	-12	-1	10
mknosl	-23	-10	-12	-1	3
mknots*	-22	-14	-15	-8	11
dknofr*	-27	-17	-18	-11	0
dknofp*	-46	-37	-38	-28	0
dknoww*	-22	-10	-12	-3	11
dknosl	-38	-26	-27	-17	4
dknots*	-28	-20	-21	-17	7
auton*	1	-4	-3	-8	-5
blcurf*	20	13	14	15	-2
mblcurf	12	10	10	9	-3
blmovi*	23	14	15	7	2
mblmovi	22	14	15	8	2
blpeer*	-7	-14	-13	-17	-4
mblpeer	-8	-13	-12	-15	-2

Variable	Unadjusted Bias	classify-analyze	expected value	trajectory kernel	propensity score
blimits	14	3	4	-2	-1
perm	9	3	3	3	-22
induct	-7	-7	-7	-6	2
mfrcomp*	-20	-16	-16	-18	-5
mfrscrm*	-7	0	-1	-3	-15
mfrlove*	0	5	4	3	-11
mfrcrit*	-3	4	3	6	-6
mfrenco*	-18	-13	-14	-10	-9
mfrblme*	-2	5	4	8	-17
mfrel	-13	-5	-6	-5	-16
fmrcomp*	-22	-17	-18	-18	-8
fmrscrm*	-9	-4	-5	-1	-13
fmrlove*	-4	0	-1	-1	-12
fmrcri*	-8	-3	-4	3	-6
fmrenco*	-16	-13	-13	-10	-10
fmrblme*	-7	1	0	9	5
fmrel	-16	-9	-10	-5	-11
allwnce96*	0	-3	-3	1	5
allwncei96*	9	7	7	4	3
heat1*	-4	-5	-4	-7	-5
computer1*	-61	-59	-60	-56	1
study1*	-18	-13	-14	-13	-1
webster1*	-16	-15	-15	-12	7
innice1	-51	-48	-48	-44	-7
infair1*	31	28	28	28	12
inpoor1*	33	31	32	25	0
inunk1*	5	8	8	7	-11
outnice1	-52	-51	-51	-48	-3
outfair1*	28	26	26	28	3
outpoor1*	43	43	43	34	1
outunk1*	-10	-9	-10	-3	-2
med_do*	49	47	47	48	-14
med_hs*	-2	1	0	-3	10
med_co	-48	-48	-48	-43	3
med_un*	13	13	13	8	-1
fed_do*	36	35	35	32	-3
fed_hs	-4	-4	-4	-4	7
fed_co*	-43	-42	-42	-38	-1
fed_un*	20	19	19	16	-5
pov000*	33	34	33	39	-3
pov100*	23	22	23	21	7
pov200*	-5	-4	-5	-9	-1
pov300*	-23	-24	-24	-22	1
pov400*	-29	-30	-30	-28	0
povunk	-5	-3	-3	-5	-5
income1	-55	-54	-54	-50	-6
govaid*	58	54	54	51	1
afdc*	42	39	39	36	0

Variable	Unadjusted Bias	classify-analyze	expected value	trajectory kernel	propensity score
mcaid*	47	44	44	43	4
ssi*	15	14	14	24	8
fdaid*	47	44	44	41	3
afdc5y*	24	23	23	30	0
mcaid5y*	33	32	32	36	-2
ssi5y*	17	16	16	28	8
fdaid5y*	36	34	35	41	-6
assetbp*	-17	-16	-16	-15	1
assetop*	-27	-27	-26	-21	4
assettu*	-18	-17	-17	-18	-2
assetpr*	-50	-50	-50	-47	2
assetcs*	-31	-31	-31	-36	1
assetbb*	-34	-37	-37	-32	3
assetsm*	-33	-32	-32	-29	-5
assetve*	-25	-24	-24	-32	3
assetot*	-7	-9	-9	-8	3
assets	-55	-55	-55	-55	2
massets*	-3	-1	-1	-6	-6
ghealth1*	24	22	22	22	10
weight1*	11	13	13	16	8
weight1*	10	7	7	0	-1
height1*	-4	-7	-7	-9	0
bmi1	17	15	15	9	1
conds*	13	9	10	8	5
slscale*	5	6	6	4	4
ccscale*	11	12	12	9	-5
arr1*	24	20	20	15	2
parr1	28	21	22	17	8
propt0*	12	-3	-2	-7	-6
theft0*	13	6	7	10	-7
oprop0*	11	0	1	-6	4
attak0*	5	-3	-3	-9	-21
sell0*	20	13	14	13	-10
c1freq1*	22	8	10	1	10
c2freq1*	29	9	11	0	8
c3freq1*	11	5	6	-3	-12
c4freq1*	1	-13	-12	-19	3
c5freq1*	19	6	7	-3	-10
c6freq1*	15	7	8	4	3
c1prev1	40	14	17	1	21
c2prev1	34	12	15	-1	10
c3prev1	16	4	6	-1	-3
c4prev1	25	10	12	-3	2
c5prev1	45	28	30	13	8
c6prev1	28	22	23	20	7
burg1*	13	8	8	5	-1
sawshot1*	32	25	25	21	0
bullied1*	20	10	11	1	1

Variable	Unadjusted Bias	classify-analyze	expected value	trajectory kernel	propensity score
prchrch*	-40	-37	-38	-33	-5
prsmoke*	-42	-35	-35	-20	-3
prdrunk*	-30	-25	-25	-22	-7
prsport*	-19	-19	-19	-11	5
prgangs*	-35	-31	-32	-24	9
prcollg*	-29	-27	-27	-21	-6
prvolun*	12	16	15	20	4
prdrugs*	-47	-38	-38	-27	-11
prtruan*	-44	-38	-38	-32	2
propeer	-53	-44	-45	-31	-1
antipeer	49	43	43	32	3
gangn1*	23	14	15	14	-4
gangp1*	48	39	40	31	-6
cgang1*	26	18	19	11	-3
sbnthret*	18	14	14	8	3
sb_thret	35	25	27	21	8
sbnstole*	7	3	3	13	15
sb_stole	6	0	1	3	-1
sbnfight*	28	21	22	16	2
sb_fight	39	28	30	24	0
sbnlate*	20	15	15	11	0
sb_2late	31	25	26	26	1
sbnabsent*	39	34	35	38	-2
sb_2wkab	23	20	20	20	-16
satgood*	-33	-25	-26	-23	-4
satintr*	-26	-20	-21	-13	0
sadisru*	4	4	4	1	-9
safairg*	-29	-22	-23	-18	-2
sacheat*	-19	-14	-14	-6	4
safaird*	-17	-8	-10	-7	-2
sassafe*	-33	-27	-28	-22	-4
schattach	-34	-24	-26	-19	-2
schenviro	-22	-17	-17	-12	-4
sus1*	57	48	49	50	9
cretain1*	38	39	39	42	11
cskip1*	2	2	2	0	0
t_coll1*	-3	-3	-3	1	6
t_voc1*	12	13	13	10	5
gifted*	-56	-50	-51	-50	-7
msgrd*	-87	-77	-78	-70	-14
hga1*	-15	-15	-16	-21	-2
hgc1	-13	-12	-13	-17	-6
r_gs*	-42	-41	-41	-47	-1
r_ar*	-46	-44	-44	-49	-4
r_wk*	-42	-40	-40	-45	0
r_pc*	-53	-49	-49	-53	-6
r_no*	-40	-37	-38	-39	-9
r_cs*	-44	-41	-41	-39	-2

Variable	Unadjusted Bias	classify-analyze	expected value	trajectory kernel	propensity score
r_ai*	-19	-19	-18	-24	-3
r_si*	-19	-19	-19	-26	1
r_mk*	-64	-60	-60	-61	-4
r_mc*	-43	-40	-40	-43	0
r_ei*	-37	-34	-34	-39	-6
r_ao*	-45	-41	-42	-41	0
m_asvab*	25	26	26	21	2
ysex1*	32	27	28	27	-5
alcho1*	31	12	14	6	9
smoke1*	49	30	32	17	8
marij1*	45	33	35	22	5
cocan2*	30	20	21	15	9
ac30dy01*	8	-3	-1	-3	3
ac30rt01*	10	0	1	1	5
ac30sw01*	4	-3	-2	-3	-13
ac30bn01*	5	-6	-5	-7	-9
cg30dy01*	31	19	21	11	1
cg30rt01*	29	22	23	13	-1
mj30dy01*	21	17	17	11	-4
mj30sw01*	11	8	8	5	-1
ckyr02*	19	13	14	8	8
ck30sw02*	11	7	8	7	1
relcath*	-11	-10	-11	-5	4
relbapt*	13	12	12	14	-4
relprot	-7	-5	-5	-8	-1
relothr*	-19	-17	-18	-15	-2
relnone*	17	14	15	8	2
prelath*	-13	-14	-14	-7	5
prelbapt*	3	1	1	4	-13
prelprot	14	16	16	9	2
prelothr*	-18	-15	-15	-13	-2
prelnone*	2	1	1	0	8
prupath*	-15	-16	-16	-11	1
prupbapt*	17	16	16	18	-7
pruprot	0	2	2	-5	1
prupothr*	-10	-9	-9	-8	0
prupnone*	6	6	6	4	10
hw	-30	-26	-27	-21	-4
hwfreq*	-43	-37	-38	-28	2
hwmweek*	-21	-16	-17	-11	4
hwmwknd*	-13	-10	-10	-6	0
ls	-19	-16	-17	-14	6
lsfreq*	-15	-13	-13	-12	-2
lsmweek*	-14	-11	-12	-6	8
lsmwknd*	-15	-13	-13	-10	1
tv	-7	-7	-7	-13	-3
tvfreq*	1	-2	-1	-11	-3
tvmweek*	19	15	15	13	-2

Variable	Unadjusted Bias	classify-analyze	expected value	trajectory kernel	propensity score
tvnwknnd*	20	17	17	9	3
rd	0	3	3	3	5
rdfreq*	-2	1	1	0	1
rdmweek*	6	7	7	9	4
rdmwknnd*	-5	-2	-2	-3	1
dinner1*	-9	-3	-4	-4	-5
housework1*	-14	-10	-10	-8	-6
funfam1*	4	10	10	12	3
relig1*	-18	-13	-14	-1	-1
shots1*	20	15	16	21	3
hours96*	-5	-8	-8	-10	0
hours97*	-12	-14	-14	-16	0
urtlo1*	14	15	15	8	8
urtme1	-13	-13	-13	-9	1
urthi1*	8	8	8	10	1
urtvh1*	-8	-9	-9	-11	-14
weightu1	-13	-12	-12	-11	-12
weighto1	-2	-5	-5	-10	-1
puber1*	6	1	1	-5	3
piithf96*	12	2	3	-4	-4
piithf97*	19	8	10	5	-1
pii prp96*	25	14	15	4	2
pii prp97*	8	-6	-5	-5	8
piidrg96*	24	20	21	18	2
piidrg97*	28	16	18	10	-6
pii96	28	13	15	5	3
pii97	26	7	9	3	-7
balanced/314	183	222	218	228	310
% balanced	58%	71%	69%	73%	99%

*included in propensity score model

Appendix C. Unadjusted and adjusted standardized bias (IRT models)

Variable	Unadjusted Bias	Classify-Analyze	Expected Value	Trajectory Kernel	Propensity Score
dpre3 (IRT)	34	6	10	-1	0
dpre2 (IRT)	37	12	13	10	13
dpre1 (IRT)	33	7	10	0	15
grp1prb*	-36	0	-4	0	-10
grp2prb*	6	-1	0	0	3
grp3prb*	5	-4	-3	0	3
grp4prb	38	4	8	0	8
male*	19	13	14	6	-3
hisp*	10	10	10	9	-26
white	-3	-2	-2	-16	14
black*	12	9	10	18	-20
other*	-7	-6	-6	1	2
age1*	3	3	2	1	-22
ur_rur0*	-1	3	2	1	12
ur_rur1*	-8	-6	-6	-7	7
ur_urb0*	-2	-4	-4	2	8
ur_urb1*	3	1	1	4	-11
ur_unk0	3	2	2	-3	-22
ur_unk1	10	10	10	5	8
msacty0*	12	10	10	14	11
msacty1*	12	10	10	18	13
msancc0*	-7	-5	-6	-7	10
msancc1*	-1	0	0	-10	-13
msanot0*	-10	-8	-9	-5	-2
msanot1	-13	-12	-12	-8	2
msaunk0	4	3	3	-2	-22
dwellh*	-46	-45	-45	-39	-17
dwella*	29	27	27	27	6
dwello	30	30	30	22	17
moves0	14	14	14	9	1
mobility0*	8	8	8	2	0
northel1*	-18	-18	-18	-7	6
northcl1*	-7	-4	-4	-13	0
southl1*	35	34	34	32	7
westl	-19	-21	-21	-19	-15
hhbb1	-53	-48	-48	-44	-19
hhbs1*	25	24	24	15	4
hhbn1*	31	27	27	30	24
hhoo1*	9	8	8	9	-19
adopted	-15	-16	-16	-16	-4
indep1*	-6	-4	-4	-3	-6
deadp*	-3	-5	-5	-7	-2
fborn*	-9	-9	-9	-6	-4
hhszsl1*	9	11	11	23	-27
mom19*	42	41	41	39	1
headst*	36	35	35	37	-18
noinsl1*	29	30	30	23	-3

Variable	Unadjusted Bias	Classify-Analyze	Expected Value	Trajectory Kernel	Propensity Score
cfpar1	-11	-3	-4	-1	-18
cfrel1*	13	11	11	9	9
cfrnd1*	0	-6	-5	-9	12
cfoth1*	-3	-6	-5	-1	10
cfnon1*	8	6	6	6	-1
cfmis1	-8	-6	-6	-4	-4
momhigh*	-5	0	0	2	19
momwant*	-13	-4	-5	-12	16
momenjy*	8	16	15	12	18
mattach	-5	4	3	1	22
dadhigh*	-16	-8	-10	-9	14
dadwant*	-16	-9	-10	-13	4
dadenjy*	-18	-13	-13	-16	-9
dattach	-19	-12	-13	-15	4
msupprai*	-12	-4	-5	0	14
msupcrit*	-8	-1	-2	-6	10
msuphelp*	-10	-4	-4	-5	-13
msupblam*	-1	6	5	0	-1
msupplan*	-19	-12	-13	-11	1
msupscale	-15	-4	-6	-7	4
msup*	-9	-1	-2	1	-9
mstrict*	-6	-4	-4	-3	-16
dsupprai*	-23	-15	-16	-17	10
dsupcrit*	-16	-13	-13	-13	17
dsuphelp*	-12	-6	-7	-14	-13
dsupblam*	-11	-4	-5	-3	18
dsupplan*	-37	-29	-30	-32	3
dsupscale	-29	-19	-21	-23	11
dsup*	-15	-8	-9	-5	0
dstrict*	12	12	12	5	-12
mknofr*	-5	2	1	-1	16
mknofp*	-9	-5	-5	-3	11
mknoww*	-6	6	5	7	10
mknosl	-9	1	0	2	16
mknots*	-3	2	2	-1	2
dknofr*	-24	-15	-17	-17	19
dknofp*	-51	-43	-44	-40	-3
dknoww*	-12	0	-2	-6	10
dknosl	-35	-24	-26	-25	10
dknots*	-16	-9	-10	-12	-7
auton*	-12	-15	-15	-14	-1
blcurf*	15	9	10	14	-1
mblcurf	12	8	9	9	4
blmovi*	16	9	9	7	-6
mblmovi	15	9	10	8	2
blpeer*	-23	-28	-28	-26	-2
mblpeer	-24	-28	-28	-27	7
blimits	-1	-10	-9	-8	-5

Variable	Unadjusted Bias	Classify-Analyze	Expected Value	Trajectory Kernel	Propensity Score
perm	13	9	9	-2	0
induct	-15	-10	-11	-9	32
mfrcomp*	-19	-16	-15	-16	15
mfrscrm*	-8	2	0	-10	12
mfrlove*	4	8	8	6	4
mfrcrit*	3	9	8	6	35
mfrenco*	-8	-3	-4	-3	-31
mfrblme*	6	15	13	6	-4
mfrel	-6	3	1	-4	7
fmrcomp*	-24	-20	-21	-21	21
fmrscrm*	-11	-3	-5	-9	7
fmrlove*	3	7	6	6	-8
fmrcri*	0	7	6	5	15
fmrenco*	-21	-18	-18	-15	-26
fmrblme*	-20	-12	-14	-6	-6
fmrel	-16	-9	-10	-9	0
allwnce96*	4	2	3	6	11
allwncei96*	3	2	2	5	10
heat1*	1	0	0	-12	-14
computer1*	-52	-50	-51	-46	3
study1*	-18	-16	-17	-14	36
webster1*	-8	-8	-8	-6	20
innice1	-51	-48	-48	-44	-9
infair1*	27	25	25	25	14
inpoor1*	28	27	27	23	11
inunk1*	20	21	21	17	-19
outnice1	-50	-48	-48	-46	-13
outfair1*	27	26	26	26	9
outpoor1*	41	40	40	35	6
outunk1*	-6	-7	-7	0	2
med_do*	41	39	40	46	14
med_hs*	8	9	9	2	9
med_co	-50	-49	-49	-44	-9
med_un*	12	12	12	5	-21
fed_do*	32	31	31	31	-17
fed_hs	-2	-3	-3	-1	13
fed_co*	-43	-41	-42	-37	0
fed_un*	22	22	22	13	1
pov000*	35	35	35	44	17
pov100*	16	17	17	17	15
pov200*	-13	-12	-13	-17	2
pov300*	-17	-16	-16	-17	0
pov400*	-31	-33	-33	-32	3
povunk	3	3	3	-1	-34
income1	-58	-57	-57	-56	-8
govaid*	47	45	45	42	-3
afdc*	29	26	27	27	11
mcaid*	36	36	36	35	17

Variable	Unadjusted Bias	Classify-Analyze	Expected Value	Trajectory Kernel	Propensity Score
ssi*	17	16	16	26	5
fdaid*	36	35	35	34	14
afdc5y*	22	21	21	27	20
mcaid5y*	26	26	26	31	24
ssi5y*	21	21	21	29	4
fdaid5y*	29	28	28	37	15
assetbp*	-14	-13	-13	-11	3
assetop*	-38	-38	-38	-31	0
assettu*	-9	-9	-9	-14	-3
assetpr*	-44	-42	-42	-42	0
assetcs*	-33	-31	-31	-42	-3
assetbb*	-27	-30	-30	-26	3
assetsm*	-24	-22	-22	-20	5
assetve*	-19	-18	-18	-31	0
assetot*	-13	-14	-14	-15	5
assets	-50	-49	-49	-52	2
massets*	9	9	9	1	-38
ghealth1*	26	25	25	26	3
weighti1*	9	12	12	16	7
weight1*	10	7	8	5	5
height1*	-3	-5	-5	-5	9
bmi1	17	15	15	12	5
conds*	18	15	16	10	12
slscale*	2	3	3	-1	8
ccscale*	15	16	16	6	-3
arr1*	26	23	23	17	6
parr1	28	24	24	18	13
propt0*	16	7	8	-2	-19
theft0*	21	17	18	22	17
oprop0*	19	14	14	6	13
attak0*	6	-2	-1	-2	2
sell0*	8	3	4	4	4
c1freq1*	19	9	10	1	11
c2freq1*	21	6	8	-1	-17
c3freq1*	7	2	2	-3	-94
c4freq1*	-6	-15	-14	-18	1
c5freq1*	17	8	9	7	-1
c6freq1*	0	-4	-3	-5	3
c1prev1	28	7	9	-5	9
c2prev1	24	8	10	0	-13
c3prev1	14	6	7	5	-14
c4prev1	6	-4	-3	-11	7
c5prev1	36	23	25	14	-12
c6prev1	13	8	9	12	16
burgl1*	20	16	16	10	-13
sawshot1*	29	22	23	26	7
bullied1*	27	20	21	8	3
prchrch*	-25	-21	-21	-23	1

Variable	Unadjusted Bias	Classify-Analyze	Expected Value	Trajectory Kernel	Propensity Score
prsmoke*	-26	-18	-19	-7	4
prdrunk*	-4	2	2	-6	2
prsport*	-6	-4	-5	-2	-8
prgangs*	-35	-28	-29	-21	-1
prcollg*	-18	-14	-15	-11	-6
prvolun*	21	26	25	26	0
prdrugs*	-20	-11	-12	-14	1
prtruan*	-26	-20	-21	-21	-4
propeer	-28	-18	-19	-14	0
antipeer	23	16	17	13	0
gangn1*	29	22	23	23	7
gangp1*	41	34	35	34	-13
cgang1*	18	11	11	10	-7
sbnthret*	14	11	11	5	3
sb_thret	32	24	25	21	12
sbnstole*	15	11	11	20	16
sb_stole	15	10	11	7	-16
sbnfight*	14	5	6	24	0
sb_fight	24	15	16	27	6
sbnlate*	20	17	18	18	7
sb_2late	22	18	18	17	-16
sbnabsent*	24	18	18	27	8
sb_2wkab	6	3	3	9	-11
satgood*	-30	-23	-23	-25	-1
satintr*	-15	-8	-9	-5	15
sadisru*	-1	-1	-1	-2	-7
safairg*	-20	-15	-16	-17	8
sacheat*	-12	-9	-9	-6	6
safaird*	-3	3	2	2	-3
sassafe*	-37	-31	-31	-33	1
schattach	-22	-13	-14	-13	8
schenviro	-22	-18	-18	-18	1
sus1*	45	35	36	42	-11
cretain1*	39	39	39	40	23
cskip1*	-6	-5	-5	-6	-1
gifted*	-61	-54	-54	-57	-25
msgrd*	-87	-77	-78	-71	-12
hga1*	-22	-23	-23	-25	-5
hgc1	-18	-17	-18	-17	-13
r_gs*	-47	-45	-45	-47	7
r_ar*	-38	-35	-36	-40	1
r_wk*	-47	-45	-46	-48	2
r_pc*	-52	-48	-49	-52	-10
r_no*	-38	-36	-36	-37	-13
r_cs*	-42	-38	-38	-36	-10
r_ai*	-28	-25	-26	-33	-13
r_si*	-21	-21	-21	-24	8
r_mk*	-67	-63	-63	-63	8

Variable	Unadjusted Bias	Classify-Analyze	Expected Value	Trajectory Kernel	Propensity Score
r_mc*	-40	-38	-38	-40	5
r_ei*	-29	-26	-26	-31	-4
r_ao*	-41	-35	-36	-33	6
m_asvab*	27	27	27	23	-8
ysex1*	21	18	18	30	5
alcho1*	15	2	4	3	3
smoke1*	41	27	29	12	0
marij1*	24	16	17	10	-6
cocan2*	14	6	7	4	10
ac30dy01*	-8	-15	-14	-12	-2
ac30rt01*	-8	-13	-12	-12	-2
ac30sw01*	-4	-7	-6	-1	3
ac30bn01*	-10	-16	-15	-12	0
cg30dy01*	11	3	3	-3	-6
cg30rt01*	4	-4	-3	2	6
mj30dy01*	3	-1	-1	-6	-8
mj30sw01*	2	0	1	-2	1
ckyr02*	11	6	7	3	7
ck30sw02*	7	3	4	3	3
relcath*	-1	-1	-1	4	14
relbapt*	6	3	3	7	10
relprot	-8	-5	-6	-13	-3
relotthr*	-13	-12	-12	-9	-10
relnone*	11	11	11	10	-21
prelath*	0	-1	-1	2	16
prelbapt*	7	5	5	2	-12
prelprot	2	4	4	0	-12
prelothr*	-17	-16	-16	-12	-1
prelnone*	-3	-3	-3	1	12
prupath*	-9	-12	-12	-9	0
prupbapt*	16	13	14	16	0
prupprot	-8	-4	-4	-8	-6
prupotthr*	-5	-3	-3	-3	5
prupnone*	10	10	10	9	11
hw	-17	-15	-15	-16	-3
hwfreq*	-29	-25	-26	-17	0
hwmweek*	-4	0	0	-1	5
hwmwknd*	-5	-3	-3	-2	-10
ls	-29	-27	-27	-19	1
lsfreq*	-18	-17	-18	-13	-3
lsmweek*	-14	-12	-13	-4	8
lsmwknd*	-6	-5	-5	-1	2
tv	-5	-7	-6	-13	-18
tvfreq*	3	2	2	-12	-17
tvmweek*	21	17	17	14	-2
tvmwknd*	35	32	33	23	12
rd	-6	-2	-3	3	17
rdfreq*	-5	-1	-2	5	19

Variable	Unadjusted Bias	Classify-Analyze	Expected Value	Trajectory Kernel	Propensity Score
rdmweek*	-9	-7	-7	-4	7
rdmwknd*	-14	-11	-12	-11	4
dinner1*	-7	-3	-4	-2	0
housework1*	-5	0	-1	7	4
funfam1*	8	12	12	12	7
relig1*	-21	-19	-19	-9	-4
shots1*	17	11	12	18	8
hours96*	9	10	10	6	8
hours97*	-13	-12	-12	-12	1
urtlo1*	17	15	15	10	10
urtme1	-7	-5	-6	-9	14
urthi1*	-6	-7	-7	0	1
urtvh1*	0	1	1	4	-41
weightu1	-6	-7	-6	-6	-9
weighto1	-6	-9	-9	-14	0
puber1*	3	-1	0	1	18
piithf96*	1	-8	-7	-8	-1
piithf97*	24	19	19	11	8
piiprp96*	9	1	2	-4	10
piiprp97*	14	5	6	-2	-2
piidrg96*	8	4	5	5	9
piidrg97*	20	13	14	12	-1
pii96	13	2	4	-1	7
pii97	25	14	15	7	-5
balanced/311	198	230	226	229	288
% balanced	64%	74%	73%	74%	93%

REFERENCES

- Abbott, Andrew. 1997. On the concept of turning point. *Comparative Social Research* 16:85-105.
- Alexander, Karl L., Doris R. Entwisle, and Carrie S. Horsey. 1997. "From First Grade Forward: Early Foundations of High School Dropout." *Sociology of Education* 70:87-107.
- Alexander, Karl L., Doris R. Entwisle and Nader S. Kabbani. 2001. The dropout process in life course perspective: Early risk factors at home and school. *Teachers College Record* 5:760-822.
- Apel, Robert, Robert Brame, Shawn D. Bushway, Amelia Haviland, Daniel S. Nagin, and Raymond Paternoster. February 2006. "Unpacking the relationship between adolescent employment and antisocial behavior: A matched samples comparison." Manuscript.
- Bachman, Jerald G., and Patrick O'Malley. 1978. *Youth in Transition Volume VI: Adolescence to Adulthood: Change and Stability in the Live of Young Men*. Ann Arbor, MI: University of Michigan Press.
- Barton, Paul E. 2005. *One-Third of a Nation: Rising Dropout Rates and Declining Opportunities*. Princeton, NJ: Educational Testing Service.
- Battin-Pearson, Sara, Michael D. Newcomb, Robert D. Abbot, Karl G. Hill, Richard F. Catalano, and David J. Hawkins. "Predictors of early high school dropout: A test of five theories." *Journal of Educational Psychology* 92:568-582.
- Bernburg, Jön Gunnar and Marvin D. Krohn. 2003. Labeling, life chances, and adult crime: The direct and indirect effects of official intervention in adolescence on

- crime in early adulthood. *Criminology* 41:1287-1318.
- Blumstein, Alfred, Jacqueline Cohen, Jeffrey Roth, and Christy Visher (Eds.) 1996. *Criminal Careers and "Career Criminals."* Volume 1. Washington, D.C.: National Academy Press.
- Brame, Robert, Shawn Bushway and Raymond Paternoster. 1999. Can panel designs tell us anything about continuity and change in offending over time. *Criminology* 37:588-641.
- Cairns, Robert B., Beverley D. Cairns, and Holly J. Neckerman. 1989. "Early school dropout: Configurations and determinants." *Child Development* 60:1437-1452.
- Cloward, Richard A., and Lloyd E. Ohlin. 1960. *Delinquency and Opportunity: A Theory of Delinquent Gangs*. New York: Free Press.
- Cohen, Albert K. 1950. *Delinquent Boys*. New York: Free Press.
- Crum, Rosa M., Kathleen K. Bucholz, John E. Helzer, and James C. Anthony. 1992. "The Risk of Alcohol Abuse and Dependence in Adulthood: The Association with Educational Level." *American Journal of Epidemiology* 135:989-999.
- Crum, Rosa, Margaret E. Ensminger, Marguerite J. Ro and Joan McCord. 1998. "The Association of Educational Attainment and School Dropout with Risk of Alcoholism: A Twenty-Five-Year Prospective Study of Inner-City Children." *Journal of Studies on Alcohol* 59:318-326.
- Dannefer, Dale. 1984. Adult development and social theory: A paradigmatic reappraisal. *American Sociological Review* 49:100-116.
- De Li, S. 1999. Legal sanctions and youths' status achievement: A longitudinal study. *Justice Quarterly* 16:377-401.

- Dehejia, Rajeev and Sadek Wahba. 1999. "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs." *Journal of the American Statistical Association* 94:1053-1062
- , 2002. "Propensity score matching methods for nonexperimental causal studies." *The Review of Economics and Statistics* 84:151-161.
- Donovan, John E., and Richard Jessor. 1985. "Structure of Problem Behavior in Adolescence and Young Adulthood." *Journal of Consulting and Clinical Psychology* 53:890-904.
- Donovan, John E., Richard Jessor, and Frances M. Costa. 1988. "Syndrome of Problem Behavior in Adolescence and Young Adulthood: A Replication" *Journal of Consulting and Clinical Psychology* 56:762-765.
- Drapela, Laurie A. 2005. "Does dropping out of high school cause deviant behavior? An analysis of the National Education Longitudinal Survey." *Deviant Behavior* 26:47-62.
- Eckstein, Zvi and Kenneth I. Wolpin. 1999. "Why youths drop out of high school: The impact of preferences, opportunities, and abilities." *Econometrica* 67:1295-1339.
- Elder, Glen H. Jr. 1985. Perspectives on the life course. In Elder, Glen H. Jr. (ed.) *Life Course Dynamics*. Ithaca: Cornell University Press.
- , 1994. Time, human agency, and social change: Perspectives on the life course. *Social Psychology Quarterly* 57:4-15.
- , 1998. The life course as developmental theory. *Child Development* 69:1-12.
- Elliott, Delbert S. 1966. "Delinquency, school attendance, and dropout." *Social Problems* 13: 307-314.

- Elliott, Delbert S., David Huizinga, and Suzanne S. Ageton. 1985. *Explaining Delinquency and Drug Use*. Beverly Hills, CA: Sage Publications.
- Elliott, Delbert S., and Harwin L. Voss. 1974. *Delinquency and Dropout*. Lexington, MA: Lexington Books.
- Ensminger, Margaret and Anita Slusarick. 1992. "Paths to high school graduation or dropout: A longitudinal study of a first grade cohort." *Sociology of Education* 65:95-113.
- Farrington, David P., Bernard Gallagher, Lynda Morley, Raymond J. St. Ledger, and Donald J. West. 1986. "Unemployment, school leaving, and crime." *British Journal of Criminology* 26:335-356.
- Farrington, David P., Lloyd Ohlin, and James Q. Wilson. 1986. *Understanding and Controlling Crime*. New York: Springer-Verlag.
- Finn, Jeremy D. 1989. Withdrawing from school. *Review of Educational Research* 59:117-142.
- Giordano, Peggy C., Stephen A. Cernkovich and Donna D. Holland. 2003. "Changes in friendship relations over the life course: Implications for desistance from crime." *Criminology* 41:293-327.
- Glueck, Sheldon and Eleanor Glueck. 1950. *Unraveling Juvenile Delinquency*. New York: Commonwealth Fund.
- Gottfredson, Michael R. and Travis Hirsch. 1990. *A General Theory of Crime*. Stanford: Stanford University Press.
- Greene, Jay P. 2002. *High School Graduation Rates in the United States*. New York: The Manhattan Institute for Policy Research.

- Guttman, Louis L. 1950. "The bases for scalogram analysis," pp. 60-90 in Samuel A. Stouffer, Louis Guttman, Edward A. Suchman, Paul F. Lazarsfeld, Shirley A. Starr and John Clausen (eds.). *Measurement and Prediction: Studies in Social Psychology in World War II*. Princeton, NJ: Princeton University Press.
- Hareven, Tamara K. and Kanji Masaoka. 1988. "Turning points and transitions: Perceptions of the life course." *Journal of Family History* 13:271-289.
- Haviland, Amelia M. and Daniel S. Nagin. 2005. "Causal inferences with group based trajectory models." *Psychometrika* 70(3):1-22.
- Haviland, Amelia, Daniel S. Nagin, and Paul R. Rosenbaum. 2005. Manuscript.
- Heckman, James J. and V. Joseph Hotz. 1989. Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the American Statistical Association* 84:862-874.
- Hindelang, Michael J., Travis Hirschi, and Joseph G. Weis. 1981. *Measuring Delinquency*. Beverly Hills: Sage Publications.
- Hoffmann, John P. and Felicia Gray Cerbone. 1999. "Stressful life events and delinquency escalation in early adolescence." *Criminology* 37:343-373.
- Horney, Julie D., Wayne Osgood and Ineke Haen Marshall. 1995. Criminal careers in the short-term: Intra-individual variability in crime and its relationship to local life circumstances. *American Sociological Review* 60:655-673.
- Janosz, Michel, and Marc Le Blanc. 2000. "Predicting different types of school dropouts: A typological approach." *Journal of Educational Psychology* 92:171-190.
- Jarjoura, G. Roger. 1993. Dropping out of school enhance delinquent involvement?

- Results from a large-scale national probability sample. *Criminology* 31:149-172.
- Jarjoura, G. Roger. 1996. The conditional effect of social class on the dropout-delinquency relationship. *Journal of Research in Crime and Delinquency* 33:232-255.
- Jessor, Richard, and Shirley L. Jessor. 1977. *Problem Behavior and Psychosocial Development*. New York: Academic Press.
- Kandel, Denise B., Ronald C. Kessler, and Rebecca Z. Margulies. 1978. "Antecedents of Adolescent Initiation into Stages of Drug Use: A Developmental Analysis." *Journal of Youth and Adolescence* 7:13-40.
- Katz, Jack. 1988. *Seductions of Crime*. New York: Basic Books.
- Kaufman, Philip, Martha Naomi Alt, and Christopher D. Chapman. 2004. *Dropout Rates in the United States: 2001* (NCES 2005-046). U.S. Department of Education. National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Krohn, Marvin D., Terence P. Thornberry, Lori Collins-Hall, and Alan J. Lizotte. 1995. School dropout, delinquent behavior, and drug use: An examination of the causes and consequences of dropping out of school. In Kaplan, Howard B. (ed.) *Drugs, Crime, and Other Deviant Adaptations: Longitudinal Studies*. New York: Plenum Press.
- Laub, John H. and Robert J. Sampson. 1993. Turning points in the life course: Why change matters to the study of crime. *Criminology* 31:301-325.
- , 2003. *Shared Beginnings, Divergent Lives: Delinquent boys to Age 70*. Cambridge, Mass: Harvard University Press.

- Laub, John H., Daniel S. Nagin, and Robert J. Sampson. 1998. "Trajectories of change in criminal offending: Good marriages and the desistance process." *American Sociological Review* 63:225-238.
- LeBlanc, Marc, and Marcel Frechette. 1989. *Male Criminal Activity From Childhood Through Youth*. New York: Springer-Verlag.
- LeBlanc, Marc, Évelne Vallières, and Pierre McDuff. 1993. "The prediction of males' adolescent and adult offending from school experience." *Canadian Journal of Criminology* 35:459-478.
- Mandelbaum, David G. 1973. The study of life history: Gandhi. *Current Anthropology* 14:177-206.
- Mendell, David. April 17, 2006. "Gates Foundation Gives City Schools \$21 Million." *Chicago Tribune*.
- Mensch, Barbara S., and Denise B. Kandel. 1988. "Dropping Out of High School and Drug Involvement." *Sociology of Education* 61:95-113.
- Moffitt, Terrie E. 1993. Adolescence-limited and life-course persistent adolescent behavior: A developmental taxonomy. *Psychological Review* 100:674-701.
- , 1994. Natural histories of delinquency. In Weitekamp, Elmer G. M., and Hans-Jurgen Kerner (Eds.) *Cross-National Longitudinal Research on Human Development and Criminal Behavior*. Dordrecht: Kluwer Academic.
- Morrow, George. 1986. "Standardizing practice in the analysis of school dropouts." *Teachers College Record* 87:342-355.
- Mukherjee, Satyanshu Kumar. 1971. "A typological study of school status and delinquency." Unpublished doctoral dissertation. University of Pennsylvania.

- Murnane, Richard J. 1999. "Do Male Dropouts Benefit From Obtaining a GED, Postsecondary Education, and Training?" *Evaluation Review* 23:475-503.
- Nagin, Daniel S. 1999. Analyzing developmental trajectories: Semi-parametric, group-based approach. *Psychological Methods* 4:139-177.
- , 2004. Response to "methodological sensitivities to latent class analysis of long-term criminal trajectories." *Journal of Quantitative Criminology* 20:27-35.
- , 2005. *Group-based modeling of development*. Cambridge, MA: Harvard University Press.
- Nagin, Daniel S., Linda Pagani, Richard E. Tremblay, and Frank Vitaro. 2003. Life course turning points: The effect of grade retention on physical aggression. *Development and Psychopathology* 15:343-361.
- Nagin, Daniel S. and Richard E. Tremblay. 1999. Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development* 70:1181-1196.
- Newcomb, Michael, and Peter M. Bentler. 1988. *Consequences of Adolescent Drug Use: Impact on the Lives of Young Adults*. Newbury Park, CA: Sage.
- Nye, F. Ivan and James F. Short, Jr. 1957. "Scaling Delinquent Behavior." *American Sociological Review* 22:326-331.
- Osgood, D. Wayne, Laura L. Finken, and Barbara J. McMorris. 2002. "Analyzing multiple-item measures of crime and deviance II: Tobit regression analysis of transformed scores." *Journal of Quantitative Criminology* 18:319-347.
- Osgood, D. Wayne, Barbara J. McMorris, and Maria T. Potenza. 2002. "Analyzing multiple-item measures of crime and deviance I: Item response theory scaling."

Journal of Quantitative Criminology 18:267-296.

Osgood, D. Wayne, and David C. Rowe. 1994. "Bridging criminal careers, theory, and policy through latent variable models of individual offending." *Criminology* 32:517-554.

Paternoster, Raymond, Shawn Bushway, Robert Brame, and Robert Apel. 2003. "The effect of teenage employment on delinquency and problem behaviors." *Social Forces* 82:297-335.

Paternoster, Raymond, Charles W. Dean, Alex Piquero, Paul Mazzerolle and Robert Brame. 1997. Generality, continuity, and change in offending. *Journal of Quantitative Criminology* 13:231-266.

Polk, Kenneth, Christine Adler, Gordon Bazemore, Gerald Blake, Sheila Cordray, Gary Coventry, James Galvin, and Mark Temple. 1981. "Becoming Adult: An Analysis of Maturational Development from Age 16 to 30 of a Cohort of Young Men." Final Report of the Marion County Youth Study. Eugene, OR: University of Oregon.

Raudenbush, Steven W. 2001. "Toward a coherent framework for comparing trajectories of individual change." Pp. 35-64 in L. Collins and A. Sayer (Eds.) *New methods for the analysis of change*. Washington, DC: The American Psychological Association.

Robins, Lee N. 1978. Sturdy childhood predictors of adult antisocial behavior: Replications from longitudinal studies. *Psychological Medicine* 8:611-622.

Rönkä, Anna, Sanna Oravala and Lea Pulkkinen. 2002. "'I met this wife of mine and things got onto a better track' Turning points in risk development." *Journal of*

Adolescence 25:47-63.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The central role of the propensity score in observation studies for causal effects." *Biometrika* 70:41-55.

-----, 1985. "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score." *American Statistician* 39:33-38.

Rowe, David C., D. Wayne Osgood, and Alan W. Nicewander. 1990. "A latent trait approach to unifying criminal careers." *Criminology* 28:237-70.

Rumberger, Russell W. 1987. "High school dropouts: A review of issues and evidence." *Review of Educational Research* 57:101-121.

Rutter, Michael. 1987. "Psychosocial resilience and protective mechanisms." *American Journal of Orthopsychiatry* 57:316-331.

-----, 1996. Transitions and turning points in developmental psychopathology: As applied to the age span between childhood and mid-adulthood. *International Journal of Behavioral Development* 19:603-626.

Rutter, Michael and Marjorie Rutter. 1993. *Developing Minds: Challenge and Continuity across the Life Span*. New York: BasicBooks.

Samejima, Fumiko. 1969. "Estimation of latent ability using a response pattern of graded scores." *Psychometrika Monograph Supplement* 17.

Sampson, Robert J. and John H. Laub. 1993. *Crime in the Making: Pathways and Turning Points Through Life*. Cambridge, MA: Harvard University Press.

-----, 1997. A life-course theory of cumulative disadvantage and the stability of delinquency. In Terence P. Thornberry (ed.), *Developmental Theories of Crime and Delinquency*. New Brunswick, N.J.: Transaction.

- Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, and Milagros Nores. 2005. *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*. Ypsilanti, MI: High/Scope Press.
- Sellin, Thorsten and Marvin E. Wolfgang. 1964. *The measurement of delinquency*. Montclair, NJ: Patterson Smith.
- Smith, Jeffrey and Petra Todd. 2005. "Does matching overcome LaLonde's critique of nonexperimental methods?" *Journal of Econometrics* 125:305-353.
- Staff, Jeremy and Christopher Uggen. 2003. The fruits of good work: Job quality and adolescent deviance. *Journal of Research in Crime and Delinquency* 40:263-90.
- Swanson, Christopher B. and Duncan Chaplin. 2005. *Counting High School Graduates When Graduates Count: Measuring Graduation Rates Under the High Stakes of NCLB*. Washington, D.C.: The Urban Institute.
- Sweeten, Gary. 2004. "School dropout and subsequent offending: Distinguishing selection from causation." Unpublished masters thesis. University of Maryland.
- , 2006. Who will graduate? Disruption of high school education by arrest and court involvement. *Justice Quarterly*, forthcoming.
- Sweeten, Gary, Shawn Bushway, and Raymond Paternoster. 2005. "Does dropping out of high school mean dropping into delinquency?" Unpublished manuscript.
- The Sentencing Project. 2004. *Facts about Prisons and Prisoners*.
- Thornberry, Terence P., Melanie Moore, and R.L. Christenson. 1985. The effect of dropping out of high school on subsequent criminal behavior. *Criminology* 23:3-18.
- Thornburgh, Nathan. April 17, 2006. "Dropout Nation." *Time* Pp. 30 -40.

- Tyler, John H., Richard J. Murnane, and John B. Willett. 2003. "Who Benefits From a GED? Evidence for Females from High School and Beyond." *Economics of Education Review* 22:237-247.
- U.S. Census Bureau. 2002. *Statistical Abstract of the United States*.
- U.S. Department of Education. 2001. *Dropout Rates in the United States: 2000*.
- Uggen, Christopher. 2000. Work as a turning point in the life course of criminals: A duration model of age, employment, and recidivism. *American Sociological Review* 65:529-546.
- Voelkl, Kristin E., John W. Welte, and William F. Wieczorek. 1999. "Schooling and delinquency among white and African American adolescents." *Urban Education* 34:69-88.
- Warr, Mark. 1998. Life-course transitions and desistance from crime. *Criminology* 36:183-216.
- Williams, Christine L., and Ernst L. Wynder. 1993. "A Child Health Report Card: 1992." *Preventative Medicine* 22:604-628.
- Wright, John Paul and Francis T. Cullen. 2004. "Employment, Peers, and Life-Course Transitions." *Justice Quarterly* 21:183-205.
- Zagorsky, Jay L., and Rosella Gardecki. 1998. What have researchers learned from the National Longitudinal Surveys? *Journal of Economic and Social Measurement* 25:35-57.