University of Cincinnati

Date: 2/11/2019

I, Ian A Silver, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Criminal Justice.

It is entitled:

Genetically adjusted propensity score matching: A proposal of a novel analytical tool to help close the gap between non-experimental designs and true experiments in the social sciences.

Student's name: **Ian A Silver**

This work and its defense approved by:

Committee chair: Joseph Nedelec, Ph.D.

Committee member: J.C. Barnes, Ph.D.

Committee member: Joshua Cochran, Ph.D.

Committee member: George Richardson, Ph.D.

32852

Genetically adjusted propensity score matching: A proposal of a novel analytical tool to help close the gap between non-experimental designs and true experiments in the social sciences.

A dissertation submitted to the
Graduate School
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

In the School of Criminal Justice
of the College of Education, Criminal Justice, and Human Services

By

Ian A. Silver

Dissertation Committee: Joseph L Nedelec, Ph.D. (Chair)
J. C. Barnes, Ph.D.
Joshua C. Cochran, Ph.D.
George B. Richardson, Ph.D.

ABSTRACT

Objectives: Social scientists employ various statistical techniques to approximate the causal association between two interrelated constructs. Although these methodologies have been useful for the advancement of knowledge, the limitations associated with preceding statistical techniques limit the ability of scholars to approximate causal associations within some conditions. As such, the current study provides a new statistical technique designed to approximate causal associations independent of observed genetic and environmental confounders.

Methods: Genetically adjusted propensity score matching (GAPSM) represents an innovative iteration of propensity score matching (PSM) designed to integrate environmental and genetic factors into the matching process. By using polygenic risk scores, future scholars can estimate genetically adjusted propensity scores (GAPS) through the implementation of two distinct statistical processes. To demonstrate the validity of the GAPSM approach, the current study employs simulation analyses to compare the point estimates derived from a post-GAPSM model to the point estimates derived from a post-PSM model and an MZ difference score model.

Results: The results of the simulation analyses demonstrated that when environmental measures that explain a larger portion of the variance in a treatment condition are introduced into the GAPSM approach, post-GAPSM models approach the true point estimate more closely than the point estimates derived from a post-PSM model and an MZ difference score model.

Conclusions: Overall, the findings demonstrate that the GAPSM approach can be useful when assessing the causal effects of treatment conditions on subsequent phenotypes by adjusting for observed environmental and genetic factors. Within the social sciences, this method could provide substantive advancements in our understanding of causal effects. Specifically, GAPSM represents another tool social scientists can use to conduct rigorous genetically sensitive examinations of the etiological influence of environmental factors on human behavior.

Dedication Page

To my Mom, Dad, and Wife
You gave me everything and I gave you a book.

ACKNOWLEDGEMENTS

I would like to begin by thanking the friends that stood by me throughout my graduate studies. I would like to thank my friends from Brooklyn, from my undergraduate studies at the University of Cincinnati, and the friends that I have made during my graduate studies at the University of Cincinnati.

I would also like to thank the University of Cincinnati Corrections Institute (UCCI), with a special thank you to Myrinda Schweitzer, Carrie Sullivan, and Jamie Newsome. The opportunities you afforded to me and the guidance you have given me was essential to my development as a scholar. I will happily always be a part of the UCCI family.

Furthermore, I would like to acknowledge all of the professors who have had an impact on me. Of which, I would like to directly thank Edward Latessa, Chris Sullivan, and John Wooldredge for their hard work and devotion to my academic career.

JC Barnes, Joshua Cochran, and George Richardson. Thank you for agreeing to be on my dissertation committee and challenging me every step of the way. I know that a methodological dissertation isn't the most ideal read, but the importance of the methodology outweighs the convoluted discussions. Additionally, without you my dissertation wouldn't have succeeded at demonstrating the validity and usefulness of the GAPSM methodology. Thank you for all of your help on my dissertation and in my career. Don't expect for this to be the last time you hear from me.

Joe, Kate, and Clementine, and Rob. You have played an indescribable role in my life. Joe and Rob, you were always role models to me and helped me develop into the person I am today. Additionally, Kate you have been an amazing sister-in-law who has done nothing but support me. Thank you for everything that you have done and everything you will do in the future. Additionally, I would like to thank everyone in my family who has been there for me when I needed it the most.

Mom, I could not have done this without your support. My doctoral degree signifies as much my hard work as yours. Thank you for being there every time I needed you, thank you for trusting in me even when I wasn't the best child, and thank you for supporting me even when I told you that I was going to school for another four years. I could not have asked for a better mother, friend, and role model. As you have done, I will always be there for you if you ever need anything. Thank you for always loving me.

Dad, thank you for everything you have done for me throughout my education and my life. You have always encouraged and supported my dreams. No matter what idea, goal, or dream I brought to you, I always knew that I would be reassured that I could achieve whatever goal I put my mind to. Lynn, I would also like to thank for your encouragement and enthusiasm throughout my life. Thank you for always loving me.

Six years ago you agreed to be my undergraduate research advisor, stupidly might I add, and since then you have not been able to get rid of me. Joseph Nedelec, while it wasn't always easy, it has been a remarkable six years, in which we have accomplished so much. Although I don't believe this will be the pinnacle of our accomplishments, this dissertation and my graduate studies as a whole could not have been accomplished without your guidance. You have been an amazing mentor and friend, who has encouraged me to develop my own research interests, create manuscripts outside the realm of your own expertise, and integrate biological and social mechanisms to better explain antisocial behavior in unique environments. This dissertation is a product of my hard work and your expertise. Thank you so much. I would also like to thank Amanda, Anna, and Robert Nedelec for allowing me to take time away from their husband/dad so I can be a successful student and scholar.

Last but not least I would like to thank my wonderful wife, Danielle Silver. Although I might have a Ph.D., I will always be the second smartest person in our household. You are one of the most amazing people that I have ever met and I could not have achieved this without your love and support. Additionally, I am happy that you have had faith and trusted me since day one of my doctoral studies. I know that it has been difficult at times, but we have always been there for each other and I swear there is a light at the end of the tunnel. You have supported me throughout this process and similarly I will go to the end of the world to help you achieve your dreams. This Ph.D., however, should be awarded to both of us for our hard work and devotion. If only you registered for classes! I love you and I am happy that I can share my life and Ph.D. with you.

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

CHAPTER 1: STATEMENT OF THE PROBLEM

Arguably, contemporary criminology is no closer to determining the causal effects between variables than criminological scholarship in the early 20th century. Although theories and statistical analyses have become more complex, criminologists generally choose more simplistic analyses that often demonstrate little more than the association between two variables (e.g., Gray et al., 2015; Intravia et al., 2017; Makarios, Cullen, and Piquero, 2017). Furthermore, the abundance of scholarship employing simple statistical techniques has manifested itself in over-interpretations of statistical associations (e.g., Restivo and Lanier, 2015; Watts, 2018).[1] For example, the simplistic statistical techniques (e.g., statistical control cross-sectional models) employed by Gottfredson and Hirschi (1990) generated the broad conclusion that "high self-control effectively reduces the possibility of crime" (pg. 89) and "the major cause of low self-control thus appears to be ineffective child-rearing" (pg. 97). Besides interpreting statistical associations causally, Gottfredson and Hirschi (1990) effectively ignored all of the potential biological and environmental predispositions associated with childrearing, self-control, and antisocial behavior (Harris, 2011).

Furthermore, these over-interpretations of statistical associations have generated the inability to disprove theoretical perspectives. For instance, in their analysis of the contextual effects and mitigating factors of labeling theory, Restivo and Lanier (2015) estimated an OLS regression analysis where future criminal behavior was regressed on arrest and attitudes of self, which resulted in the conclusion that "This finding is consistent with the current labeling literature and continues to provide confirmation of the idea that formal labeling will result in

[1] In the current context, over-interpretation refers to the process in which scholars portray their findings as causal when causality cannot be determined.

increases of future delinquent behavior" (pg. 132). Though there is a statistical association between arrest and future criminal behavior, these results do not confirm the idea that formal labeling will increase future criminal behavior. Specifically, by analyzing non-experimental data the scholars allowed for substantive biases to remove their ability to confirm or identify a causal association between two or more concepts. Moreover, the theoretical model of labeling theory is highly confounded by individual predispositions preceding criminal justice interventions, which substantially increase the likelihood of self-selection influencing the observed association and restricts the potential causal pathway between criminal labels and future criminal activity (Farrington, 2005; Piquero et al., 2007; Smith and Paternoster, 1990; Nedelec and Silver, 2018).

As a substantial source of bias in such analyses, self-selection refers to the conscious or unconscious process in which individuals' predispositions influence their probability of selecting into an environment or situation. Due to the inability to implement true experimental designs that allow scholars to examine causality, self-selection has had a substantive influence on the accumulation of knowledge within the field of criminology. Currently, various advanced statistical techniques are used to address this limitation through the approximation of a counterfactual condition. The approximation of a counterfactual condition allows scholars to approach causality by emulating the processes common within quasi-experimental designs. Remarkably, without these advanced statistical techniques scholars cannot interpret causal associations between concepts, which often influences the observed over-interpretation of results within criminology.

In an effort to address this issue beyond contemporary methodologies, the current study creates a statistical methodology *genetically adjusted propensity score matching* (*GAPSM*) applicable to the difficulties facing contemporary criminological research. To be specific, whereas statistical associations provide some information on the relationship between two or

more variables, advances statistical techniques can generally adjust for the biases associated with non-experimental designs. *GAPSM* is a method developed to adjust for some biases biological and social self-selection[2] associated with non-experimental designs.

1.1. Current Focus

The current study focuses on two mechanisms influencing antisocial behavior that are theoretically and empirically supported by an abundance of scholarship: intelligence and educational attainment (e.g., Devenish, Hooley, and Mellor, 2017; Herrnstein and Murray, 1994; Hirschi and Hindelang, 1977; Gonzalez et al., 2016; Hagan and Parker, 1999; Mears and Cochran, 2013; Merton, 1968; Nye, Short, and Olson, 1958; Schwartz and Beaver, 2018; Sutherland and Cressey, 1970). Fundamentally, the theoretical assertions are generally well formulated and empirical examinations provide evidence of a robust statistical association between intelligence and antisocial behavior, and educational attainment and antisocial behavior (e.g., Akers, 2011; Backman, 2017; Demmler et al., 2017; Elliot, and Voss, 1974; Larzelere and Patterson, 1990; Piotrowska et al., 2015; Thornberry and Farnworth, 1982; Moffitt, 1993).Thus, two fundamental research questions are addressed in the current study: (1) What is the causal effect of intelligence on future antisocial behavior?; and, (2) what is the causal effect of educational attainment on future antisocial behavior? These questions are addressed by estimating the relationship between each variable (intelligence and educational attainment) and antisocial behavior using three different statistical techniques. Succeeding the examination of these questions with preexisting methodologies, *genetically adjusted propensity score matching* (*GAPSM*) is proposed as an innovative statistical technique to examine causality while holding both observed genetic and observed environmental factors constant at varying levels of exposure

[2] In the current context, biological self-selection refers to the process in which genetic and neurological processes predispose individuals to a higher probability of experiencing a phenotype or environmental condition. While social self-selection refers to the process in which environmental and societal processes predispose individuals to a higher probability of experiencing a phenotype or environmental condition.

to the treatment condition.[3] The GAPSM proposal is supplemented by a simulation analysis to demonstrate when the GAPSM approach is superior to or inferior to the preexisting methodologies used in contemporary criminology. Section 1.2. provides an outline for the current study.

1.2. Outline

The advancements in knowledge provided by these methodologies permit scholars to draw conclusions beyond the majority of contemporary scholarship. One collection of statistical methodologies allows for the more direct approximation of causal effects between a treatment condition and an outcome of interest by establishing an approximate counterfactual condition. Chapter 2 will demonstrate that only true experimental designs can establish a counterfactual condition where all unobserved variables are held constant across varying levels of exposure to a treatment condition. By holding all unobserved variables constant, scholars can estimate the true causal effects underlying the association between a treatment condition and an outcome of interest. As a result of a general inability to implement true experiments, criminologists are commonly hindered by the potential effects of self-selection bias when estimating causal associations (Heckman, 1990a; Guo and Fraser, 2015). When considering a counterfactual condition, self-selection refers to the process in which predispositions increase or decrease your probability of exposure to the treatment condition. In an effort to approximate the counterfactual condition established by a true experimental design, scholars have generated various statistical techniques to adjust for the effects of self-selection and other potential sources of bias. While various sources of bias exist, the current study will focus on self-selection.

[3] Throughout the current dissertation, the term treatment condition refers to any variable influencing phenotypic differences (e.g., intelligence, educational attainment, and peer delinquency). Furthermore, any concept that is hypothesized to be causally associated with an outcome could be considered as a treatment condition.

Chapter 3 will explain five statistical methodologies commonly used to identify and adjust for the effects of self-selection into a treatment condition. These methodologies include statistical control models, propensity score matching, ACE decomposition models, MZ difference score models, and polygenic risk score models. Statistical control models, such as ordinary least squares (OLS) or binary logistic regression (BLR), are a class of statistical models that employ standard social science methodologies (SSSMs) to adjust for the potential effects of self-selection. Although these models are widely employed in the field of criminology, they generally provide inadequate approximation of the causal association between two variables. Propensity score matching (PSM) enhances upon statistical control models by holding all observed environmental conditions constant at varying levels of the treatment condition, but it is still another example of an SSSM. In contemporary criminology the use of statistical control models and PSM often corresponds with the assumption that biological factors do not predispose individuals to self-select into a treatment condition (Plomin et al., 2013).

Contrasting with this assumption are the various quantitative genetic methodologies used to approximate a counterfactual condition. ACE decomposition models are used primarily to identify the potential genetic and unobserved environmental factors influencing self-selection. These models decompose phenotypic variance in a treatment condition into genetic (h^2), shared (c^2), and non shared environmental factors (e^2). Divergent from ACE decomposition models, MZ difference score models adjust for the unobserved genetic factors increasing or decreasing the probability of exposure to the treatment condition.[4] In contemporary scholarship, MZ difference score models are perceived as the gold standard for establishing genetically-sensitive counterfactual conditions (Plomin et al., 2013; Vitaro et al., 2009). Although perceived as the

[4] It should be noted that scholars using these techniques assume that the model adjusts for shared non-genetic factors (e.g., race, age, sex; Plomin et al., 2013). Generally, beyond basic demographics (e.g., race, age, sex) scholars have rarely assessed how well MZ difference score models adjust for some environmental factors (i.e., the shared environments). For a more detailed discussion of this issue please see Plomin et al., 2013.

gold standard, MZ difference score models suffer limitations regarding statistical power and potential limitations in terms of generalizability (Plomin et al., 2013; Vitaro et al., 2009). In an effort to address these limitations, scholars have developed polygenic risk score models. Polygenic risk score models are analytical strategies that statistically control for the influence of observed genetic factors on the association between a treatment condition and an outcome of interest. When used as a statistical control, polygenic risk scores suffer the same limitations associated with the majority of statistical control models (i.e., they provide inadequate approximations of the causal association between two variables). Specifically, without satisfying the assumptions associated with the statistical control models, the introduction of a polygenic risk score cannot provide an approximation of the true counterfactual condition.

The final section of Chapter 3, proposes a new methodological strategy to approximate the counterfactual condition that would be established by a true experiment. This technique combines polygenic risk scores with propensity scores to create *genetically adjusted propensity scores* (*GAPS*). As designed, GAPS capture the variance in the treatment condition predicted by both the polygenic risk score and the propensity score. Consistent with propensity scores, participants should be matched using GAPS to hold the observed genetic and observed environmental factors constant across varying levels of exposure to the treatment condition. This technique is labeled as *genetically adjusted propensity score matching* (*GAPSM*). As outlined, GAPSM should provide a superior counterfactual condition to the other techniques described because participants are matched using estimates produced from predicting exposure to a treatment condition with *both* observed genetic *and* observed environmental factors.

As reviewed in Chapter 4, the current study demonstrates the validity of the GAPSM technique by addressing two research questions: First, what is the causal effect of intelligence on antisocial behavior; And second, what is the causal effect of educational attainment on antisocial

behavior? These research questions can be assessed using the four statistical techniques described in Chapter 3. Furthermore, the two research questions were developed from contemporary theoretical and empirical literature in criminology suggesting that both intelligence and educational attainment are causally associated with antisocial behavior (e.g., Akers, 2011; Backman, 2017; Demmler et al., 2017; Gonzalez et al., 2016; Hagan and Parker, 1999; Merton, 1968).

The current study is demarcated into four distinct studies. First, linear regression models and ACE decomposition models will be employed to empirically examine the potential factors influencing self-selection into intelligence or educational attainment (i.e., the TC). Furthermore, the first study will estimate the baseline effects of intelligence and educational attainment on future antisocial behavior. Second, propensity score matching (PSM) will be used to adjust for social self-selection during the estimation of the causal association between intelligence and antisocial behavior, and the causal association between educational attainment and antisocial behavior.[5] More specifically, all observed environmental factors will be held constant at varying levels of intelligence and educational attainment. Third, MZ difference score models will be used to adjust for the effects of unobserved genetic factors and shared environmental factors that may influence exposure to varying levels of intelligence and educational attainment. Fourth, a simulation analysis will be conducted, where post-GAPSM estimates are compared to post-MZ difference estimates and unconfounded post-PSM estimates to evaluate the relative proximity to a true point estimate (1.00). The results of which should demonstrate the conditions in which the GAPSM technique approaches the true point estimate closer than the preexisting methodologies.

Chapter 5 and Chapter 6 provide an overview of the data, sample, methods, and results produced from studies 1,2,3, and 4. Studies 1, 2, and 3 employ data collected during the National

[5] Without the introduction of a genetically informed covariate (e.g., polygenic score), PSM cannot adjust for genetic self-selection.

Longitudinal Study of Adolescent to Adult Health (Add Health). Studies 1, 2, and 3 will primarily focus on Wave I, Wave III, and Wave IV to establish temporal ordering between the measures. The methods and results sections will be demarcated into four studies (following the template from Chapter 4). Chapter 7 will discuss the theoretical and empirical implications associated with the findings. Furthermore, Chapter 7 will discuss the empirical validity of the GAPS and GAPSM technique, the applicability of GAPSM to examinations of contemporary criminological perspectives, and the limitations associated with the current study and the GAPSM technique.

CHAPTER 2: CAUSALITY AND SELF-SELECTION

The complexity of causality is fundamentally difficult to comprehend. Encompassed in the complexity is the inability to agree upon a stringent definition of causality. Generally, scholars define causality as the necessary, sufficient, and noncircular condition, where an exogenous variable generates the presence of an endogenous variable (Holland and Rubin, 1988; Lewis, 1973; Spirtes, Glymour, and Scheines, 2000). To state differently, the endogenous variable is only present when the exogenous variable occurs first (Lewis, 1973). The definition of a causal condition can be demonstrated through four figures.

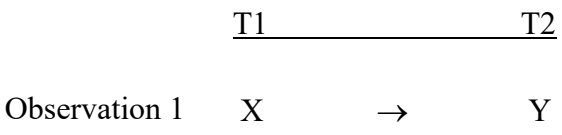


Figure 2.1
A Demonstration of a Causal Condition.

Figure 2.1 displays the common theoretical perception of causality. In this condition, *X* causes the existence of *Y*.[6] Note that "T1" and "T2" reference the time that *X* and *Y* were observed. This simplification is difficult to determine because of the three conditions associated with theoretically establishing causality (i.e., necessary, sufficient, and noncircular association; Lewis, 1973).[7] First, necessity refers to the idea that variance in *Y* cannot be observed without the existence of variance in *X* (Kun and Weiner, 1973). Note that this statement assumes that multiple *X*s do not cause *Y*.

[6] Throughout the current dissertation *X* refers to exogenous variables and *Y* refers to endogenous variables.
[7] The theoretical establishment of causality is divergent from the empirical establishment of causality (i.e., statistical association, temporal ordering, non-spurious association).

	T1		T2
Observation 1	X	→	Y
Observation 2			Y

Figure 2.2
A Demonstration of a Condition Failing to Achieve Necessary Causal Association.

Figure 2.2 represents the failure to achieve necessity. In this condition, we observe variance in *Y* during two distinct observations (i.e., Observation 1 and Observation 2). As displayed above, Observation 1 demonstrates a "causal" association between *X* and *Y*. Nevertheless, Observation 2 determines that *Y* can occur independent of *X*. The independent occurrence of *Y*, or the presence of *Y* without *X*, demonstrates a failure to achieve necessity (Kun and Weiner, 1973; Spirtes et al., 2000).

	T1		T2
Observation 1	X	→	Y
Observation 2	Z	→	Y
Observation 3		→	Y

Figure 2.3
A Demonstration of a Condition Failing to Achieve Sufficient Causal Association.

Figure 2.3 displays a condition where *X* is not a sufficient cause of *Y*. While Observation 1 demonstrates a potentially causal association between *X* and *Y*, Observation 2 demonstrates a potentially causal association between *Z* and *Y*, and Observation 3 demonstrates that *Y* can be observed independent of both *X* and *Z*. In all three observations, *Y* can be observed as the product of a preceding variable, but *X* is not a sufficient cause of *Y* given the similar outcome when *Z* is introduced or when no independent variable is introduced (Lewis, 1973; Suppes, 1970). This is a

function of the inability to determine if variance in *X* or if variance in *Z* is causing the observed variance in *Y*.

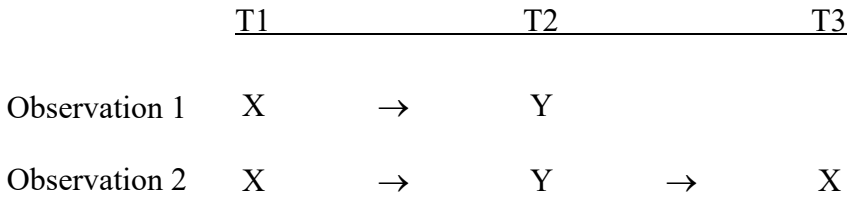


Figure 2.4
A Demonstration of a Circular Condition.

Finally, as demonstrated by Figure 2.4, *X* is not a sufficient cause of *Y* because the association between the two variables is circular. A circular association refers to the condition in Observation 2, where *X* causes *Y,* but *Y* also causes *X*. In a circular condition, *X* cannot be a cause of *Y* because *Y* can precede *X* in some observations. Although the definition of causality provides a theoretical framework for identifying causal mechanisms, empirically demonstrating a causal association between two variables requires scholars to satisfy three criteria of causality: *X* and *Y* must have a logical theoretical and empirical association, *X* must precede any observation of *Y*, and *X* and *Y* must possess a non-spurious association (Shadish et al., 2002; Suppes, 1970; Vigen, 2015). Again, these assumptions can be best demonstrated through a series of figures.

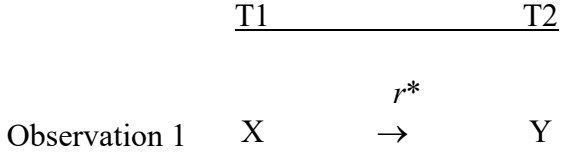


Figure 2.5A
Demonstration of a Theoretical and Empirical Association.

Figure 2.5 – which is almost identical to Figure 1 – represents an empirical association between two variables: *X* and *Y*. As indicated by the *r** (the common symbol for a statistically significant correlation coefficient) Observation 1 results in evidence suggesting an empirical association between two variables. Note that unlike Figure 1, we are examining the association between *X* and *Y* in an empirical, rather than theoretical, framework. Although an empirical association is observed in Figure 5, this association could remain illogical.[8]

	T1		T1
		*r**	
Observation 1	X	→	Y

Figure 2.6
A Demonstration of Temporal Order.

Figure 2.6 demonstrates a condition where the empirical association between *X* and *Y* does not satisfy the temporal ordering criterion. Let us assume that the association between *X* and *Y* is theoretically supported. As indicated by *r**, an empirical association between *X* and *Y* was also observed. While the association between *X* and *Y* is both theoretically and empirically supported, the observation of *X* and *Y* during the same time period (T1) typically generates an inability to understand the direction of causality (Miller, 1999; Shadish et al., 2002). Consistent with cross-sectional examinations, we are unable to determine if *X* caused *Y* or if *Y* caused *X* (Morgan and Winship, 2015).

[8] For example, as demonstrated by Vigen (2015), average per-capita cheese consumption is highly correlated (r = .94) with the number of people who died becoming tangled in their own bed sheets. Although an empirical association was observed (r = .94), it is theoretically illogical to suggest that cheese consumption causes individuals to become tangled and die in their bed sheets (Vigen, 2015).

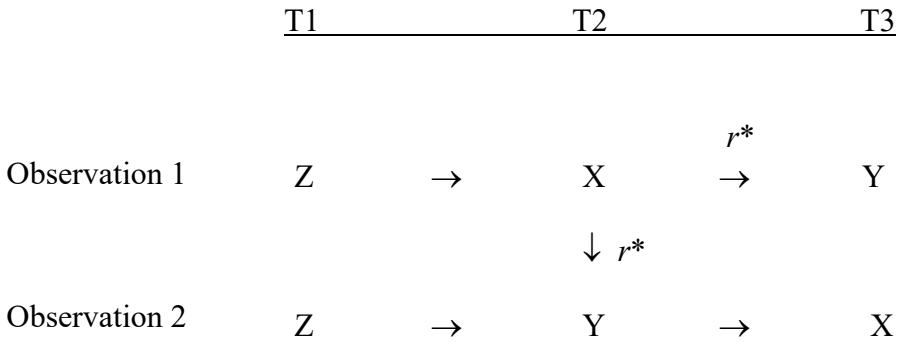


Figure 2.7
A Demonstration of a Spurious Association.

Figure 2.7 demonstrates a spurious association. First, let us assume that the association between *X* and *Y* satisfy the other two criteria associated with empirically demonstrating causality. The third criterion of causality requires the demonstration of a non-spurious association between the *X* and *Y* (Shadish et al., 2002). A non-spurious association is conceptualized as a relationship that is not influenced by any additional variables (Singleton and Straits, 2010). The letter "Z" generally represents the hypothetical confounding variable. In Figure 7, we observe an empirical association between *X* and *Y* (represented by *r**) but this association appears to be spurious because of an unobserved variable (*Z*). Generally, the third criterion of causality (a non-spurious association) is the most difficult to satisfy because a potentially infinite number of unobserved variables could render any observed association spurious (Singleton and Straits, 2010). In a theoretical sense, the only way to satisfy the third criterion of causality is through an experimental design which relies on the counterfactual framework to theoretically exclude the influence of any unobserved variables (Morgan and Winship, 2015).

2.1. Counterfactual Framework: A Theoretical Proposition

Simply put, the counterfactual framework is the parenthetical what if statement.[9] In a research sense, the counterfactual refers primarily to "what would we observe if *X* had never occurred?" (Morgan and Winship, 2015; Roese, 1997). This statement can theoretically dissolve into a series of hypothetical situations where *X* (hereafter referred to as a treatment condition)[10] had never occurred (Morgan and Winship, 2015). For illustration, let us conduct a medical study, where 50 individuals with a treatable illness are exposed to a treatment condition. We observed that the illness disappeared succeeding exposure to the treatment condition. Though this observation appears to causally link the treatment condition to the disappearance of the illness, we have failed to demonstrate that the treatment condition is casually associated with the disappearance of the illness. In the counterfactual framework, we can determine that the treatment condition is causally associated with the disappearance of the illness by observing a parallel universe, where the only difference is exposure to the treatment condition. Causality can be established by observing the counterfactual condition because only variation in the treatment condition exists. Figure 2.8 represents the parallel universe counterfactual condition:

Assignment (A)	T1	T2	T3
50 participants$_A$	Y_{A1}	TC	Y_{A2}
50 participants$_B$	Y_{B1}		Y_{B2}

Figure 2.8
Parallel Universe Example of a Counterfactual Condition.

Figure 2.8 represents our parallel universe example, where the same 50 participants are observed in each universe ("A" or "B") and the only difference between the two universes is

[9] The counterfactual framework represents the primary theoretical guidance for the development of the GAPSM methodology.
[10] A treatment condition refers to a single independent variable hypothesized to influence an outcome of interest. The term treatment condition is used to conform with later discussions about statistically establishing a counterfactual condition.

exposure to the treatment condition at T2. Consistent with our example, Y_1 represents the observation of the illness, Y_{A2} represents the disappearance of the illness, and Y_{A2} represents the persistence of the illness. In universe A the 50 participants are exposed to the treatment condition and we observe the disappearance of the illness at T3, but in universe B the 50 participants are not exposed to the treatment condition and we do not observe the disappearance of the illness at T3. Theoretically, this parallel universe example establishes a causal relationship between the treatment condition and the disappearance of the illness because it establishes that the treatment condition is both a necessary and a sufficient cause of the disappearance of the illness, when holding all unobserved variables (*Z*) constant. To state differently, without exposure to the treatment condition the disappearance of the illness did not occur. Universe B represents the counterfactual condition in this example. However, given that the observation of a parallel universe is currently impossible scholars are required to establish counterfactual conditions in less ideal ways.

The fundamental strategy for establishing a counterfactual condition is the experimental design (Lewis, 1979; Roese, 1997). The experimental design requires the random assignment of participants into two groups: treatment and control (Jackson, 1977). The treatment group includes the participants exposed to the TC, while the control group represents the counterfactual condition (i.e., what would be observed if *X* had never occurred?; Roese, 1997; Shadish et al., 2002). Furthermore, since a parallel universe cannot be observed, we assume that all other unobserved variables (*Z*) are normally distributed in the population and are held constant through the process of random selection and random assignment (Shadish et al., 2002). This assumption is conditional on the hypothesized sampling distribution, where the aggregation of multiple samples would provide the true association between the treatment condition and an observed outcome (Morgan and Winship, 2015). Returning to the example provided above, let us take a

sample of 100 participants with an illness. We randomly assign 50 participants to the treatment group and 50 participants to the control group under the assumption that their characteristics are normally distributed in the population. Below represents the experimental design:

	A_R	T1	T2	T3
Treatment	50_p	Y_1	TC	Y_2
Control	50_p	Y_1		Y_2

Figure 2.9
Example of a True Experiment: Participants are Randomly Assigned.

Figure 2.9 represents the true experimental design with randomized assignment (A_R) to the treatment and control groups. At T1 we observe the illness for both the treatment and control groups, at T2 the treatment group is exposed to the treatment condition (TC), and at T3 we observe differences in the persistence of the illness (ΔY) between the treatment and control groups. Consistent with the parallel universe example, the true experiment suggests a causal association between the treatment condition and the disappearance of the illness (Shadish et al., 2002). This conclusion can be derived from the assumption that our control group represents the counterfactual condition for our treatment group (Roese, 1997). While threats to validity can still impact the observed association, it is safe to assume that the random assignment of participants to the treatment and control groups randomly distributes unobserved characteristics potentially associated with both Y_1 and Y_2 (indicated by Z in the figures above) between the two groups (Shadish et al., 2002). The random distribution of unobserved characteristics to both the treatment and control groups allows scholars to assume that these factors will have limited

impact on the observed association succeeding replication of the study (Spirtes et al., 2000).[11] Of importance, however, is that although one true experiment suggests a causal association between a treatment condition and an outcomes of interest, the aggregation of replications can provide evidence indicating a causal association, or a lack thereof, between a treatment condition and an outcome of interest (Spirtes et al., 2000).

The ability to implement a true experimental design is quite limited in the social sciences, and this is especially germane in criminology (Singleton and Straits, 2010). Primarily, this limitation is a result of the potential unethical circumstances associated with randomly assigning individuals to a treatment condition. Generally, as a consequence of this inability social scientists rely on non-experimental methodologies to examine the association between a treatment condition and various outcomes of interest. Unlike experimental research, non-experimental research designs remove the ability to establish causal associations between a treatment condition and outcome of interests as a function of the inability to control for unobserved predictors of the treatment condition (Spirtes et al., 2000). If unobserved characteristics do influence the observed association between a treatment condition and an outcome of interest it is likely a consequence of various threats to validity (e.g., regression, maturation, and attrition; Shadish, Cook, and Campbell, 2002). Perhaps the most salient threat to validity in non-experimental designs is that of self-selection bias (Lewis, 1979). The issue of self-selection bias only arises in non-experimental designs and is imperative to control if one wishes to establish of causality.

2.2. Self-Selection Bias

In the context of non-experimental designs, self-selection refers to the process in which individual predispositions increase or decrease the probability of selection into the treatment and

[11] Various factors (e.g., attrition, history, and limited generalizability) could serve as limitations for experimental designs and impact the validity of the findings.

control groups, which biases the observed association between a treatment condition and an outcome of interest (Spirtes et al., 2000). Both genetic and non-genetic predispositions influence the probability of selection into treatment and control groups (Heckman, 1990b). Let us rely on the illness example again.

	A_{NR}	T1	T2	T3
Treatment	50_p	Y_1	TC	Y_2
Control	50_p	Y_1		Y_2

Figure 2.10
Example of a Non-Experimental Design: Participants Were Not Randomly Assigned.

Figure 2.10 represents a non-experimental design, where non-random assignment (A_{NR}) was used to classify the 100 participants into the treatment and control groups. Note that even though 50 participants were assigned to the treatment and control groups, their selection into these groups was partially based on the individuals' proclivities. The same outcome is observed, the disappearance of an illness after exposure to the treatment condition, but causality cannot be determined. The fundamental reason causality cannot be determined is that individual predispositions could have increased the probability of exposure to the treatment condition and the probability of the disappearance of an illness. Stated differently, non-experimental designs remove the ability to establish causality, resulting from the inability to hold all unobserved variables constant between the two groups (Shadish et al., 2002). Theoretically, the unobserved variables increasing or decreasing the probability of exposure to the treatment condition can be demarcated into biological and environmental factors (Morgan and Winship, 2015).

Formula 2.1 is a simple representation of the potential confounding factors theoretically establishing a spurious association between a treatment condition and Y.

[Formula 2.1]

$$Y = TC + Z(G + E)$$

In this equation, $Z(G + E)$ represents the genetic (G) and environmental (E) factors predisposing an individual to the treatment condition and the observed outcome, TC represents the treatment condition, and Y represents the observed outcome. As specified, any observed variation in Y is the sum of $TC + Z$, where Z could be theoretically observed or unobserved (Morgan and Winship, 2015). It is impossible to account for all observed or unobserved potentially confounding variables in a non-experimental design given the infinite number of variables that could account for causal association between treatment condition and Y (Morgan and Winship, 2015). In contemporary scholarship there are various methods to approximate the effects of $Z(G + E)$ on the observed association between the outcome of interest and the treatment condition.

The most common method is to establish a quasi-experimental design, where some of the potentially confounding variables could account for association between the treatment condition and the outcome of interest are held constant (Morgan and Winship, 2015). A quasi-experimental design refers to an empirical assessment where a counterfactual condition is approximated using statistical, or methodological, techniques (Singelton and Straits, 2010). Generally, quasi-experimental designs create counterfactual conditions where the estimates produced resemble the estimates that would be observed in a true experimental design (Morgan and Winship, 2015). Thus, even though causality remains unattainable, quasi-experimental designs allow scholars to approach (to a certain degree) the causal effects of the treatment condition on the outcome of interest (Shadish et al., 2002).

2.2. Self-Selection in Criminology

In the social sciences, especially criminology, scholars generally make assumptions about the effects of social self-selection that may be unwarranted (e.g., Agnew, 1992; Akers, 2011;

Gottfredson and Hirschi, 1990; Sampson and Laub, 1995). While self-selection refers to both sociological and biological factors that could potentially confound an association between two variables, social self-selection refers only to self-selection based upon sociological factors. Specifically, scholars generally assume that social self-selection has limited, or no, impact on the theoretical association between a treatment condition and an outcome of interest. Remarkably, the effects of these assumptions are commonly overlooked in contemporary scholarship. For example, Sampson and Laub (1995) assume that social self-selection has limited impact on the association between marriage and desistance from criminal activity and empirical examinations generally rely heavily on this assumption (e.g., Forrest, 2014; Simons and Barr, 2014; Warr, 1998). To quote Sampson and Laub (1995):

> "And because these relationships obtain within the delinquent group, it is difficult to dismiss the results on the basis of a stability or self-selection argument that antisocial children simply replicate their antisocial behavior as adults – that delinquents invariably continue their interactional styles in adult life, and hence have incompatible relations with family, work, and other institutions of social control." (p. 147)

Thus, some criminological scholars rarely consider establishing a counterfactual condition to examine if social self-selection influences the association between marriage and desistance from criminal activity (e.g., King, Massoglia, and Macmillian, 2007; Sampson, Laub, and Wimer, 2006). Although this is just one example, a variety of the observed associations in criminology could be influenced by the effects of social self-selection. Since a majority of the literature in criminology assumes that social self-selection has limited influence on observed associations, it is possible that poor replication would be observed when employing quasi-experimental designs (Ioannidis, 2005).

In addition to assuming that social self-selection has limited influence on observed associations, criminologists also assume that biological factors rarely increase or decrease the probability of exposure to a treatment condition (see Barnes et al., 2014). Nevertheless, this assumption is generally invalid (Plomin, DeFries, and Loehlin, 1977). Quantitative genetic methodologies have provided important insight into how biological and social self-selection could potentially increase or decrease the probability of exposure to a treatment condition. To reanalyze the validity Sampson and Laub's (1995) assumption, an analysis conducted by Johnson and colleagues (2004) demonstrated that marriage is highly heritable (h^2 = .61) and meta-analyses conducted by Mason and Frick (1994), Rhee and Waldman (2002), and Bergen and colleagues (2007) demonstrated that on average 50 percent of the variation in antisocial behavior is accounted for by genetic factors. In combination, these two results generate the expectation that the observed association between marriage and desistance from criminal activity would be attenuated when accounting for the effects of biological self-selection. Consistent with these expectations, Barnes and Beaver (2012) demonstrated that even though marriage still influenced desistance from criminal activity, the strength of the association was attenuated by approximately 60 percent. Examples of self-selection potentially confounding the association between two sociological concepts exist throughout criminology (e.g., Barnes, Beaver, and Boutwell, 2011; Nedelec, Park, and Silver, 2016; Nedelec, Richardson, and Silver, 2017).

Commonly, scholars can adjust for the effects of self-selection by approximating quasi-experimental research designs through the implementation of various statistical techniques. In the current context, a quasi-experimental research design refers to the process in which participants are not randomly assigned to the treatment condition, but the design has similar structural characteristics to that of a true experiment (Shadish et al., 2002). Generally, quasi-experimental designs can be identified by three characteristics, the effort to identify threats to

internal validity, the effort to control by design, and the effort to make coherent pattern of matches (Shadish et al., 2002). Additionally, various statistical techniques can be implemented where a counterfactual condition is considered to identify threats to internal validity, control by design, and make a coherent pattern of matches can be approximated and used to assess the causal association between multiple concepts. To demonstrate and advance upon the statistical techniques used to establish counterfactual conditions, chapter 3 reviews the common standard social science methodologies (SSSMs) and the quantitative genetic methodologies used to establish counterfactual conditions and proposes a new statistical technique for establishing a counterfactual condition.

CHAPTER 3: STRATEGIES IDENTIFYING AND ADJUSTING FOR SELF-SELECTION

As reviewed in the previous chapter, self-selection bias is the one of the fundamental factors influencing the inability to establish causality within the social sciences. To reiterate, self-selection refers to the process in which biological and environmental conditions increase or decrease the probability of exposure to a treatment condition and an outcome of interest (Guo and Fraser, 2015). Although random assignment into the treatment condition is a method for handling self-selection, sociologists and criminologists can rarely rely on experimental designs to examine the association between a treatment condition and an outcome of interest. Nevertheless, various scholars have developed statistical techniques to establish counterfactual conditions that could potentially approximate the counterfactual condition of a true experiment (Purcell et al., 2009; Rosenbaum and Rubin, 1983; Rosenbaum and Rubin, 1984; Pike et al., 1996). The current chapter reviews five statistical methodologies (i.e., linear regression analysis, propensity score matching, ACE decomposition models, MZ difference score models, and polygenic risk score models) used to approximate a counterfactual condition. Additionally, the final section of the current chapter proposes a new methodology – *genetically adjusted propensity score matching* (*GAPSM*) – which can be used to approximate a counterfactual condition that adjusts for the observed genetic and observed environmental factors influencing self-selection into a treatment condition.

3.1. Standard Social Science Methodologies

Self-selection is fundamental issue in contemporary sociological scholarship. As outlined in Chapter 2, the inability to randomly assign individuals to a treatment or control condition, which is common in sociology and criminology, drastically reduces the capacity to establish causality. Furthermore, those same predispositions likely increase the probability of spuriousness

between a treatment condition and an outcome of interest.[12] To estimate the causal association between a treatment condition and an outcome of interest, non-experimental designs necessitate the establishment of a true counterfactual condition. A true counterfactual condition is the hypothetical situation were only variation in a treatment condition can account for variation in an outcome of interest. Due to the inability to examine criminological treatment conditions with experimental designs, sociologists have typically employed two statistical techniques to adjust for the effects of self-selection.

3.1.1. Ordinary least squares and binary logistic regression (statistical control method)

The current discussion will focus on the introduction of statistical controls into two regression models (e.g., ordinary least squares (OLS) and binary logistic regression (BLR)), which provide the foundational components for the majority of advanced regression models (Fox, 2016). The term statistical control refers to an exogenous variable with a theoretical or empirical association with both the treatment condition and the outcome of interest (Draper and Smith, 2014). Regarding functionality, statistical controls empirically associated with the treatment condition and the outcomes of interest will remove all covariance between the three variables (Draper and Smith, 2014). Only the remaining variation in the treatment condition can predict the remaining variation in the outcomes of interest. Fundamentally, this technique attempts to establish causality by decreasing the probability of spuriousness. Notably, to assure that the introduction of a statistical control removes the appropriate amount of variation in the treatment condition and the outcomes of interest, OLS and BLR requires scholars to satisfy the respective assumptions of the statistical tools (i.e., linearity, heteroscedasticity, and non-random

[12] Throughout the current chapter, I describe the statistical techniques regarding a treatment condition and an outcome of interest. While seemingly convoluted, this provides uniformity when describing divergent statistical analyses and the ability to establish causality.

error; Fox, 2016). If a statistical control violates the respective assumptions an imprecise amount of variation in the treatment condition or the outcome of interest might be removed (Draper and Smith, 2014).

Formula 3.1 represents the bivariate scalar specification for an OLS regression model. Commonly, OLS regression is used to predict variation in a continuous outcome measure. OLS regression minimizes the sum of the squared differences between the dependent and independent variables to probabilistically estimate a linear function (Montgomery, Peck, and Vining, 2012). This calculation provides scholars a linear function with the highest probability of representing the population given the data (Fox, 2016). The bivariate calculation assumes that variation in the treatment condition (TC) can be used to predict the variation in the outcome of interest (Y; Montgomery, Peck, and Vining, 2012). In a multivariate model (represented by Formula 3.2) the variation in the treatment condition (TC) and the variation in the independent variable (statistical control; X) is used to predict the variation in the outcome of interest (Y; Fox, 2016). Consistent with probability theory, any covariance between *treatment condition* and *X*, and *X* and *Y* would be removed preceding the estimation of the linear function, ensuring the establishment of the best linear unbiased estimate (aka, BLUE; Seber and Lee, 2012). Consistent with standard social science methodologies, social scientists often only adjust for the effects of observed environmental factors but not the observed or unobserved genetic and the unobserved environmental factors (Draper and Smith, 2014).

[Formula 3.1]

$$Y = b_0 + b_1 TC_{1i}$$

[Formula 3.2]

$$Y = b_0 + b_1 TC_{1i} + b_2 X_{2i}$$

Statistical controls function similarly in BLR models. Formula 3.3 represents the bivariate estimation of a BLR model. Divergent from OLS technique, BLR is used to predict variation in dichotomous outcomes (Draper and Smith, 2014). Furthermore, unlike OLS regression, BLR relies on the iterative maximum likelihood estimation technique to determine the linear function that best represents the observed data (Montgomery et al., 2012). To estimate a BLR model, the dependent variable must be transformed into a logged odds ($\log(\frac{Y}{1-Y})$), which creates an unbounded distribution of scores (Draper and Smith, 2014). Regarding statistical controls, any covariance between *treatment condition* and *X*, and *X* and *Y* would be removed prior to predicting variation in the outcome of interest (*Y*; Fox, 2016). Again, social scientists commonly only adjust for the effects of observed environmental conditions.

[Formula 3.3]

$$\log(\frac{Y}{1-Y}) = b_0 + b_1 TC_{1i}$$

[Formula 3.4]

$$\log(\frac{Y}{1-Y}) = b_0 + b_1 TC_{1i} + b_2 X_{2i}$$

The statistical control method is widespread in criminology and sociology. While the simplicity of statistical controls makes the technique appealing, two substantive limitations exist when considering the potential effects of self-selection (Fox, 2016).[13] First, the method suffers from an inability to establish a counterfactual condition (Guo and Fraser, 2015).[14] As described in Chapter 2, a counterfactual condition is addressed by the question "what would have occurred

[13] To reiterate, social self-selection refers to the process in which sociological factors increase or decrease the probability of exposure to a treatment condition.
[14] Counterfactual conditions can be estimated with regression based techniques but require the user to satisfy all of the assumptions associated with said technique (Vanderweele, 2015).

in terms of the outcome if the treatment condition was not introduced?" Furthermore, the only way to establish a counterfactual is by randomly distributing, or holding constant, all other exogenous variables at varying levels of exposure to the treatment condition. Predictably, the statistical control technique cannot establish a counterfactual condition, which results from the inability to randomly distribute, or hold constant, all other variables at varying levels of exposure to the treatment condition. Consequently, the statistical control technique is ineffective at adjusting for social self-selection and establishing a causal association between the treatment condition (TC) and the outcome of interest (Y). Second, since the statistical control technique was created for experimental conditions, scholars often ignore the inability to adjust for the observed or unobserved genetic predictors of self-selection (i.e., biological self-selection; Guo and Fraser, 2015). This is a particular problem in SSSMs, where it is assumed that biological factors have limited influence on selection into the treatment condition (Plomin et al., 2013). This limitation is only addressed when moving from SSSMs to quantitative genetic methodologies.

3.1.2. Propensity score matching and generalized propensity score matching

To address the first limitation associated with the statistical control technique, Rosenbaum and Rubin (1983) developed propensity score matching (PSM). PSM was designed as a data reduction matching technique. Prior to the development of PSM, scholars relied on exact matching to establish counterfactual conditions (Guo and Fraser, 2015). Exact matching required scholars to find an individual who experienced the TC, identify key characteristics, and match the individual who experienced the treatment condition to an individual who did not experience the treatment condition. This allowed scholars to establish counterfactual conditions where the key characteristics were held constant at varying levels of exposure to the treatment condition (i.e., the treatment and control case had the same score on the key characteristics;

Rosenbaum and Rubin, 1983). Although beneficial, exact matching is cumbersome when scholars match participants on an increasing number of key characteristics. Furthermore, when the number of key matching characteristics is increased the probability of identifying an exact match decreases. The difficulty to identify exact matches encouraged Rosenbaum and Rubin (1983) to develop a technique where scores on key characteristics could be aggregated, weighted, and matched between individuals at varying levels of exposure to the treatment condition.

As determined by Rosenbaum and Rubin (1983; 1984; 1985), the simplest technique to aggregate and weight scores on key characteristics was through the estimation of a binary logistic regression (BLR).[15] Distinct from OLS regression, BLR analysis allows scholars to estimate the probability of experiencing a dichotomous outcome given an individual's score on an independent variable. These probabilities can be aggregated, where an individuals' probability of experiencing a dichotomous outcome is representative of their observed characteristics. Furthermore, the influence of an observed characteristic on the treatment condition (i.e., the slope estimate) can be employed as a weight. By using the slope estimate as a weighting mechanism, Rosenbaum and Rubin (1983; 1984; 1985) ensured that aggregated probability accurately represented the contribution of each independent variable to the prediction of the dichotomous outcome. While labeled as a propensity score, the score represents an individuals' predicted probability of experiencing a dichotomous outcome given their characteristics. In principle, propensity scores were a well-developed data reduction technique designed to allow scholars to match participants on numerous key characteristics (Rosenbaum and Rubin 1983).

In contemporary scholarship, the estimation of a propensity score with a dichotomous treatment condition remains consistent with Rosenbaum and Rubin's (1983; 1984; 1985) original

[15] During the development of this technique, scholars were primarily interested in dichotomous TCs such as the diagnosis of a disease.

technique. Formula 3.5 represents the estimation of the binary logistic regression. Distinct from our statistical control model the estimation process requires scholars to regress the dichotomous treatment condition (TC) on the key characteristics of interest ($X_{1i} + X_{2i} + X_{ki}$). Succeeding this estimation, individuals' scores for each key characteristic are aggregated, weighted ($b_0 + b_1X_{1i} + b_2X_{2i} + b_kX_{ki}$), and transformed ($1 + \exp(b_0 + b_1X_{1i} + b_2X_{2i} + b_kX_{ki})$) into a propensity score (P; Formula 3.6). The calculations associated with estimating a propensity score for a dichotomous treatment condition makes the technique favorable as compared to exact matching techniques (Apel and Sweeten, 2010; Austin, 2011; Guo and Fraser, 2015). Despite being favorable, the estimation process described above requires scholars to use a dichotomous treatment condition.

[Formula 3.5]

$$\log\left(\frac{TC}{1 - TC}\right) = b_0 + b_1X_{1i} + b_2X_{2i} + b_kX_{ki}$$

[Formula 3.6]

$$P = \sum_{i=1}^{n} [1 + \exp(b_0 + b_1X_{1i} + b_2X_{2i} + b_kX_{ki})]$$

To address this limitation, multiple scholars in a series of papers (Hirano and Imbens, 2004; Imbens, 2000; Joffee and Rosenbaum, 1999) developed techniques to estimate propensity scores for ordinal, multinomial, and continuous treatment conditions. Due to the focus of the current study, only Hirano and Imbens' (2004) generalized propensity score estimator (GPS) for continuous treatment conditions is discussed. Generalized propensity score (GPS) estimation is a two-step process. First, as presented in Formula 3.7, GPS requires scholars to assume a normally distributed treatment condition (Hirano and Imbens, 2004). Commonly this necessitates the

transformation of the treatment condition represented by $g(TC_i)$. Given that the treatment condition is a function of the independent variables (X_i) a simple maximum likelihood regression can be estimated. As it should be noted, N represents the normality and σ^2 represents the variance.

[Formula 3.7]

$$g(TC_i)|X_i \sim N(b_0 + b_1X_{1i} + b_2X_{2i} + b_kX_{ki}, \sigma^2)$$

Second, the estimation of the GPS requires the establishment of covariate balance prior to the estimation of propensity scores (Formula 3.8). Balance refers to the correspondence of covariate distributions between treatment and control groups. Due to the inability to establish precise treatment and control groups, GPS uses a six-step process to estimate the equivalence of key independent variables ($X_{1i} + X_{2i} + X_{ki}$) at varying levels of exposure to the treatment condition. First, GPS necessitates the binning of the treatment distribution into *n* intervals. Second, GPS is computed at each *n* interval. Third, GPS scores are binned using *m* intervals. Fourth, GPS calculates the mean differences on each covariate for individuals that belong to the same binned treatment condition (*n*), and individuals that belong to the same binned GPS score (*m*), but a different binned treatment condition (*n*). Fifth, the mean differences are aggregated and weighted using the number of observations at each GPS interval. Sixth, balance is assessed by using Student's *t* or Bayes factor. As indicated by the six-step process, the propensity score (P) and balance are estimated simultaneously. The conditional expectation of the outcome given the treatment and GPS scores can be estimated with Formula 3.9, whereas the dose-response function to discern treatment effects and 95 percent confident intervals can be estimated with Formula 3.10. These are described in detail in Hirano and Imbens (2004).

[Formula 3.8]

$$\hat{P}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{1}{\sqrt{2\hat{\sigma}^2}}[(\mathrm{g}(\mathrm{TC}_i) - b_0 + b_1 X_{S_{1i}} + b_2 X_{2i} + b_k X_{ki})]^2\right)$$

[Formula 3.9]

$$P = \mathrm{E}[\mathrm{Y}_i|\mathrm{TC}_i, R_i] = \alpha_0 + \alpha_1 \mathrm{TC}_i + \alpha_2 \mathrm{TC}_i^2 + \alpha_3 \mathrm{R}_i + \alpha_4 \mathrm{R}_i^2 + \alpha_5 \mathrm{TC}_i \mathrm{R}_i$$

[Formula 3.10]

$$P = \mathrm{E}[\mathrm{Y}_i] = \frac{1}{N}(\hat{\alpha}_0 + \hat{\alpha}_1 \mathrm{TC} + \hat{\alpha}_2 \mathrm{TC}^2 + \hat{\alpha}_3 \hat{\mathrm{r}}_i(\mathrm{TC}_i, Z_{S_{1i}}) + \hat{\alpha}_4 \hat{\mathrm{r}}(\mathrm{TC}_i, Z_{S_{1i}})^2 + \hat{\alpha}_5 TC\hat{\mathrm{r}}(\mathrm{TC}_i, Z_{S_{1i}}))$$

Succeeding the estimation of the propensity score (P), two categories of matching techniques are generally employed within contemporary literature: greedy matching and optimal matching.[16] Greedy matching refers to matching techniques that create a subsample of cases with the same probability of being assigned to the treatment condition (i.e., propensity score). Included in this category are Mahalanobis metric matching, nearest neighbor matching, nearest neighbor matching in a caliper, and nearest available Mahalanobis metric matching. Due to the overlap in the matching techniques, descriptions of Mahalanobis metric matching and nearest neighbor matching within a caliper can generalize to the other techniques listed above.

Mahalanobis metric matching was invented preceding the creation of propensity scores and matches participants using Formula 3.11. The letters "u" and "v" represent vectors of values for matched variables for treatment (i) and control (j) cases, respectively. C represents the sample covariance matrix. While not present in Formula 3.11, the estimated propensity score (P) is

[16] These matching techniques are only valid with a dichotomous treatment condition. For continuous TC, matching is performed as part of the GPS statistical calculation. Other matching techniques do exist in the literature (e.g., genetic matching; Dehejia and Wahba, 2002; Diamond and Sekhon, 2013). Genetic matching refers to a non-genetically sensitive form of identifying participants with similar propensity scores.

treated as an additional covariate in the calculation and is intended to reduce the average distance between individuals on all covariates. As the number of covariates increases, the average distance between the treatment and control cases increases.

[Formula 3.11]

$$d(i,j) = (\mathrm{u} - \mathrm{v})^T \mathrm{C}^{-1} (\mathrm{u} - \mathrm{v})$$

Nearest neighbor matching within a caliper is represented by Formula 3.12. As indicated by the formula, matches are determined by the difference between the estimated propensity scores for the treatment (P_i) and the control (P_j) cases. Distinct from nearest neighbor matching, nearest neighbor matching within a caliper requires the difference between propensity scores to be less than the specified caliper value (ε). Generally, the specified value should be less than .10 and determined through trial and error (Apel and Sweeten, 2010; Guo and Fraser, 2015). The specified caliper value has a direct impact on the balance of covariates.

[Formula 3.12]

$$C(P_i) = \left|P_i - P_j\right| < \varepsilon$$

Optimal matching is a more cumbersome mathematical process designed to minimize the total sample propensity scores distance (Δ). Formula 3.13 represents optimal matching. In Formula 3.13, "s" represents a matched pair, where "A_s" equals the propensity score for a treatment case, "B_s" equals the propensity score for a control case, and the difference between two cases is represented by "($|A_s|B_s|$)". "ω" represents the weight of the individual difference and "δ" represents the weight of the average distance (A_s, B_s). This calculation should optimize matches by reducing the average sample distance on the propensity score to the absolute minimum. Scholars have encouraged the employment of multiple control cases for each treatment case when conducting greedy or optimal matching (Guo and Fraser, 2015).

[Formula 3.13]

$$\Delta = \sum_{s=1}^{S} \omega(|A_s|B_s|)\delta(A_s, B_s)$$

To ensure the establishment of an empirically valid counterfactual condition, scholars must demonstrate balance between the treatment and control groups on the specified covariates. To reiterate, balance is the observed equivalence of the distribution for key covariates between treatment and control cases. To truly demonstrate balance, it must be determined that the distribution of scores on specified covariates is not divergent between treatment and control cases. Generally, scholars establish balance by estimating the standardized effect size for the mean difference and a Student's *t* (Guo and Fraser, 2015). Notably, if balance cannot be demonstrated using the standardized effect size for the mean difference and a Student's *t*, quartile differences can be estimated. If balance cannot be achieved, scholars often recommend transforming the covariates into semi-continuous or categorical variables (Guo and Fraser, 2015).

Consistent with the counterfactual logic, PSM allows scholars to establish a counterfactual condition where all observed environmental factors are held constant in post-matching analysis. Thus, PSM allows scholars to assume that observed environmental characteristics have limited influence on the outcome of interest because both the treatment and control conditions possess the same distribution of characteristics. In a theoretical sense, holding all the environmental factors constant should produce a counterfactual condition approximating that of a true experiment. PSM identifies a subsample of similar cases (i.e., cases matched on the specified covariates) and estimates post-matching analyses to determine if variation in the treatment condition corresponds to variation in the outcome of interest. These post-matching

analyses are primarily bivariate analyses (e.g., outcome of interest regressed on the TC) of only the subsample identified by PSM. Although cumbersome, scholars have demonstrated the substantial advancements, when considering causality, offered by PSM when compared to the statistical control technique (Guo and Fraser, 2015).

Within the social sciences, PSM is used to examine the association between a variety of treatment conditions and observable outcomes. For instance, scholars have examined the association between gang membership and victimization (DeLisi et al., 2009), college programming and educational attainment (Melguizo, Kienzl, and Alfonso, 2011), special education and learning/behavioral outcomes (Morgan et al., 2010), and a variety of other treatment outcome associations (e.g., Braga, Piehl, and Hureau, 2009; Harding, 2003; Levine and Painter, 2003; Pompoco et al., 2017).[17] Furthermore, evident by the wide variety of scholarship, PMS can be conducted with numerous treatment conditions (McCormick et al., 2013; Mears and Cochran, 2013; Mears, Cochran, and Beaver, 2013; Mears et al., 2012). PSM is highly regarded in the social sciences (Dehejia and Wahba, 2002; Guo and Fraser, 2015), resulting from the ability to establish a counterfactual condition where all observed environmental conditions are held constant at varying levels of exposure to the treatment condition. Nevertheless, the ability of PSM to achieve a true counterfactual condition is limited by the SSSMs used to estimate the propensity scores. Consistent with the discussion regarding statistical controls, propensity scores are generally used to only adjust for the effects of observed environmental influences, never

[17] Considering the treatment conditions of interest – intelligence and educational attainment – a limited quantity of research has been conducted using PSM. More precisely, while intelligence has been widely examined using the statistical control technique, scholars have rarely considered intelligence as a treatment condition. Regarding educational attainment, scholars have demonstrated that educational achievement is associated with future criminal behavior (Backman, 2017; Blomberg, Bales, and Piquero, 2012). Remarkably, the majority of scholarship assessing the effects of educational attainment and future antisocial behavior examines the effects of prison education programs and subsequent recidivism (e.g., Blomberg et al., 2012; Duwe and Clark, 2014; Kim and Clark, 2013). Overall, as demonstrated by various scholars (e.g., Apel and Sweeten, 2010; Dehejia and Wahba, 2002; Guo and Fraser, 2015) the results of post-PSM analyses are primarily regarded as more accurate and conservative than statistical control models.

considering the potential effects of biological self-selection. As demonstrated in the subsequent sections, scholars in quantitative genetics have developed various techniques to examine and adjust for biological self-selection into a treatment condition.

3.2. Quantitative Genetic Methodologies: Conceptual Background[18]

Distinct from standard social science methodologies, quantitative genetic methodologies rely on a theoretical foundation provided by evolution and molecular genetics to adjust for the effects of biological self-selection (Falconer and McKay 1996). To briefly provide the theoretical foundation for behavioral genetics, Darwin (1859) and subsequent scholars (Churchill, 1980; Eldredge, 1985; Fisher, 1919; Mayr and Provine, 1998; Mendel, 1866; Weismann, 1893) provided a theoretical and empirical basis in which the hypotheses associated with phenotypic heredity could be integrated into statistical analyses through various assumptions. Specifically, scholars could limit the confounding effects of genes by implementing statistical analyses on data collected from genetically related portions of the population (Plomin et al., 2013). One portion of the population important to behavior geneticists are monozygotic twins (Plomin et al., 2013). Evident by research on zygosity, MZ twins share 100 percent of their genetic makeup (Plomin et al., 2013; Watson et a., 2008). In statistical analyses, the knowledge that MZ twins share 100 percent of their genetic makeup allows scholars to assume that phenotypic differences between MZ twins can only be accounted for by differences in exposure to environmental conditions (the non shared environment; Barnes, Beaver, and Boutwell, 2011; Barnes et al., 2014; Plomin et al., 2013). As discussed further in sections 3.3.1. and 3.3.2., the conceptual framework outlined above provides the foundation for developing the mathematical assumptions associated with the ACE decomposition model and MZ difference models.

[18] Please see Appendix A for a detailed description of evolution, natural selection, heredity, genetic relatedness, and candidate gene research.

While behavioral genetics offers insight into the genetic factors underlying phenotypic variance within the population, these methodologies cannot identify the specific genes influencing the expression of a phenotype. To address the limitations of behavioral genetics and candidate gene research (see Appendix A), contemporary scholarship has employed various technological advancements to conduct Genome Wide Association Studies (GWAS; Conley and Fletcher, 2017). GWAS have been the foremost technique for examining the association between molecular genetic factors and phenotypic variation (Bush and Moore, 2012; Chabris et al., 2015). Generally, GWAS are performed by drawing information from the entire genome[19] of each participant and isolating the specific genetic markers associated with the phenotype of interest (Bush and Moore, 2012). While the statistical analysis is straightforward, the phenotype of interest is regressed on the whole genome and therefore GWAS needs thousands of participants to satisfy the power requirements associated with an enormous number of independent variables (Conley and Fletcher, 2017).

One of the foremost assessments employing GWAS was the recent study conducted by Lee and colleagues (2018), which examined the association between the whole genome and educational attainment. The research team collected and coded the whole genome of over 1.1 million participants. The findings indicated that 1,271 genome-wide loci[20] were predictive of educational attainment. Furthermore, these findings suggested that educational attainment had a heritability estimate of .11 to .13, which was divergent from behavioral genetic estimates (.52;

[19] In the current context, the entire genome refers to approximately .5 percent of the genetic material that varies between individuals.
[20] A locus (plural: loci) is the position on the specified chromosome that could correspond to coding genes.

Polderman et al., 2015).[21] Although GWAS have provided an important advancement in our understanding of the association between molecular genetics and phenotypic variation, limitations persist. The primary limitations associated with GWAS are the inability to directly estimate the genetic contribution when controlling for environmental conditions and the inability to test gene-environment interactions. As discussed in subsection 3.3.3., scholars generated a conceptual and mathematical formula estimating the aggregated genetic effects associated with phenotypic variation (Conley and Fletcher, 2017).

3.3. Quantitative Genetic Methodologies: Adjusting for Self-Selection

To review, quantitative genetic methodologies are statistical techniques designed to adjust for the potential influence of biological predispositions. When considering self-selection, these quantitative genetic methodologies allow scholars to adjust for the observed and unobserved effects of biological self-selection into a treatment condition. Echoing the sentiments in Chapter 2, two types of self-selection can confound the association between a treatment condition and an outcome of interest: social self-selection and biological self-selection. To adjust for the potential effects of self-selection, scholars must approximate a true counterfactual condition, in which all observed and unobserved biological and environmental factors are constant across the treatment and control conditions. As outlined in section 3.1, while SSSMs adjust for the influence of observed environmental factors on the TC, they often ignores the potential biological factors influencing the probability of exposure to a treatment condition. To

[21] These findings – and others – from GWAS have alluded to the issue of missing heritability. The issue of missing heritability refers the divergence between the heritability estimates in behavioral genetics (i.e., twin and adoption studies) and the heritability estimates produced by molecular genetics. Various scholars have suggested that this divergence could be accounted for by the inability to achieve satisfactory power (e.g., Okbay et al., 2016; Chabris et al., 2015; Conley and Fletcher, 2017), which is evident by the recent reduction in the missing heritability for various phenotypes. Overall, the importance of this is to outline that divergences between quantitative and behavioral genetic methodologies exist (Chabris et al., 2015; Conley and Fletcher, 2017).

address these limitations, the three techniques discussed in the current section are quantitative genetic methodologies designed to adjust for any potential observed and unobserved biological self-selection.

3.3.1. ACE decomposition model

The ACE decomposition model has remained as one of the foremost methodologies employed to quantify the influence of genetics on phenotypic variation (Coyne and Wright, 2014; Jang, Livesley, and Vemon, 1996, Loehlin, 1992; Plomin et al., 2013). One can estimate the quantity of variance in a phenotype attributable to genetic (h^2), shared (c^2), and non shared environmental factors (e^2; Loehlin, 1992; Neale and Cardon, 2013). This can be achieved through the employment of latent variable analysis (Neale and Cardon, 2013).[22] The ACE decomposition model was established through reliance on the principals of Mendelian inheritance (Plomin, Chipuer, and Neiderhiser, 1994). To review, Mendel (1866) proposed that heredity followed three laws: the law of segregation, the law of independent assortment, and the law of dominance. Following Mendel's (1866) guidance, a simple formal accounting for the variance in a phenotype can be produced.

Three terms can be observed within Formula 3.14: V_P, V_G, V_E. Simplistically, V_P refers to the variance in the phenotype, which is equal to the variance of genetic factors (V_G) plus the variance of the environmental factors (V_E). This formula indicates that the variance in a phenotype in a population is equal to the observed variance in genetic factors and the observed variance in environmental factors (Rushton et al., 1986; Neale and Cardon, 2013). As alluded to

[22] A discussion of general latent variable analysis and variance decomposition models can be found in McCutcheon (1987). McCutcheon's (1987) discussion outlines the statistical principals associated with general latent variable analysis and variance decomposition models, which preceded the ACE model.

within previous discussions, the formula presented below does not accurately capture all the elements associated with observed variance.

[Formula 3.14]

$$V_P = V_G + V_E$$

Researchers have identified that the variance in the environmental component (Formula 3.15) can be demarcated into to sub-categories: the shared environment (V_C) and the non shared environment (V_E; Plomin et al., 2013). The shared environment refers to the commonalities experienced by individuals and the non shared environment refers to the distinct environmental conditions experienced by individuals (Scarr, Scarf, and Weinberg, 1980; Hetherington, Reiss, and Plomin, 2013; Plomin et al., 1994). While the shared environment has little influence on unrelated individuals, it is especially important when discussing the equal environments assumption associated with twin research (EEA; Barnes et al., 2014; Derks, Dolan, and Boomsma, 2006). Similarly, the variance associated with genetic factors can be demarcated into three categories: additive genetic effects (V_A), dominant genetic effects (V_D), and epistatic genetic effects (V_I; Formula 3.16; Plomin et al., 2013). Additive genetic effects are when multiple alleles or multiple genes have an aggregated influence on the phenotype and dominant genetic effects are when only one allele or gene influences the phenotype (Carey, 2003; Plomin et al., 2013). Epistatic genetic effects are when the influence of one gene is conditioned upon the presence or absence of another gene (Christian, Kang, and Norton, 1974; Isik, Li, and Frampton, 2003). The full equation is presented in Formula 3.17 and indicates that variance in phenotype (V_P) is equal to the aggregated variance of the shared environment (V_C), the non shared environment (V_E), additive genetic effects (V_A), dominant genetic effects (V_D), epistatic genetic effects (V_I), and any covariance between the terms $2(cov_{x1,x2})$. Consistent with basic statistics, any overlap in the variance of two distinct terms must be adjusted for to ensure that the variance

does not show up twice (Fox, 2016; Henly, 1993). Furthermore, Formula 3.17 theoretically implies that the variance in phenotype can be estimated by observing the terms on the right side of the equation.

[Formula 3.15]

$$V_E = V_C + V_E$$

[Formula 3.16]

$$V_G = V_A + V_D + V_I$$

[Formula 3.17]

$$V_P = V_C + V_E + V_A + V_D + V_I + 2(cov_{v1,v2}) + \cdots$$

Although Formula 3.17 outlines the complex calculations used to estimate the variance in a phenotype, various assumptions can be relied on to simplify the formula. As presented in Formula 3.18, four major assumptions help simplify the mathematical equation. First, as represented by $V_A = V_A + V_D + V_I$, behavioral and molecular geneticists have directly illustrated that additive genetic effects are the principal mechanism influencing the establishment of a complex phenotype (e.g., Davies et al., 2015; Delvin, Daniels, and Roeder, 1997; Hill, Goddar, and Visscher, 2008). Second, the covariance between the genetic and the shared environment is equal to zero ($Cov_{A,C} = 0$). Third, the covariance between the genetic and the non shared environment is equal to zero ($Cov_{A,E} = 0$). Fourth, the covariance between the shared environment and the non shared environment is equal to zero ($Cov_{C,E} = 0$).

[Formula 3.18]

$$V_G = V_A + V_D + V_I$$

$$Cov_{A,C} = 0$$

$$Cov_{A,E} = 0$$

$$Cov_{C,E} = 0$$

As presented in the simplified Formula 3.19, variance in the shared environment (V_C), the non shared environment (V_E), and additive genetic effects (V_A) should account for the variance associated with a phenotype. Observing the genetic and environmental factors associated with the variance in a phenotype is difficult to achieve with even the most advanced technology. To address this issue, scholars approached the analysis from the opposite direction, decomposing the variance in a phenotype into latent constructs through the employment of various theoretical assumptions (Kaplan, 2009; Plomin et al., 2013; Rijsdiik and Sham, 2002).

[Formula 3.19]

$$V_P = V_C + V_E + V_A$$

This is best presented in Figure 3.1, where the variance in the phenotype is measured (represented by the square) and demarcated into various latent constructs, which are represented by the circles. Foundationally, the principals of genetic relatedness underlie the latent statistical processes associated with ACE modeling (Plomin et al., 2013). Scholars developed a method reliant on two subsamples of genetically related individuals to estimate the variance in a phenotype (Boomsma, Busiahn, and Peltonen, 2002; McGue, Osler, and Christensen, 2010; Vandenberg, 1996). Commonly, the two subsamples are MZ and DZ twins, but the ACE decomposition model can be estimated with other genetically related individuals (Plomin et al., 2013).[23] MZ twins are identical twins who share 100 percent of their genetic makeup and DZ twins are fraternal twins who share 50 percent of their genetic makeup. These two subsamples are commonly selected for various reasons, the foremost being that twins have a highly overlapping shared environment (Plomin et al., 2013).

[23] While the ACE decomposition model can be estimated with other genetically related individuals, the validity of the estimates drastically decreases resulting from the inability to accurately assume the existence of the shared environment (Plomin et al., 2013).

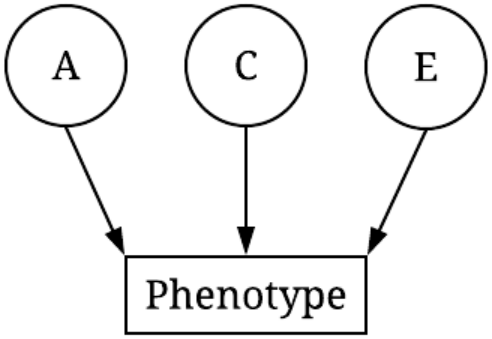


Figure 3.1
Visual Demonstration of Decomposition of Variance Model.[24]

To use twin modeling to estimate an ACE decomposition model, additional assumptions are necessitated. The foremost assumptions made by behavioral geneticists are the equal environments assumption (EEA) and the random mating assumption. Briefly, the EEA argues that the effects of the shared environment on the variance of a phenotype is consistent across MZ and DZ twins and the random mating assumption argues that individuals randomly mate with other individuals (Rijsdiik and Sham, 2002; Rowe, 1983). Although discussions have appeared in prior literature (Barnes et al., 2014), strong evidence suggests that violating these assumptions has a limited effect on the estimates produced (Bailey, Dunne, and Martin, 2000; Barnes et al., 2014; Littvay, 2012; Rijsdiik and Sham, 2002; Felson, 2014). With these additional assumptions scholars constructed calculations to estimate the amount of variance in the phenotype attributable to genetic and environmental factors.

Formula 3.20 and Formula 3.21 present the mathematical calculation used to estimate the correlation in phenotype for MZ and DZ twins. Notably, only the genetic (h^2) and shared

[24] In the visual demonstration A represents the additive genetic effects, C represents the shared environmental effects, and E represents the non shared environmental effects.

environment (c^2) are officially estimated within the formulas (all additional variance and error is attributable to the non shared environment, e^2, in the final calculations; Plomin et al., 2013). Furthermore, the genetic correlation between individuals in the twin pairs are represented by the $r(h^2)$, where MZ twins share 100 percent of their genes ($1(h^2)$) and DZ twins share 50 percent of their genetic material ($.5(h^2)$).

[Formula 3.20]

$$r_{MZ} = r(h^2) + c^2$$

$$r_{MZ} = 1(h^2) + c^2$$

[Formula 3.21]

$$r_{dZ} = r(h^2) + c^2$$

$$r_{dZ} = .5(h^2) + c^2$$

By combining these formulas, the variance in the phenotype can be attributed to the variance in genetic and shared environmental factors (h^2 and c^2; Plomin et al., 2013). Formula 3.22 presents the calculation for estimating the variance in a phenotype associated with the variance in genetic factors (h^2). h^2 can be estimated by subtracting the correlation in a phenotype between DZ twins (r_{DZ}) from the correlation in a phenotype for MZ twins (r_{MZ}) and multiplying that value by two. The multiplication of the final value by two accounts for the known difference in genetic relatedness of MZ and DZ twins. MZ twins are two times more genetically related than DZ twins (i.e., 100 percent genetic relatedness for MZ twins and 50 percent genetic relatedness for DZ twins). Formula 3.23 demonstrates the calculation for estimating the variance in a phenotype associated with the variance in shared environmental factors (c^2). c^2 is estimated by multiplying the correlation in a phenotype between DZ twins

(r_{DZ}) by two and subtracting the correlation in a phenotype for MZ twins (r_{MZ}) from that value. Formula 3.24 presents the calculation for estimating the variance in a phenotype associated with the variance in non shared environmental factors (e^2), which is calculated as 1 minus the variance in genetic factors (h^2) plus the variance in shared environmental factors (c^2). This calculation allows any statistical or measurement error to be captured in the estimate for the non shared environment (Plomin et al., 2013).

[Formula 3.22]

$$h^2 = 2(r_{MZ} - r_{DZ})$$

[Formula 3.23]

$$c^2 = 2(r_{DZ}) - r_{MZ}$$

[Formula 3.24]

$$e^2 = 1 - (h^2 + c^2)$$

The ACE decomposition model is the foremost technique for identifying and estimating the influence of genetic factors and environmental factors on a phenotype or a treatment condition.[25] In regards to biological self-selection, the ACE decomposition model can be estimated in combination with SSSMs to observe the disjunction between the variance explained by environmental factors between the two methodologies. Any substantive difference between the two terms (i.e., the $e^2 + c^2$ value for the ACE decomposition model and the r^2 value for the SSSMs) would be suggestive of potential biological self-selection into a treatment condition and

[25] While the term treatment condition alludes to a predetermined treatment, in the current context any variable influencing phenotypic differences can be considered a treatment condition (e.g., intelligence, educational attainment, peer delinquency).

the differences between SSSMs and genetically sensitive methodologies, which will be discussed in subsequent subsections (Pinker, 2003). A disjunction commonly exists between these terms due to the estimation process of SSSMs, which is suggestive of the inability to account for genetic factors within the model. Generally, the estimation of an r^2 value for the SSSMs captures all three components associated with a treatment condition ($h^2 + e^2 + c^2$; Pinker, 2003).

Empirical assessments employing the ACE decomposition model have directly demonstrated the disjunction between quantitative genetics and sociological research using SSSMs (e.g., Branigan, McCallum, and Freese, 2013; Delvin et al., 1997; Wright and Beaver, 2005). One of the largest meta-analyses ever conducted further solidified the validity of the hypothesis that the majority of complex phenotypes are subjected to some genetic influence (Polderman et al., 2015). The majority of 17,804 traits examined had average heritability estimates that centered around .49. Furthermore, the average heritability estimate suggests that the employment of SSSMs might overestimate the effects of a treatment condition on antisocial behavior. Evidence suggests that the counterfactual established during standard social science analyses might not represent reality, where biological and social self-selection influences the probability of exposure to a treatment condition (Beaver, 2009; Plomin et al., 2013). Overall, the ACE decomposition model allows scholars to partially observe the degree to which biological self-selection could potentially alter the estimates associated with SSSMs (Pinker, 2003).

3.3.2. MZ difference model

Divergent from the ACE decomposition model, the MZ difference model is a common behavioral genetic technique designed to adjust estimates for the unobserved effects of biological

self-selection (Pike et al., 1996; Rovine, 1994; Rowe and Plomin, 1981).[26] In the current context, biological self-selection refers to the process in which an individuals' probability of exposure to a treatment condition is influenced by biological predisposition. As illustrated mathematically below, the MZ difference model adjusts for the unobserved effects of biological self-selection through reliance on Mendelian inheritance and genetic relatedness (Burt, McGue, and Iacono, 2009; Pike et al., 1996). Unlike other kinship pairs, MZ twins share 100 percent of their genetic material, which results from the meiosis of a single zygote. Since it can be assumed that MZ twins share 100 percent of their genetic material, all phenotypic differences between MZ twins should result from differences in environmental factors (Asbury et al., 2003; Caspi et al., 2004). This can be represented by the Formula 3.25 and Formula 3.26.[27] As demonstrated in Formula 3.25, any phenotypic differences between MZ twins (Δp_{mz}) is equal to the genetic differences (Δh_{mz}) and the environmental differences ($\Delta c_{mz} + \Delta e_{mz}$). Accordingly, since MZ twins share 100 percent of their genetic material, we can safely assume that the genetic difference between our MZ twins is equal to zero ($\Delta h^2{}_{mz} = 0$; represented by Formula 3.26; Pike et al., 1996; Plomin 2011; Vitaro, Bredgen, and Arseneault, 2009). With this assumption in mind, the phenotypic differences between MZ twins (Δp_{mz}) is equal to the shared (Δc_{mz}) and the non shared (Δe_{mz}) environmental differences. In addition to the assumption that the genetic difference between our MZ twins is equal to zero ($\Delta h_{mz} = 0$), scholars generally assume that any phenotypic differences (Δp_{mz}) could be safely accounted for by differences in the non shared environment (Δe_{mz}; Neale and Cardon, 2013; Vitaro et al., 2009). Notably, contained within the Δe_{mz}, are observed differences in the treatment condition (ΔTC_{mz}) and independent variables of interest (ΔX_{mz}).

[26] MZ difference model also adjust estimates for the unobserved effects of shared environment.
[27] In the current context, Δ represents the difference between MZ twins for the specified term.

[Formula 3.25]

$$\Delta p_{mz} = \Delta h_{mz} + \Delta c_{mz} + \Delta e_{mz}$$

[Formula 3.26]

$$\Delta h_{mz} = 0$$

The divergence between MZ twins regarding a phenotype and the non shared environment can be calculated using the simple formulas presented below. As indicated by Formula 3.27, the discordance between MZ twins on a phenotype ($\Delta p_{mz} = p_{t1} - p_{t2}$), the treatment condition ($\Delta TC_{mz} = TC_{t1} - TC_{t2}$), or observed independent variables ($\Delta X_{mz} = X_{t1} - X_{t2}$) can be calculated by simply subtracting the score for twin two from the score for twin one.

[Formula 3.27]

$$\Delta p_{mz} = p_{t1} - p_{t2}$$

To ensure unbiased estimates scholars often randomly assign the MZ twins the label of twin one or twin two (Nedelec, Park, Silver, 2016; Viding et al., 2009). The logic of MZ difference scores has led to the development of a variety of statistical estimation techniques adjusting for the influence of unobserved genetic confounders (Rovin, 1994). In addition to MZ discordance estimation provided in Formula 3.27 scholars can estimate MZ discordance through a residual gain score, a relative versus absolute difference score, a variance score, a fixed effects model, and various other techniques (Fujiwara and Kawachi, 2008; Rovin, 1994).

As indicated below, the raw MZ difference scores can be introduced into basic regression models and adjust for the effect of the unobserved genetic confounders. Formula 3.28 presents the simple formula for a bivariate ordinary least squares (OLS) regression model. This formula

illustrates the assumption that scores on the dependent variable (Y) are a function a vector of coefficients (b_1), which is influenced by a $m\ x\ i$ matrix of independent variables (X_i).[28] As indicated by Formula 3.29, scholars can merely replace scores on the dependent variable (Y_i) and the independent variables (X_i) with MZ difference scores to adjust for the effects of unobserved genetic and shared environment confounders. The dependent variable (Y_i) is now represented by MZ difference scores on the specified phenotype (Δp_{mz}) and the independent variables (X_i) is now represented by MZ difference scores on the treatment condition (ΔTC_{1mz}) and the specified independent variables (ΔX_{2mz}; Rovine, 1994). Consistent with the assumptions presented above, the genetic influence on the phenotype is assumed to be zero because $\Delta h_{mz} = 0$ (Neale and Cardon, 2013). Furthermore, it can be assumed that the shared environmental difference has zero influence on the estimations because $\Delta c_{mz} = 0$. The same logic can be applied to binary logistic regression.

[Formula 3.28]

$$Y = b_0 + b_1 X_{1i}$$

[Formula 3.29]

$$\Delta p_{mz} = b_0 + b_1 \Delta TC_{1mz} + b_2 \Delta X_{2mz}$$

As noted by scholars, the MZ difference score technique is the gold standard to adjust for the effects of biological self-selection in a statistical model (Vitaro et al., 2009). As demonstrated above, the MZ difference score technique allows scholars to assume that genetic predispositions are held constant within the statistical model. The MZ difference score establishes a natural

[28] m represents the number of columns (i.e., variables) within the dataset and i represents the number of rows (i.e., individuals) within the dataset.

counterfactual condition where only the differences on environmental conditions are examined for two genetically identical individuals. This means that any observed differences on an outcome of interest can only be attributable to the treatment condition, observed environmental conditions, and unobserved environmental conditions (and measurement error). Regarding biological self-selection, the beneficial nature of the MZ difference score technique is evident by the conservative statistical estimates produced within various statistical analyses (Turkheimer and Waldron, 2000).

MZ difference scores have been employed to empirically assess the association between a treatment condition and an outcome of interest when holding unobserved shared genetic and shared environmental factors constant. Generally, when compared to standard social science methodologies, MZ difference scores produce more conservative estimates (e.g., Asbury et al., 2003; Beaver, 2008; Beaver, Vaugh, and DeLisi, 2013; Caspi et al., 2004; Pike et al., 1996). Evident by recent scholarship (e.g., Burt et al., 2009; Nedelec, Park and Silver, 2016; Nedelec, Richardson, and Silver, 2017; Nedelec et al., 2012), when adjusting for unobserved genetic factors scholars tend to demonstrate null associations between various treatment conditions and observed outcomes. These null associations generally support the suggestion that genetically sensitive methodologies approximate a true counterfactual condition more closely than standard social science methodologies (Plomin et al., 2013). The demonstration of a null association in a genetically sensitive model suggests that the treatment condition has limited influence on the observed outcome when biological predispositions are held constant (Pike et al., 1996). Furthermore, these findings suggest that a superior counterfactual can be established by accounting for the effects of biological self-selection.

Although MZ difference method is one of the preeminent methodologies adjusting for genetic confounders in behavioral genetics, limitations associated with the technique persist

(Vitaro et al., 2009). The MZ difference score technique relies on a sample of MZ twins, which make up a limited portion of the general population. Critics (e.g., see Boardman and Fletcher, 2015) have suggested the limited number of MZ twins within the general population could reduce the representativeness of statistical associations to non-twins (Barnes and Boutwell, 2012). Furthermore, even though these critiques are unsubstantiated (see Barnes et al., 2014), a well-known limitation of the MZ difference score technique is the assumption that twin modeling techniques are underpowered . Specifically, sufficient variation in MZ difference scores and a sufficient sample size must be achieved prior to observing a statistically significant association between a phenotype (Δp_{mzi}) and a treatment condition (ΔTC_{mzi}; Neale and Cardon, 2013). With recent advancements in technology and the difficulties associated with MZ difference scores, scholars in behavioral genetics and molecular genetics have implemented various methodologies to estimate the effects of observed genetic factors on a treatment condition (Purcell et al., 2009). One innovative technique has allowed scholars to estimate the probability of exposure to a treatment condition with observed genetic variants (Evans, Visscher, and Wray, 2009). As demonstrated below, polygenic risk scores allow scholars to adjust for the influence of observed genetic variants by estimating and controlling for the effects of genetic loci (Evans et al., 2013).

3.3.3. Polygenic risk scores

Similar to propensity scores, the polygenic risk score approach is a data reduction technique designed to aggregate the effects of multiple genetic alleles in an effort to examine or control for the effects of multiple variables within a single model. Briefly, exposure to a treatment condition can result from two distinct genetic mechanisms of biological self-selection: polygenic effects or dominant genetic effects (Purcell et al., 2009). In the current context, a

dominant genetic effect refers to the hypothesis that a single genetic locus influences the probability of exposure to the treatment condition. Furthermore, scholarship has consistently demonstrated that it can be safely assumed that the probability of exposure to treatment condition is generally a function of aggregated effects (Bush and Moore, 2012; Chabris et al., 2015). Notably, early genome wide association studies (GWAS) commonly isolated individual locus, creating an examination of dominant genetic, rather than polygenic effects (Hirschhorn and Daly, 2005; McCarthy et al., 2008). In other words, GWAS examine the association between a single locus and a treatment condition, rather than the aggregate effects of multiple loci (Price et al., 2006). To address this issue, scholars created polygenic risk scores, which account for the aggregated genetic effects of the whole genome on the observed variation in the treatment condition (Purcell et al., 2009).[29] The estimation of a polygenic risk score is a three-step process requiring multiple GWA datasets.[30]

First, polygenic risk scores necessitate the estimation of a genome wide association on an independent dataset (Purcell et al., 2009). Specifically, if the GWA model and polygenic scores are estimated on the same sample, the polygenic scores will artificially amplify the estimated effects of specific locus on the treatment condition (Evans et al., 2009). Formula 3.30 demonstrates the estimation of a GWA model for a single treatment condition (TC). As indicated by the formula, the treatment condition (TC) is a function of the vector of estimates (b_i), a m x n matrix of genetic markers (G_i) – otherwise known as loci – and error (e; Dudbridge, 2013). Captured within the error term is the influence of the non shared environment. Not evident by the formula above, GWA analysis uses an iterative estimation process, where the treatment condition

[29] While simplistically stated above, the technique requires scholars to specify the loci of interest. While some scholars have generated polygenic risk scores from only loci significantly associated with the phenotype of interest, evidence suggests that the inclusion of the whole genome provides a superior statistical control for the observed genetic effects (Dudbridge, 2013).
[30] While not common, polygenic scores can be estimated with a single dataset.

(TC) is regressed on genetic markers (G_i) and mutations at each locus (m). The estimates for each locus (m) represent the association between the genetic markers (G_i) and the treatment condition (TC), given the variation observed in the dataset (Purcell et al., 2009). Due to the iterative process, scholars are required to achieve significance levels well beyond the standard social science study (i.e., $\alpha < .05 \times 10^{-5}$; Evans et al., 2013). An additional term ($\sum_{i=1}^{m} b_{i2}G_i + E_2$) can be introduced into the model to estimate the genetic correspondence between two phenotypic traits or a treatment condition and an outcome of interest (Dudbridge, 2013).

[Formula 3.30]

$$TC = \sum_{i=1}^{m} b_i G_i + e$$

Second, succeeding the estimation of a GWA model on an independent sample, the polygenic risk scores associated with a treatment condition can be estimated (Purcell et al., 2009). Using the estimates ($\hat{b}_i$) generated from the initial GWA model, scholars simply sum the weighted coefficients associated with each genetic marker (i.e., loci; Dudbridge, 2013). In other words, the vector of estimates from the independent sample ($\hat{b}_i$) GWA model is multiplied by a m x n matrix of genetic alleles (G_i) and summed ($\sum_{i=1}^{m} \hat{b}_i G_i$; Formula 3.31). When the effects are aggregated, a raw score representing the amount of variance in the treatment condition accounted for by observed genetic factors is created (Purcell et al., 2009). Again, Formula 3.31 represents the formula for estimating polygenic scores for a single trait, and the complexity of the estimation is greatly enhanced by the introduction of additional terms (Dudbridge, 2013).

[Formula 3.31]

$$\hat{S} = \sum_{i=1}^{m} \hat{b}_i G_i$$

An alternative approach is to use the unweighted polygenic score. As indicated by Formula 3.32, the unweighted polygenic risk score is the aggregate effects of genetic markers (i.e., loci) without the presumption of a specified effect size or direction of association (Dudbridge, 2013). $sgn(\hat{b}_i)$ refers to the absolute value for the slope estimate divided by the true value for the slope estimate. The unweighted polygenic risk score technique can be considered more robust against various estimation errors (e.g., limited sample size, population heterogeneity), but the limitations outweigh the benefits (Palla and Dudbridge, 2015). Specifically, the unweighted polygenic risk scores have limited guidance from prior scholarship and could potentially overestimate the variance in the treatment condition predicted by each locus (Palla and Dudbridge, 2015). Overall, scholars are encouraged to compare the two estimation techniques, but the weighted technique is generally preferred (Dudbridge, 2013).

[Formula 3.32]

$$\hat{S} = \sum_{i=1}^{m} sgn(\hat{b}_i) G_i$$

$$sgn(\hat{b}_{i1}) = \frac{|\hat{b}_i|}{\hat{b}_i}$$

Third, similar to MZ discordance scores polygenic risk scores are generally introduced into a statistical model as a predictor or statistical control. As indicated by Formula 3.33 and

Formula 3.34, the polygenic risk score functions as a standard independent variable, where Y is regressed on $b_1\hat{S}_{1i}$. When introduced as a statistical control the polygenic risk score is calculated using the two-trait, rather than the one-trait, GWA estimation technique (Dudbridge, 2013).[31] The two-trait estimation technique allows scholars to estimate the genetic factors (i.e., shared polygenic factors) potentially accounting for the covariance between the dependent and independent variables of interest. As indicated by scholars (e.g., Middeldorp et al., 2011; Okbay et al., 2016; Plomin, 2013; Power et al., 2015), the introduction of polygenic scores into a statistical model should adjust for the observed genetic predispositions influencing the covariance between two environmental variables.

[Formula 3.33]

$$Y = b_0 + b_1\hat{S}_{1i} + b_2X_{2i}$$

[Formula 3.34]

$$\log\left(\frac{Y}{1-Y}\right) = b_0 + b_1\hat{S}_{1i} + b_2X_{2i}$$

two statistical expectations are associated with polygenic scores. First, as outlined above, polygenic scores mathematically estimate exposure to the treatment condition given the observed genetic markers (b_iG_i; Evans et al., 2013). Predictably, this estimation technique removes any variation accounted for by the non shared environment or error (e; Evans et al., 2013). Second, since the polygenic scores are a function of a GWA model, they are often subjected to the effects of missing heritability (Plomin, 2013). The variance in a treatment condition explained by a polygenic score is generally lower than heritability estimates produced by ACE decomposition

[31] Please note that the two-trait estimation technique is only used when scholars want to control for the genetic covariance between two variables. Generally, it is acceptable to introduce a polygenic score estimated with the one-trait estimation technique and other environmental variables when predicting outcomes of interest (Dudbridge, 2013).

models (Plomin, 2013). To outline this effect, if an ACE decomposition model estimates an h^2 = .60, but a GWAS estimates an r^2 = .10, the polygenic scores will only predict 10 percent of the variance in the treatment condition of interest (Plomin, 2013). As sample sizes increase, the amount of missing heritability in polygenic scores should approach (Dudbridge, 2013; Chatterjee et al., 2013; Dudbridge, 2013). As noted by various scholars (e.g., Dudbridge, 2013; Plomin, 2013; Purcell et al., 2009), variance explained by polygenic scores is a function of statistical power of the sample. Samples with limited statistical power will explain less variance in the treatment condition than samples with adequate statistical power, which is generally determined through a power analysis (Cohen, 1992).

Due to their recent development, polygenic risk scores have generally been implemented as a covariate in various regression analyses (Domingue et al., 2016; Hamshere et al., 2011). Polygenic risk scores have either been the primary independent variable or a statistical control when assessing the predictors of an outcome of interest (Domingue et al., 2016; Dominque et al., 2018; Hamshere et al., 2011; Lori et al., 2017; Trotta et al., 2017). When regarding biological self-selection, the latter technique is more common. Specifically, an outcome of interest is regressed on a treatment condition while controlling for the effects of the polygenic risk score. Generally, when employed as a statistical control, more conservative estimates of the association between a treatment condition and an observed outcome are produced (Agerbo et al., 2015; Derks et al., 2012; French et al., 2015; Jansen et al., 2018). Due to data limitations scholars have not employed polygenic risk scores as statistical controls to examine the causal effects of intelligence or educational attainment on future antisocial behavior.

Remarkably, while polygenic risk scores represent the most recent advancements in genetically sensitive methodologies, contemporary scholarship employing polygenic risk scores as a covariate in a statistical control model cannot adjust for biological self-selection.

Specifically, this model cannot establish a counterfactual condition where the polygenic score remains constant at varying levels of exposure to the treatment condition. To address this limitation, a counterfactual condition – that adjusts for biological self-selection – can be established by matching participants on the polygenic risk score. The polygenic risk score must remain balanced at varying levels of exposure to the treatment condition.[32] In theory, post-matching analyses should produce estimates similar to that of an MZ discordance model because the influence of biological predispositions on the outcome of interest should be equal to zero. Nevertheless, matching participants on polygenic risk scores only adjusts for the influence of observed genetic factors on the treatment condition. To establish a more precise counterfactual condition, theory dictates that all biological and environmental factors should be held constant at varying levels of exposure to the treatment condition (Lewis, 1979). Although difficult to achieve, genetically adjusted propensity score matching (*GAPSM*) could approximate the true counterfactual more accurately than the methodologies reviewed by matching participants on the observed genetic and observed environmental predictors of a treatment condition.

3.4. Genetically Adjusted Propensity Scores (*GAPS*)

To review, the primary focus of the current chapter was to establish a comprehensive understanding of the statistical methodologies employed to adjust for the effects of self-selection. Self-selection, the process in which an individual's predispositions increase or decrease the probability of exposure to a treatment condition and an outcome of interest, is the primary factor limiting the ability of social scientists during assessments of causality. While the term "treatment condition" alludes to a predetermined treatment, in the current context any variable influencing phenotypic differences can be considered a treatment condition (e.g., intelligence, educational

[32] Balance, or common support, the necessity to maintain an equivalent distribution of scores (on a variable of interest) at different levels of the treatment condition.

attainment, peer delinquency). Self-selection limits our ability to establish causality by introducing non-random error into the statistical models, which in-turn increases the probability of observing a spurious association between the treatment condition and the outcome of interest. Other than random assignment (i.e., a true experiment) scholars can eliminate the effects of self-selection by holding the observed and unobserved genetic and environmental conditions constant. Stated differently, a causal association between the treatment condition and the outcome of interest can only be achieved by establishing the perfect counterfactual condition. Predictably, this is extremely difficult to accomplish in the social sciences.

The purpose of the current dissertation is to advance beyond contemporary statistical techniques and develop the capacity to construct estimates nearly uninfluenced by the effects of observed confounding variables. As noted in prior sections, only a limited body of scholarship in the social sciences can adjust for the effects of self-selection through the implementation of random assignment (Shadish et al., 2002; Singelton and Straits, 2010), which is an unavoidable consequence of the complexity social science research questions. For all other scholarship, I suggest *genetically adjusted propensity score matching* (*GAPSM*), which is a technique designed to obtain estimates virtually uninfluenced by the effects of observed genetic and observed environmental factors on self-selection. The superiority of this technique is highly dependent upon the observed environmental conditions.

If only demographic characteristics (e.g., race, gender, and age) and shared environmental variables (e.g., parent income, parent education, and childhood living situation) are available, then estimates similar to those of MZ difference scores should be produced. If the available data possesses additional environmental conditions (e.g., peer groups, previous behavioral measures, dietary habits, etc.) then *GAPSM* should provide estimates more conservative than MZ difference scores. Consistent with other statisticians (e.g., Fox, 2016),

theory and available data should heavily influence the employment of a statistical technique to examine a specified research question. *GAPSM* estimates and adjusts for the effects of observed genetic and environmental factors through the aggregation of propensity scores and polygenic risk scores into a single estimate (referred to as a *genetically adjusted propensity score)*. As indicated in previous subsections, propensity scores adjust for the observed environmental predictors and polygenic risk scores adjust for the observed genetic predictors. The combination of propensity scores and polygenic risk scores should allow scholars to approach causality beyond the techniques identified above by controlling for both observed genetic and observed environmental predictors of the treatment condition.

3.4.1. Estimating genetically adjusted propensity scores (GAPS)

As outlined below, the primary distinction between propensity scores and *genetically adjusted propensity scores* (*GAPS*) is the inclusion of a genetic risk variable (i.e., polygenic risk score) during the estimation of the propensity score. The estimation process for a GAPS can be outlined in two steps. First, scholars are required to estimate a polygenic risk score using the method described in subsection 3.3.3. In the current context, the polygenic risk score should be estimated using the one-trait model (e.g., the treatment condition estimation technique), where the treatment condition (TC) is regressed on the observed genetic alleles as presented in Formula 3.35.

[Formula 3.35]

$$TC, = b'G + E = (\sum_{i=1}^{m} b_{i1}G_i + E_1)$$

A polygenic risk score is estimated using coefficients ($\hat{b}_1$) from an independent GWA, where the treatment condition (TC) was regressed on the observed genetic alleles and an m x n

matrix of genetic loci (G_i; Formula 3.36). Every cell of the genetic matrix (G_{1i}) represents the alleles associated with an individual at a specified genetic locus. Overall, the polygenic risk score represents an individual's observed genetic susceptibility to exposure to the treatment condition. Scholars generally standardize the raw scores on a *z* distribution, which is presented in Formula 3.37. The *z* standardization of the polygenic score is equal to the raw score of an individual (S_i) minus the mean of the distribution ($\bar{S}$) divided by the standard deviation of the distribution (s).

[Formula 3.36]

$$\hat{S} = \sum_{i=1}^{m} \hat{b}_1 G_i$$

[Formula 3.37]

$$Z_{S_i} = \frac{S_i - \bar{S}}{s}$$

Second, two methods exist for combining a polygenic risk score with a propensity score. The first method estimates the propensity score with the polygenic score as a predictor of the treatment condition. The polygenic score and any observed environmental covariates would be introduced into a regression model. This technique requires the standardization of the raw polygenic scores on a *z* distribution and produces GAPS estimates for the entire sample. Although only two estimation techniques (one for a dichotomous treatment condition and one for a continuous treatment condition) are reviewed below, polygenic scores can be introduced into any statistical method designed to estimate propensity scores.

The dependent variable in the regression analysis represents the treatment condition (TC) and the symbol ζ represents the GAPS. Consistent with contemporary knowledge, binary logistic

regression is the primary technique used for estimating the predicted probabilities associated with dichotomous treatment conditions. As outlined by Formula 3.38, the log odds of the treatment condition ($\log\left(\frac{TC}{1-TC}\right)$) is regressed on the standardized polygenic score ($Z_{S_{1i}}$) and any observed environmental conditions ($X_{2i} - X_{ki}$). Succeeding the estimation, the odds associated with standardized polygenic scores ($Z_{S_{1i}}$) and observed environmental conditions ($X_{2i} - X_{ki}$) are summed and exponentiated to estimate GAPS (Formula 3.39). This method allows the newly estimated GAPS to approximate the influence of observed genetic and observed environmental factors on the treatment condition. A similar estimation technique is employed with continuous outcome variables.

[Formula 3.38]

$$\log\left(\frac{TC}{1-TC}\right) = b_0 + b_1 Z_{S_{1i}} + b_2 X_{2i} + b_k X_{ki}$$

[Formula 3.39]

$$\zeta = \sum_{i=1}^{n} [1 + \exp(b_0 + b_1 Z_{S_{1i}} + b_2 X_{2i} + b_k X_{ki})]$$

To review, Formula 3.40 and Formula 3.41 present the estimation of propensity scores for continuous treatment conditions. The only difference in the estimation of a propensity score for a continuous treatment condition and a GAPS for a continuous treatment condition is the replacement of one independent variable (X_{1i}) with the standardized polygenic score ($Z_{S_{1i}}$) in formulas 3.40 and 3.41. As indicated by the formulas, the polygenic score is introduced as an independent variable and does not require the further complication of advanced statistical models. Furthermore, the GAPS produced by the two estimation techniques will structurally

mimic that of a propensity scores (i.e., a normal distribution of effects ranging between .00 and 1.00, where more extreme values are observed less often), but adjust for both the observed genetic and environmental factors influencing the treatment condition. To reiterate, scholars can employ ordinal regression models, multi-level models, and longitudinal models to estimate a GAPS. The only requirement is the introduction of a polygenic score into the regression model.

[Formula 3.40]

$$\mathrm{g}(\mathrm{TC}_i)|\mathrm{X}_i \sim N(b_0 + b_1 Z_{S_{1i}} + b_2 X_{2i} + b_k X_{ki}, \sigma^2)$$

[Formula 3.41]

$$\hat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} exp\left(-\frac{1}{\sqrt{2\hat{\sigma}^2}}(\mathrm{g}(\mathrm{TC}_i) - b_0 + b_1 Z_{S_{1i}} + b_2 X_{2i} + b_k X_{ki})^2\right)$$

The second method, which requires the normalization of the raw polygenic scores (indicated by Formula 3.42), involves the independent estimation of the polygenic score and the propensity score. In the current context, independent estimation refers to the process where the propensity scores are estimated without the polygenic score as an independent variable. Scholars can independently estimate the propensity scores using binary logistic regression models, negative binomial regression model, ordinal regression models, and several alternative probability-based estimation techniques. The independent estimation method requires the treatment condition to be dichotomous or categorical, given the simultaneous estimation of balance and matching associated with generalized propensity scores.

Formula 3.42 is the method employed to normalize the raw polygenic score between 0 and 1. N_S indicates the normalized polygenic score, while S_i and $\hat{S}$ represent the score of an individual and the entire distribution of the polygenic score respectively. The normalization of

the polygenic score is required to ensure that the independently estimated polygenic risk score and propensity score are restricted to the same scale.

[Formula 3.42]

$$N_S = \frac{S_i - \min(\hat{S})}{\max(\hat{S}) - \min(\hat{S})}$$

Post-normalization, the polygenic score and propensity score can be combined using the formula presented below. In Formula 3.43, the GAPS is represented by the symbol ζ, which is equal to the addition of the weighted normalized polygenic score ($N_{\hat{S}} * \hat{S}_{s^2}$) and the weighted propensity score ($P * \left(1 - \hat{S}_{s^2}\right)$). The polygenic score and propensity score are weighted by the variance in the treatment condition explained ($\hat{S}_{s^2}$) and not explained ($1 - \hat{S}_{s^2}$) by genetic factors, respectively. The weights permit scholars to theoretically and empirically vary the influence of genetic and environmental factors during the establishment of the GAPS. Furthermore, the weights ensure that the GAPS structurally mimics the normalized polygenic risk score ($N_{\hat{S}}$) and the propensity score (P). Due to the independent estimation of the polygenic score and propensity score, all covariance between the two terms must be removed $\left(cov_{(N_{\hat{S}}),(P)}\right)$.

[Formula 3.43]

$$\zeta = (N_{\hat{S}} * \hat{S}_{s^2}) + \left(P * \left(1 - \hat{S}_{s^2}\right)\right) - \left(cov_{(N_{\hat{S}}),(P)}\right)$$

The calculation above produces GAPS, which approximates the influence of the observed genetic factors and the observed environmental factors on the treatment condition. While similar to the propensity score estimated with the polygenic risk score as a predictor, an important distinction between the two calculations does exist. The independent estimation method ensures

that an empirically valid quantity of variance in the treatment condition is attributed to environmental factors, despite the inability to quantify every environmental predictor. To state differently, the second method allows scholars to ensure that the environment is represented in the estimation of the GAPS.

3.4.2. Assumptions of genetically adjusted propensity scores (GAPS)

There are three assumptions associated with GAPS. First, the GAPS method assumes that the polygenic risk score is not associated with any environmental variance ($\zeta(c + e) = 0$). Reliably, this assumption is satisfied with the mathematical estimation of the polygenic risk score (Dudbridge, 2013). Specifically, since the variance in the treatment condition associated with the environment does not contribute to the estimation of a polygenic risk score it can be safely assumed that the polygenic risk score is not associated with any environmental variance (Dudbridge, 2013). Second, both methods described above assume that the polygenic risk score was estimated with the same treatment condition as the propensity score ($\zeta(y) = P(y)$). This assumption should be closely monitored, as the scale of the treatment condition cannot vary between the estimation of the polygenic risk score and the propensity score. Third, regarding the second method, the independent estimation of the polygenic score and the propensity score, it is assumed that the propensity score contains some variance attributed to genetic factors ($P(h^2) \neq 0$). Without assuming that the propensity score contains some variance attributed to genetic factors, the method would not necessitate the weights or the removal of covariance between the two terms. These three assumptions should be examined prior to the estimation of GAPS. It should be noted that additional assumptions need to be satisfied when introducing the polygenic score into any regression model (e.g., linearity; Fox, 2016).

3.4.3. Genetically adjusted propensity score matching (GAPSM)

Consistent with contemporary scholarship (e.g., Guo and Fraser, 2015), the superior post-estimation technique for GAPS is matching. While other post-estimation techniques do exist (e.g., the introduction of GAPS into a regression model as a control variable), they are comparable to the estimation of a regression model with both genetic and environmental controls.[33] If those estimations are desired, MZ difference scores should provide estimates consistent with GAPS. If not, *genetically adjusted propensity score matching* (*GAPSM*) should provide a superior technique adjusting for the effects of the observed genetic predictors and the observed environmental predictors of the treatment condition.[34]

To reiterate, matching in any form is a method designed to reduce a larger sample into a subsample of similar individuals. In the current context, participants are matched on a score (i.e., GAPS) approximating the genetic and environmental predictors of the treatment condition. Succeeding matching, the subsample of individuals should be similar on all the observed genetic and environmental predictors but vary when considering the treatment condition (Morgan and Winship, 2015). The variation in treatment condition contained within the subsample of individuals is dependent upon the level of measurement associated with the treatment condition. Once the subsample is generated, the individuals are compared to observe any phenotypic differences. Theoretically, only variance in the exposure to the treatment condition should account for the phenotypic differences observed at varying levels of exposure to the treatment condition (i.e., genetic predictors and environmental predictors are held constant; Shadish et al., 2002). In an empirical sense, this assumption can only be satisfied with random exposure to the treatment condition (Shadish et al., 2002). Considering the social sciences, GAPSM method provides an adjustment for phenotypic variance unaccounted for by the constituent processes

[33] A regression based approach can establish a counterfactual condition when all of the assumptions are satisfied.
[34] The validity of this claim is assessed in chapters five and six.

upon which it is based. The observed genetic and environmental confounders, which are adjusted for by GAPSM, should have no influence of the observed phenotypic differences at varying levels of exposure to the treatment condition.

Consistent with contemporary scholars, participants can be matched on the GAPS using various techniques (e.g., Mahalanobis metric matching, nearest neighbor matching, nearest neighbor matching within a caliper, optimal matching, genetic matching, among others; Guo and Fraser, 2015). It should be noted that the assumptions associated with PSM (i.e., balancing and common support) must be satisfied when employing GAPSM. To specify, the smallest possible differences, generally determined by comparing distributions of key covariates, between matched participants must exist when examining their correspondence on the polygenic score and the environmental covariates. Furthermore, the matched subsample should have superior balance when compared to the unmatched sample, which is generally determined through an examination of the distribution differences for the polygenic score and the environmental covariates between the treatment and control cases. If optimal balance cannot be achieved, prior scholarship has suggested recoding variables from continuous to semi-continuous measures (Guo and Fraser, 2015). Notably, since the polygenic score is normalized for the independent estimation method, the polygenic scores should not be transformed when using the independent estimation method. For the estimation of the propensity score with the polygenic score as a covariate, it is generally recommended for scholars to standardize or normalize the variable (Dudbridge, 2013). Overall, scholars using GAPSM should examine multiple matching techniques to ensure that the identified subsample achieves balance and common support regarding the polygenic score and the environmental covariates.

Succeeding GAPSM, scholars can employ post-matching statistical techniques to examine the association between the treatment condition and the outcome of interest. The most

common post-matching techniques used to examine the association between the treatment condition and the outcome of interest is a standard independent samples t-test or ANOVA. These post-matching techniques compare the correspondence between the differences in the outcome of interest when exposure to the treatment condition is contained on a dichotomous or semi-continuous scale. For a continuous treatment condition, scholars suggest estimating a bivariate regression model, where the outcome of interest is regressed on the treatment condition (Guo and Fraser, 2015). These analyses should be estimated on the subsample of similar individuals determined by the GAPSM. The current study provides a four-part examination of self-selection through a demonstration of how contemporary statistical techniques (i.e., SSSMs and quantitative genetic methodologies) estimate a counterfactual condition and the divergences between those approaches and the newly developed GAPSM procedure.

CHAPTER 4: THE CURRENT STUDY

As with much of the social sciences, modern criminology suffers from the general inability to establish causal associations between two variables. The inability to estimate causal associations is primarily a function of the reliance on non-experimental designs, which inhibits the approximation of a counterfactual condition. Nevertheless, the effects of non-experimental designs on the estimation of causality can be partially addressed by the implementation of advanced statistical techniques. However, these techniques (as illustrated in the discussions in prior chapters) suffer from their own limitations. To address these limitations the current study proposes a new statistical technique labeled *genetically adjusted propensity score matching* (*GAPSM*). GAPSM generates a counterfactual condition adjusting for both the observed genetic and the observed environmental factors associated with the probability of exposure to the treatment condition. In theory, the counterfactual condition created by GAPSM should hold all observed genetic and observed environmental factors constant at varying levels of exposure to the treatment condition. The *genetically adjusted propensity score* (*GAPS*) is estimated by combining polygenic risk scores with propensity scores through one of two statistical estimation techniques: combined or independent estimation.

To test the validity of the GAPSM technique, the current study addresses two independent research questions, (1) What is the causal effect of intelligence on future antisocial behavior? and (2) What is the causal effect of educational attainment on future antisocial behavior? The current study is demarcated into four distinct analyses.

First, linear regression analyses and ACE decomposition analyses will be conducted to provide an illustration of the effects of social self-selection and biological self-selection on exposure to varying levels of intelligence and educational attainment. Furthermore, a baseline model using the statistical control method will be estimated. Second, propensity score

matching (PSM) will be used to adjust for the influence of observed social factors on an individuals' exposure to the treatment conditions (intelligence and educational attainment). Subsequently, the association between the two independent variables and antisocial behavior will be estimated on the matched sample. Third, MZ difference scores will be used to adjust for the potential unobserved biological factors influencing an individuals' exposure to the treatment conditions. Similar to the PSM analysis, the twin difference score on antisocial behavior will be regressed on the twin difference scores for intelligence and educational attainment (in separate models). Finally, a proof of concept simulation analysis will be conducted, where post-GAPSM estimates are compared to post-MZ difference estimates and unconfounded post-PSM estimates to evaluate the relative proximity to a true point estimate (1.00). The results of which should demonstrate the conditions in which, the GAPSM technique approaches the true point estimate closer than the preexisting methodologies. Furthermore, these analyses should demonstrate the divergences between the counterfactual conditions when adjusting for biological self-selection, social self-selection, or both biological and social self-selection. The first three analyses will be estimated using the National Longitudinal Study of Adolescent to Adult Health (Add Health) and the final analysis will be estimated using data simulated from the Add Health.[35] The current study represents the first analysis estimating the causal effects between two variables using GAPSM.

[35] The simulation data is necessary because contemporary data collection techniques do not meet the data requirements associated with the GAPSM technique. Specifically, the GAPSM technique requires a large sample size (for both the estimation of polygenic score and the GAPSM), genome wide association data, and an extensive survey of environmental conditions to adjust for all potential factors influencing self-selection into a treatment condition. Notably, while the Add Health has yet to release their genome wide association data, it can be speculated that the first dataset satisfying the data requirements of the GAPSM technique will be the Add Health.

CHAPTER 5: METHODS

5.1. National Longitudinal Study of Adolescent to Adult Health

The data for the current dissertation was derived from the restricted version of the National Longitudinal Study of Adolescent to Adult Health (Add Health; Harris et al., 2009). As publicized, the Add Health study was designed to prospectively examine the ecology of a nationally representative sample of American adolescents between 7th and 12th grade. Although initial contact with participants were made between 7th and 12th grade, the three follow-up waves conducted by the Add Health research team spanned 14 years. The array of information collected during the Add Health study has provided numerous empirical avenues to examine sociological, biological, and economic issues within contemporary society. To ensure the national representativeness of the study, the Add Health research team used a cluster sampling design, based upon the implicit stratification of high schools within the United States. See Harris et al., (2009) and Harris, Halpren, Smolen, and Haberstick (2006), for a more detailed description of the sampling design.

The Add Health research team set out to examine of the ecology of American adolescents through the employment of two surveys: in-school questionnaire and the in-home questionnaire. The in-school questionnaire, which was a 60-minute self-administered questionnaire, was distributed to all the students within the selected high schools on a given day. Approximately, 90,000 students completed the in-school questionnaire. In addition to the in-school questionnaire, a random sample of students, within the selected high schools, stratified by grade and sex were selected to complete the in-home questionnaire (Harris et al., 2009). Overall 20,745 students were selected to participate in the in-home questionnaire, approximately 200 students (17 in each stratum) were selected from the 80 pairs of schools. The Add Health research team oversampled individuals based upon ethnic characteristics, saturation, disabled, and genetic characteristics.

For instance, siblings, dizygotic (i.e., fraternal twins) and monozygotic (i.e., identical twins) were oversampled.

The in-home questionnaires for Wave I of the Add Health were conducted between April and December of 1995. The in-home questionnaire was semi self-administered, where the interviewer read the question aloud and the participant responded on a laptop computer. Generally, the interview lasted 90 minutes long and covered a broad array of topics ranging from health status to involvement with delinquent activity. In addition to interviewing the students, the Add Health research team administered parental questionnaires to the resident mother or resident father. These questionnaires were designed to measure household characteristics, inheritable health conditions, and the relationship between the parent and the participant. In total, 17,670 parents were interviewed.

After completing the Wave I interview, the Add Health research team waited approximately 6 months to begin Wave II. Wave II questionnaires were administered approximately 1 year after the participant completed the Wave I interview. At Wave II only in-home questionnaires were administered to approximately 15,000 of the 20,745 individuals that completed the Wave I questionnaire. Notably, respondents within the 12^{th} grade at Wave I and not part of the genetic oversampling were not approached to participate in Wave II. The Wave III in-home questionnaire was administered between August 2001 and April 2002. Of the 20,745 individuals who participated in Wave I, 15,170 completed the Wave III in-home interviews. Due to the aging of the participants, the in-home questionnaire was restructured to account for the lifestyle of young adults. Specifically, topics such as participant relationship status, marital status, childbearing, educational attainment, and intelligence were introduced during the Wave III interview. Similar to Wave I, the Wave III in-home questionnaire was approximately 90 minutes long and was administered as a guided interview.

Wave IV, which was administered in 2008, collected information from 15,701 of the Wave I participants. Corresponding with Wave III, an updated survey instrument was developed to capture factors common within young adulthood. In addition to the phenotypes captured within the Wave I interview, the participants, now between the ages of 24 and 32, were asked about their roles and responsibilities as an adult, their physical and mental health status, and their overall adjustment into adulthood (Harris et al., 2009). Furthermore, personality characteristics, memory tasks, and a supplemental measure of oral vocabulary were administered.

5.1.1. The added value of the Add Health when comparing twin difference scores to PSM

The four waves of data collected during the Add Health study provide a wealth of information for any scholar interested in the biological and sociological predictors of phenotypic variance. Germane to the current dissertation, the Add Health study represents an advantageous dataset for studying genetic and sociological self-selection. Specifically, the Add Health contains at least three waves of data where robust measures of environmental, psychological, and phenotypic information was collected from participants (Beaver et al., 2009; Harris et al., 2009; Miller, Barnes, and Beaver, 2011). Furthermore, the Add Health allows scholars to establish temporal ordering during assessments of self-selection (Singleton and Straits, 2010). Specifically, the first time-period is used to measure the participants' characteristics that increase or decrease the probability of exposure to the treatment condition. The second time-period is used to operationalize the level of exposure each participant had to the treatment condition. The third time-period is used to measure phenotypic outcomes potentially associated with exposure to the treatment condition. In addition to the establishment of temporal ordering, the oversampling of genetically related individuals provides the ideal condition for assessing the correspondence in point estimates between genetically sensitive methodologies and SSSMs on a nationally representative sample (Nedelec and Beaver, 2014; Nedelec, Park, and Silver, 2017). Precisely,

the Add Heath dataset is nationally representative and has a subsample of genetically related individuals.

5.1.2. The external validity of the twin subsample

Concerns regarding the generalizability of the Add Health twin subsample have generally appeared in genetically sensitive publications or critiques of genetically sensitive methodologies (see Barnes et al., 2014). While plausible, various statistical techniques can be used to demonstrate the degree to which the twin subsample differs from the full sample. Consistent with contemporary publications (e.g., Nedelec et al., 2017), the most-often employed test for examining the similarities between the twin subsample and the full sample is an independent samples *t-test*. As such, Appendix C presents three sets of *t-tests* and the standardized difference ($Z\Delta$) value comparing the mean scores of the MZ twins on the covariates to the mean scores of the full sample. Furthermore, the mean scores for the MZ and same sex DZ twins and MZ and difference sex DZ twins on the covariates were compared to the mean scores for the full sample. The results demonstrate that there are a limited number of statistically significant differences between the MZ twin sample and the full sample (7 of 17 differences were statistically significant at $p < .05$), MZ and same sex DZ twin sample and the full sample (8 of 17 differences were statistically significant at $p < .05$), and the MZ and difference sex DZ twin sample and the full sample (9 of 17 differences were statistically significant at $p < .05$). Additionally, across all comparisons the standard difference was never larger than .207 (parental education), suggesting that the statistical significance of the differences could be a function of the large *N* size. Furthermore, while statistically significant, the magnitude of the largest difference suggests that minimal substantive differences exist between the twin subsample and the full sample.

5.2. Studies 1-3: Examining Self-Selection with Preexisting Methodologies

As highlighted in previous chapters, studies 1, 2, and 3 are designed to explore the potential existence of and adjust for the effects of self-selection using preexisting methodologies common within criminology. Furthermore, distinct from Study 4, studies 1, 2, and 3 examine the effects of self-selection on the association between intelligence and antisocial behavior as well as educational attainment and antisocial behavior using data from the restricted use version of the Add Health.

To provide further distinctions, Study 1 illustrates the existence of self-selection by providing a three-part examination of intelligence and educational attainment, and the association between the treatment conditions (i.e., intelligence and educational attainment) and subsequent antisocial behavior. First, Study 1 provides the bivariate and multivariate association between intelligence and antisocial behavior, and educational attainment and antisocial behavior to observe the point estimate confounded by self-selection (i.e., the multivariate model only controls for social factors). Second, SSSMs are used to demonstrate that social factors predict exposure to the two treatment conditions. Finally, an ACE decomposition model is used to demonstrate that a proportion of the variance in the two variables are attributable to genetic factors.

Study 2 is designed to demonstrate the validity of using various iterations of propensity score matching (PSM) when adjusting for the potential effects of social self-selection. As stipulated, the process of conducting a PSM analysis, such as pre-matching comparisons, achieving balance, adjusting caliper, and post-matching comparisons were completed to ensure that the derived point estimates from the post-matching bivariate analyses are theoretically superior to the bivariate point estimates confounded by self-selection. Study 3 is intended to demonstrate the validity of the MZ difference methodology when adjusting for the potential effects of genetic self-selection. As required, the process of conducting a MZ difference analysis,

such as cross twin correlations and cross-twin differences were completed to ensure that the derived point estimates from the post-matching bivariate and multivariate analyses are theoretically superior to the bivariate and multivariate point estimates confounded by genetic self-selection. Overall, in addition to demonstrating the validity and the efficacy of the preexisting methodologies, studies 1, 2, and 3 are designed to highlight the methodological gaps in adjusting for the effects of self-selection to set up the proof of concept simulation analysis provided by Study 4.

5.2.1. Analytical samples[36]

5.2.1.1. Analytical Sample 1: Full Sample

Derived from the restricted use version of the Add Health, a three-step process was used to generate the analytical sample for the standard social science analysis. It should be noted that when obtaining the Add Health restricted use dataset four separate files were provided, each one corresponding to each of the waves associated with the Add Health Study (i.e., Wave I, Wave II, Wave III, Wave IV). First, the data files corresponding to Wave I, Wave III, and Wave IV were merged together using the ID number provided by the Add Health research team. As noted within the subsequent subsection, the dependent variables were derived from Wave IV, the treatment condition variables were derived from Wave III, and the covariates (i.e., factors that predicted exposure to the treatment conditions) were derived from Wave I. Second, individuals who participated in Wave I of the Add Health but not Wave III were removed from the sample (N = 5,548). Third, listwise deletion was used to account for missing data on any of the measures of interest. Overall, the three-step process yielded three analytical samples ranging in size from 6,233 to 12,898.

[36] The analytical samples are presented in the order of largest sample to smallest sample and the order does not correspond to the studies in which they are used.

5.2.1.2. Analytical Sample 2: MZ and Different Sex DZ twin Subsample

Corresponding with the process outlined above, a three-step strategy was employed to develop the second analytical sample (MZ and different sex DZ twin subsample) from the genetic subsample within the restricted use version of the Add Health. First, all the respondents missing data corresponding to the genetic subsample were dropped. This step isolated known kinship pairs from general participants. Second, Wave I, Wave III, Wave IV data for the monozygotic and different sex dizygotic kinship pairs were merged together, where each *case* corresponded to a kinship pair (i.e., each row in the new dataset contained information on both twins)[37]. This creates a long file where data corresponding to twin 1 precedes data corresponding to twin 2 (i.e., a row represents a twin pair).[38] Furthermore, through this process the kinship pair – rather than the individual – becomes the unit of analysis. Finally, due to the need to identify kinship pairs, all participants without a family identification number were dropped from the analytical sample. Overall, the three-step process yielded a MZ and different sex DZ twin subsample with approximately 1,461 individuals.

5.2.1.3. Analytical Sample 3: MZ and Same Sex DZ Subsample

The same four-step strategy was employed to develop the third analytical sample (MZ and same sex DZ twin subsample) from the genetic subsample within the restricted use version of the Add Health. The only difference between the two subsamples are the exclusion of different sex dizygotic kinship pairs. Overall, the four-step process yielded a same sex MZ/DZ twin subsample with approximately 1,047 individuals.

5.2.1.3. Analytical Sample 4: MZ Subsample

[37] Please see Plomin et al., for a detailed explanation of why this step is important for analyzing the influence of genetic (h^2), shared (c^2), and non shared environment (e^2) influence on phenotypic variance.
[38] Step two resulted in dropping all kinship pairs that weren't MZ or DZ twins

A three-step strategy was employed to develop the fourth analytical sample (MZ twin subsample) from the genetic subsample within the restricted use version of the Add Health. First, all the respondents missing data corresponding to the genetic subsample were dropped. This step isolated known kinship pairs from general participants. Second, Wave I, Wave III, Wave IV data for the monozygotic kinship pairs were merged together, where each *case* corresponded to a kinship pair (i.e., each row in the new dataset contained information on both twins). This creates a long file where data corresponding to twin 1 precedes data corresponding to twin 2 (i.e., a row represents a twin pair). Furthermore, through this process the kinship pair – rather than the individual – becomes the unit of analysis. Finally, due to the need to identify MZ kinship pairs, all participants without a family identification number were dropped from the analytical sample. Overall, the four-step process yielded a MZ twin subsample with approximately 570 individuals and 282 twin pairs.

5.2.2. Measures[39]

5.2.2.1. Dependent Variables

Two dependent variables were created for the current study: *Delinquency* (Wave IV) and *Drug Use* (Wave IV). Delinquency at Wave IV was measured as the mean score on eleven items (five items had to have valid values) capturing self-reported involvement in minor and serious delinquent activities. Items such as "In the past 12 months, how often did you go deliberately damage property that didn't belong to you?" and "In the past 12 months, how often did you use or threaten to use a weapon to get something from someone?" were included in this measure. Scores on the eleven items were subjected to a reliability analysis, which indicated a moderately high level of reliability ($\alpha = 67$). Drug use at Wave IV was measured as the standardized mean

[39] Appendix B provides a list of the items (and the Add Health reference ID) used to create the measures for the current study.

score on four items (two items had to have valid values) capturing self-reported involvement in minor and serious drug use. Items such as "During the past 30 days, on how many days did you smoke cigarettes?" and "During the past 12 months, on how many days did you use [favorite drug]?" were included in this measure. Scores on the four items were subjected to a reliability analysis, which indicated a moderate level of reliability ($\alpha = 48$). Higher scores on delinquency (Wave IV) and drug use (Wave IV) represent higher levels of self-reported antisocial involvement.

5.2.2.2. Treatment Conditions

Two treatment condition variables were created for the current study: *Intelligence* (Wave III) and *Educational Attainment* (Wave III). Intelligence at Wave III was measured as the standardized score on the Peabody Picture Vocabulary Test (PPVT). The PPVT was administered during the Wave III interview and is a validate measure of verbal intelligence (Rao, Leo, Bernardin, and Unverzagt, 1991). Higher scores on intelligence (Wave III) represent higher levels of cognitive abilities. Educational attainment at Wave III dichotomously captured if the participant had completed at least one year of college at the time of the Wave III interview. At the time of the Wave III interview, the majority of the of the participants were over the age of 18 (approximately 99 percent, 149 individuals, were 18 years of age at the Wave III interview) and should have had the opportunity to enroll in higher education institutions. Educational attainment was coded where a value of "1" indicated that the participant self-reported completing at least some college and a value of "0" indicated that the participant completed 12th grade or less at the time of the interview.[40]

5.2.2.3. Covariates (Wave I)

[40] While unable to directly determine high school enrollment, only approximately 1 percent of the sample was 18 years of age at the time of the interview.

Thirteen covariates were measured at Wave I and included in the multivariate models as control variables. Evidence suggests that the covariates are empirically and theoretically related to levels of intelligence, educational attainment, and self-reported antisocial involvement (Silver and Nedelec, 2018). First, *Age* was measured as the age of the participant at the time of the Wave I interview. Age was created by subtracting the birth year for the participants from the year of the Wave I interview. Second, *Non-White* was measured as the interviewer's perception of the racial category of the participant, where a value of "0" indicated that the participant was White and a value of "1" indicated that the participant was non-White. Third, *Male* was operationalized as the interviewer's perception of the biological sex that the participant belonged to, where a value of "0" indicated female and a value of "1" indicated male.[41]

Fourth, *Parental Income* was a continuous operationalization of the parent reported amount of total income a participant's household earned before taxes in 1994. Specifically, the question asked the parents their own income, the income of everyone else in your household, and income from welfare benefits, dividends, and all other sources. Higher scores on parental income indicate higher levels of household income within the family. Fifth, *Parental Employment Status* was the self-reported employment status of the parent who completed the parent portion of the Add Health interview. Parent employment status was coded as a dichotomy where a value of "0" indicated that the parent was not employed full time in the past 12 months, and a value of "1" indicated that the parent was employed full time in the past 12 months. Sixth, *Parental Education* was a self-reported measure of the parent's educational attainment. A value of "0" on parental education signifies that the parent did not attend college, while a value of "1" on parental education signifies that the parent did attend college at some point.

[41] We deferred to the interviewer's perception of the racial category and biological sex to capture the outward reflection of the participants racial category and biological sex.

Seventh, *Maternal Conflict* was operationalized as the standardized mean of seven items (four items had to have valid scores) requiring the participants to self-report the amount of conflict they have had with their mother (or maternal caregiver). All of the items were reverse recoded (so that higher scores indicated greater maternal conflict) and were subjected to a reliability analysis, which indicated a high level of reliability between the items ($\alpha = .85$). Eight, *Paternal Conflict* was operationalized as the standardized mean of seven items (four items had to have valid scores) requiring the participants to self-report the amount of conflict they have had with their father (or fraternal caregiver). All of the items were reverse recoded (so that higher scores indicated greater paternal conflict) and subjected to a reliability analysis, which indicated a high level of reliability between the items ($\alpha = .89$). Higher scores on maternal conflict and paternal conflict indicate higher levels of parental conflict in the participants' lives.[42] Ninth, *School Attachment* measured participants' self-reported attachment to prosocial school activities. Scores on school attachment represent the standardized mean of nine items (five items had to have valid scores), and higher values indicate higher levels of attachment to school activities ($\alpha = .78$).

Tenth, *Social Support* was operationalized as the mean of eight items (four had to have valid scores) capturing participants' perceptions of support from the people involved in their lives. One item was reverse recoded to account for the structure of the question (see Appendix B). Higher scores on social support indicate higher levels of support from the people involved in the participant's life ($\alpha = .85$). Eleventh, *Peer Drug Use* was operationalized as the amount of self-reported best friends who smoked at least once a day or drank and used marijuana at least once a month. Participants had to have valid scores on all three items to receive a score on peer

[42] These were demarcated just to observe if maternal conflict and paternal conflict differentially influenced educational attainment or intelligence.

drug use. Higher values on peer drug use indicated that a higher proportion of the participant's best friends engaged in antisocial activities such as drinking and smoking ($\alpha = .76$). Twelfth, *Baseline Delinquency* was operationalized as the mean of 14 items (seven items had to have valid scores) capturing self-reported involvement in minor and serious delinquent activities during Wave I ($\alpha = .83$). Higher scores on baseline delinquency correspond to higher levels of self-reported antisocial conduct at Wave I. Thirteenth, *Baseline Drug Use* was operationalized as the mean of six items (three items had to have valid scores) capturing self-reported involvement in minor and serious drug use during Wave I ($\alpha = 68$). Higher scores on baseline drug use correspond to higher levels of self-reported drug use at Wave I.

5.2.3. Analytical strategies

5.2.3.1. Study 1: Exploring the Existence of Social and Genetic Self-Selection

A four-step analytical plan was used to evaluate the degree to which self-selection could potentially confound the association between the treatment conditions (i.e., intelligence and educational attainment at Wave III) and the dependent variables (i.e., delinquency and drug use at Wave IV). First, descriptive statistics and bivariate correlation coefficients were produced for the full sample, the MZ and same sex DZ twin subsample, and the MZ and different sex DZ twin subsample. Second, bivariate and multivariate OLS models were estimated, where the dependent variables (i.e., delinquency at Wave IV and drug use at Wave IV) were regressed on the treatment conditions (i.e., Intelligence at Wave III[43] and educational attainment at Wave III) and the covariates. Third, using OLS and binary logistic regression models the treatment conditions (i.e., Intelligence at Wave III and educational attainment at Wave III) were regressed on the 13

[43] The dependent variables were regressed on a linear, quadratic, and cubic specification of intelligence to account for the evidence suggesting that intelligence and antisocial behavior could potentially have a curvilinear association. For all subsequent models, the dependent variables were regressed on a linear, quadratic, and cubic specification of intelligence.

covariates. By regressing the treatment conditions (i.e., Intelligence at Wave III and educational attainment at Wave III) on the covariates, one can observe the degree to which environmental factors predict selection into the treatment conditions. Finally, ACE decomposition models were estimated where the variance in the treatment conditions (i.e., Intelligence at Wave III and educational attainment at Wave III) were attributed to genetic (*a*), shared-environmental (*c*), and non shared environmental factors (*e*). Two sets of ACE decomposition models were estimated, one for the MZ and same sex DZ twin subsample and one for the MZ and different sex DZ twin subsample. In sum, the five-steps provide an indication of how both genetic and social self-selection could have influenced scores on the two treatment conditions (i.e., Intelligence at Wave III and educational attainment at Wave III).

5.2.3.2. Study 2: Adjusting for Social Self-Selection

A four-step analytical strategy was employed to adjust for the confounding effects of social self-selection on the association between intelligence and antisocial behavior, and educational attainment and antisocial behavior. First, for the dichotomously coded educational attainment (Wave III) the predicted probabilities from the binary logistic regression models where educational attainment was regressed on the 13 covariates were saved. These predicted probabilities served as the propensity scores for the PSM. Second, the treatment and control cases for educational attainment (i.e., Did not complete one year of college vs. did complete one year of college) were matched using nearest neighbor matching with varying calipers (i.e., .05, .01, .005, .001, .0001). After the cases were matched, balance was assessed using *t*-tests and percent bias values.

Third, intelligence (Wave III) was subjected to generalized propensity score (GPS) matching, where participants were matched on the 13 covariates. To satisfy the assumptions associated with balancing during GPS matching, ten equal-width percentiles were created on the

intelligence scale. The range of scores on each percentile were specified as such: 10th percentile = 7 to 79, 20th percentile = 80 to 87, 30th percentile = 88 to 91, 40th percentile = 92 to 96, 50th percentile = 97 to 103, 60th percentile = 104 to 106, 70th percentile = 107 to 108, 80th percentile = 109 to 111, 90th percentile = 112 to 116, 100th percentile = 117 to 122. The percent reduction in bias (i.e., percent increase in balance) for the 13 covariates was assessed between the pre-matching and post-matching samples. Finally, various post-matching bivariate OLS regression models were estimated where delinquency and drug use was regressed on educational attainment and intelligence (linear, quadratic, cubic specifications). All analyses were estimated on the full Add Health sample.

5.2.3.3. Study 3: Adjusting for Genetic Self-Selection

A four-step analytical strategy was employed to adjust for the confounding effects of genetic self-selection on the association between intelligence and antisocial behavior, and educational attainment and antisocial behavior. First, descriptive statistics and *t*-tests were produced to evaluate if substantive mean differences existed between the MZ twins randomly assigned as twin 1 and twin 2 in the twin pairs. Second, cross-twin correlation coefficients were estimated to observe the differences on covariates within twin pairs. Furthermore, the cross-twin correlation coefficients help determine the variables in which the assumptions associated MZ difference scores could be satisfied. Third, MZ difference scores on the dependent variables (delinquency and drug use at Wave IV), the treatment conditions (intelligence and educational attainment at Wave III), and the covariates (Wave I) were created for the variables with a cross-twin correlation coefficients below 1.00.[44] Descriptive statistics for the MZ difference scores were produced. Finally, bivariate and multivariate OLS models were estimated where the MZ difference score for the dependent variables (i.e., Delinquency at Wave IV and drug use at Wave

[44] A MZ difference score for Non-White was not created because the minor discrepancy in race between MZ twins corresponds to measurement error.

IV) were regressed on the MZ difference score for the treatment conditions (i.e., Intelligence at Wave III[45] and educational attainment at Wave III) and the MZ difference scores for the covariates. All analyses were estimated on the MZ twin subsample using robust standard errors.

5.3. Study 4: GAPSM Proof of Concept

As initially proposed in Subsection 3.4. of Chapter 3, the purpose of the current dissertation is to advance upon the preexisting methodologies by adjusting for both social and genetic self-selection during the estimation of post-matching point estimates. The proposed advancement, *genetically adjusted propensity score matching (GAPSM)*, provides a novel approach that allows for the integration of polygenic risk scores and socially constructed predicted probabilities during the matching of participants. Specifically, GAPSM is designed to attain estimates virtually uninfluenced by the effects of observed genetic and observed environmental factors on self-selection into a treatment condition. Nevertheless, while the theoretical arguments presented in Subsection 3.4. of Chapter 3 are supported, the principal procedure for evaluating the initial statistical validity of a new methodology is the employment of a simulation analysis (Lewis and McKenzie, 2017).

For evaluating the potential validity of statistical methodologies, simulation analyses provide three advancements over the employment of primary or secondary data (Lewis and McKenzie, 2017). First, simulations provide the ability to compare the point estimates derived from different methodologies to a true point estimate specified by the user. Specifically, in the current context we can compare the point estimates for the bivariate association between a treatment condition and an outcome of interest derived from an unconfounded PSM, a MZ difference score, and various iterations of a GAPSM to the true point estimate. Thus, we can

[45] The dependent variables were regressed on a linear, quadratic, and cubic specification of intelligence to account for the evidence suggesting that intelligence and antisocial behavior could potentially have a curvilinear association. For all subsequent models, the dependent variables were regressed on a linear, quadratic, and cubic specification of intelligence.

statistically identify the conditions in which the point estimate for the GAPSM approach is closer (or not) to the true point estimate than the point estimate derived from the PSM approach and the MZ difference score approach.

Second, as alluded to, simulation analyses allow the user to specify an infinite number of conditions in which comparisons can occur. In the current context, the specification of divergent conditions allows for the identification of various conditions in which the point estimate for the GAPSM approach is closer (or not) to the true point estimate than the point estimates derived from the PSM approach and the MZ difference score approach. Finally, since simulation analyses are fundamentally derived from Bayesian statistics, information from previous studies can be integrated into the simulation analysis to emulate potential realities when considering the bivariate association between a treatment condition and an outcome of interest. For instance, pervious research and the findings provided in studies 1, 2, and 3 could and should guide the selection of conditions in which comparisons can occur. Specifically, considering that complex traits, such as intelligence and educational attainment, are only partially genetic the number of conditions in which comparisons should occur can be reduced to conditions that could potentially exist. Consistent with these advantages, the simulation analyses provided in Study 4 assess the empirical validity of the GAPSM approach, the conditions in which the GAPSM approach is superior to the unconfounded PSM approach, and the conditions in which the GAPSM approach is superior to the MZ difference score approach.

A superior way of demonstrating the validity of the GAPSM approach through simulation analyses is to use the theoretical logic developed by behavioral and quantitative geneticists. To revisit, the variance in a phenotype (or treatment condition) can be partitioned into three independent latent factors: genetics (a), the shared environment (*c*), and the non shared environment (e). Rationally, the two preexisting methodologies (PSM approach and MZ

difference score approach) and the GAPSM approach derive matches using information from one or more of the following factors.

In an unconfounded PSM approach, it can be theoretically assumed that all the variance in the treatment condition predicted by the shared environment (*c*) and the non shared environment (*e*) is used to match participants. While the unconfounded PSM represents an ideal situation, it is more likely that only a portion of the variance in the treatment condition predicted by the shared environment (*c*) and the non shared environment (*e*) is used to match participants. The MZ difference score approach allows scholars to assume that all the variance in the treatment condition affected by genetics (*a*) and the shared environment (*c*) is used to match participants (i.e., MZ twins). Considering that the theoretical foundation is quite strong and that the MZ difference score approach relies on adjusting for unobserved factors, it is safe to assume that the majority of MZ difference score analyses adjust for all of the variance in the treatment condition predicted by unobserved genetic factors (*a*) and the shared environment (*c*; Plomin et al., 2013).

Distinct from the unconfounded PSM and the MZ difference score, the GAPSM approach theoretically allows scholars to assume that all of the variance in the treatment condition predicted by genetics (*a*) will be used to match participants (due to the integration of a polygenic risk score)[46], but only a proportion of the variance in the treatment condition predicted by the shared environment (*c*) and the non shared environment (*e*) is used to match participants. Given these theoretical expectations for the GAPSM approach, 21 iterations of GAPSM were specified where fluctuating proportions of the variance in the treatment condition predicted by the shared environment (*c*) and the non shared environment (*e*) were used to match participants. Succeeding the matching, bivariate point estimates, where the outcome of interest was regressed on the

[46] This assumption can be satisfied if the user employs the entire genome, rather than statistically significant alleles, to create the polygenic risk score.

treatment condition, were produced for the 21 iterations and compared to conditions emulating that of the unconfounded PSM approach (all the variance in the treatment condition predicted by the shared environment (*c*) and the non shared environment (*e*) are used to match participants) and the MZ difference score approach (all the variance in the treatment condition predicted by genetics (*a*) and the shared environment (*c*) are used to match participants).

5.3.1. Simulating data

For the ease of explaining the creation of the simulation data and the subsequent simulation analyses, the *R* code for the specification of the data will be reviewed in the current subsection. The code presented can be imported into *R* and used to generally reproduce the results presented.

```
1. # Specifying the number of cases within the dataset.
2. n = 50000
3. # Specifying uncorrelated covariates of interest
4. x1 = rnorm(n,100,15)
5. x2 = rnorm(n,100,15)
6. x3 = rnorm(n,100,15)
7. x4 = rnorm(n,100,15)
8. x5 = rnorm(n,100,15)
9. x6 = rnorm(n,100,15)
10. x7 = rnorm(n,100,15)
11. x8 = rnorm(n,100,15)
```

Prior to the creation of the dataset, the number of cases included in the dataset must be specified. As specified in Line 2, the dataset was specified to contain 50,000 cases. Succeeding the specification of the N-size, lines 4 through 11 create the uncorrelated covariates that will be used to match the treatment and control participants. As indicated in subsequent code, x1–x4 will measure the non shared environment and x5–x8 will measure the shared environment. The command *rnorm* informs *R* to create a normal distribution of *n* cases (i.e., 50000) with a mean of 100 and a standard deviation of 15.

```
12. # Specifying the non shared environment.
13. e = .25*x1+.25*x2+.25*x3+.25*x4
14. # Specifying the shared environment.
```

```
15. c = .25*x5+.25*x6+.25*x7+.25*x8
16. # Specifying the genetics.
17. a = 0*s+0*ns+ rnorm(n,100,15)
18. # Specifying an error term.
19. E = 0*g+0*s+0*ns+ rnorm(n,100,15)
```

As indicated by Line 13, the simulation of the non shared environment (*e*) was equal to .25*x1+.25*x2+.25*x3+.25*x4. The *R* specification for the non shared environment (*e*; Line 13) can be rationally translated into a regression formula, where the intercept is equal to "0" and the slope of the regression line for each variable (i.e., x1, x2, x3, x4) is equal to ".25." Furthermore, evident by the specification, all the variance in the non shared environment (*e*) is equally distributed amongst the four variables (i.e., x1, x2, x3, x4). The equal distribution of the variance in the non shared environment (*e*) between the four variables (i.e., x1, x2, x3, x4) is important to specification of the 21 iterations of GAPSM. Similar to the non shared environment, the shared environment (*c*; Line 15) was specified to be equal to .25*x5+.25*x6+.25*x7+.25*x8. Again, the *R* specification for the shared environment (*c*; Line 15) can be rationally translated into a regression formula, where the intercept is equal to "0" and the slope of the regression line for each variable (i.e., x5, x6, x7, x8) is equal to ".25." Notably, due to the reliance on different variables for the specification of the non shared environment (*e*; Line 13; x1, x2, x3, x4) and the shared environment (*c*; Line 15; x5, x6, x7, x8) these terms (i.e., *e* and *c*) can be assumed to be uncorrelated.

Line 17 provides the specification for the genetic factors (*a*), which is a normal distribution of 50,000 cases with a mean of 100 and a standard deviation of 15. Furthermore, as specified (i.e., 0*s+0*ns+ rnorm(n,100,15)), genetic factors (*a*) was uncorrelated with the non shared environment (*e*) and the shared environment (*c*). Line 19 provides the specification for the error term (*E*), which is a normal distribution of 50,000 cases with a mean of 100 and a standard deviation of 15. Again, as specified (i.e., 0*g+0*s+0*ns+ rnorm(n,100,15)), the error term (*E*)

was uncorrelated with the genetic factors (*a*), the non shared environment (*e*), and the shared environment (*c*).

The specification of uncorrelated terms for the genetic factors (*a*), the non shared environment (*e*), the shared environment (*c*), and the error term (*E*) relies on theoretical knowledge that the probability of exposure to a treatment condition can be demarcated into four independent factors (genetic factors (*a*), non shared environmental (*e*), shared environmental (*c*), and error (*E*)). To review, the 19 lines of *R* code covered so far create four normally distributed variables (the genetic factors (*a*), the non shared environment (*e*), the shared environment (*c*), and the error term (*E*)), with two of those variables (the non shared environment (*e*), the shared environment (*c*)) delineated into four measures each (the non shared environment (*e*) = x1, x2, x3, x4; the shared environment (*c*) = x5, x6, x7, x8).

```
20. # Normalizing the non shared environment.
21. e_n = ((e-min(e))/(max(e)-min(e)))
22. # Normalizing the shared environment.
23. s_n = ((s-min(s))/(max(s)-min(s)))
24. # Normalizing the genetics.
25. a_n = ((a-min(a))/(max(a)-min(a)))
26. # Normalizing an error term.
27. E_n = ((E-min(E))/(max(E)-min(E)))
```

For ease of interpretation, preceding the specification of a dichotomous treatment condition it is commonly acceptable to normalize (i.e., transform the normally distributed variables to a distribution that ranges between 0 and 1) the variables of interest (genetic factors (*a*), non shared environment (*e*), shared environment (*c*), and error term (*E*)) to emulate that of a risk score. By normalizing the data, one can attribute the desired amount of variance in the treatment condition to each of the four factors and create a dichotomous outcome where scores (or probabilities) greater than .50 are equal to a value of "1" on the dichotomous treatment condition and scores (or probabilities) equal to or less than .50 are equal to a value of "0" on the dichotomous treatment condition. Lines 21, 23, 25, and 27 are the specification for normalizing

the non shared environment (*e*), shared environment (*c*), genetic factors (*a*), and the error term (*E*), respectively. As outlined in the subsequent section, the normalization of these four variables permits the specification of numerous treatment conditions in which the point estimates derived from the unconfounded PSM approach, the MZ difference score approach, and various GAPSM iterations can be compared to a true (i.e., specified) point estimate.

5.3.2. Specification of the treatment conditions (40 treatment conditions)

Consistent with contemporary scholarship on simulation analysis (e.g., Evans and Olson, 2001; Lewis and McKenzie, 2017), the comparisons between the point estimates derived from an MZ difference score approach, the point estimates derived from an unconfounded PSM approach, and the point estimates derived from various GAPSM iterations were estimated succeeding the specification of various treatment conditions. For the current study, 40 different specifications of the treatment condition were created.[47] These specifications of the treatment condition differed by the amount of variance in the treatment condition attributable to genetic factors (*a*), the non shared environment (*e*), and the shared environment (*c*).

For the first set of 13 specifications of the treatment condition, the amount of variance in the treatment condition predicted by genetic factors (*a*) increased from .05 to .65 in .05 increments, while the variance in the treatment condition predicted by the non shared environment (*e*) and the shared environment (*c*) were set equal to each other. For instance, when $a = .20$, both *e* and $c = .38$. To reduce the likelihood of estimating a perfect model, 4 percent (or .04) of the variance in the treatment condition was predicted by the error term (*E*).[48] For the second set of 13 specifications of the treatment condition, the amount of variance in the treatment condition predicted by genetic factors (*a*) increased from .05 to .65 in .05 increments,

[47] Each of the 40 specifications for the treatment conditions are provided in Appendix D.
[48] Four percent (or .04) of the variance in the treatment condition was consistently predicted by the error term (E) across the 40 specifications of the treatment conditions.

while the variance in the treatment condition predicted by the non shared environment (*e*) was approximately three times that of the variance in the treatment condition predicted by the shared environment (*c*). For example, when *a* = .20, *e* = .56 and *c* = .19.

For the third set of 13 specifications of the treatment condition, the amount of variance in the treatment condition predicted by genetic factors (*a*) increased from .05 to .65 in .05 increments, while the variance in the treatment condition predicted by the shared environment (*c*) was approximately three times that of the variance in the treatment condition predicted by the non shared environment (*e*). To provide an example, when *a* = .20, *e* = .19 and *c* = .56. The final specification was the point of equivalence, where the variance in the treatment condition predicted by genetic factors (*a*), the non shared environment (*e*), and the shared environment (*c*) were set all equal to each other. For example, the variance in the treatment condition predicted by genetic factors (*a*), the non shared environment (*e*), and the shared environment (*c*) were all set to .32. Again, the specification of these conditions is best illustrated by the *R* code, which takes the form of a simple regression formula.

```
28. ## Condition of equivalence ##
29. # Creating the treatment condition
30. tc.EQ = .32*e_n + .32*c_n + .32*a_n +.04*E_n
```

Lines 28-30 of the *R* code represents the specification for the treatment condition of equivalence. Considering that the genetic factors (*a*), the non shared environment (*e*), the shared environment (*c*), and the error term (*E*) were normalized, the specification of .32* (i.e., the slope coefficient) can be interpreted as the percentage of the variance in the treatment condition attributed to the specified factor. Furthermore, consistent with the normalized factors *tc.EQ* (i.e., the treatment condition) is on a continuous scale ranging from 0 to 1. To provide a mathematical example, if a participant's score on *e_n* (normalized non shared environmental factors) was .32, on *c_n* (normalized shared environmental factors) was .32, and on *a_n* (normalized genetic

factors) was .32, the mathematical formula presented in line 30 would provide a score of .31 on *tc.EQ*. The difference can be attributed to the existence of randomly distributed measurement error (*E*).

```
31. # creating a dichotomy, where anyone below .50 scored a 0 and anyone
    above .50 scored a 1
32. tEQ = tc.EQ
33. tEQ[tEQ <= .50] = 0
34. tEQ[tEQ > .50] = 1
```

The 40 specifications of the treatment condition were recoded into dichotomous variables where participants with scores equal to or less than .50 received a score of "0" on the *treatment dichotomy* and participants with scores greater than .50 received a score of "1" on the *treatment dichotomy.* Considering the example provided above, the participant's score on the *treatment dichotomy* for the equivalence treatment condition (i.e., variance in *t* explained by *e* = .32; variance in *t* explained by *c* = .32; variance in *t* explained by *g* = .32) would most likely be a "0" depending upon the score on the error term (*E*). Lines 31 through 34 represent this coding scheme.

5.3.3. Specification of the outcome of interest

```
35. # the dependent variable for the condition of equivalence
36. yEQ = 1.00*tEQ + 1.25*e_n + 1.25*c_n + 1.25*a_n + .005*E_n
```

Succeeding the specification of the dichotomous treatment conditions, the *Outcomes of Interest* (*y*) were specified as a function of the dichotomous treatment conditions (*t*), the non shared environment (*e*), the shared environment (*c*), the genetic factors (*a*), and the error term (*E*). The formula for the outcome of interest when the specification of the treatment condition was the point of equivalence is presented in Line 36. Due to the modifications in the dichotomous treatment conditions (i.e., the 40 specifications described in the preceding subsection) a distinct outcome of interest (*y*) was created for each specification of the treatment

condition (i.e., 40 different *y* variables). Nonetheless, all the outcomes of interest (*y*) were identical excluding the treatment condition (*t*), which varied between specifications (i.e., *t1* for specification 1 up until *t39* for specification 39, and *tEQ* for the point of equivalence). Notably, as indicated by the formula, the true point estimate for the association between the treatment conditions (*t*) and the outcomes of interest (*y*) was set at 1.00. Given that the non shared environment (*e*), the shared environment (*c*), the genetic factors (*a*), and the error term (*E*) confound the association between the treatment conditions (*t*) and the outcomes of interest (*y*) the only way to obtain the true point estimate would be to run a multivariate regression model or a post-matching regression model where all of the independent factors (i.e., *t*, *e*, *c, a, E*) are specified within the formula. Any other specification of a multivariate regression model or a post-matching regression model would produce a biased point estimate of the association between the treatment conditions (*t*) and the outcomes of interest (*y*).

5.3.4. Analytical strategy

To evaluate the validity of the GAPSM methodology, 30 point estimates were produced for each specification of the treatment condition (i.e., 40 different specifications, 120 point estimates in total). First, two pre-matching point estimates were produced, one being the true point estimate (i.e., the full model) and the other being the bivariate confounded association between the treatment condition (*t*) and the outcome of interest (*y*; i.e., *y* regressed on *t*). Second, the three distinct point estimates were produced after matching the participants on the non shared environment (*e*), the shared environment (*c*), and the genetic factors (*a*), respectively.[49]

[49] To specify, all the post-matching point estimates were derived from a bivariate analysis of the outcome of interest (*y*) regressed on the dichotomous treatment condition (*t*) using a matched sample. The matched samples were created using nearest neighbor matching with a caliper of .05, succeeding the estimation of the specified predicted probabilities derived from a binary logistic regression model where the dichotomous treatment condition (*t*) was regressed on the factors specified in the condition of interest.

Third, a non shared environment (*e*) and a shared environment (*c*; where participants were matched with scores on x1, x2, x3, x4, x5, x6, x7, and x8) post-matching point estimate was produced to emulate that of an unconfounded post-matching PSM point estimate. Fourth, a non shared environment (*e*) and genetic factors (*a*; where participants were matched with scores on x1, x2, x3, x4, and *a*) post-matching point estimate was produced to emulate that of a genetically sensitive analysis without accounting for the shared environment. Fifth, a shared environment (*c*) and genetic factors (*a*; where participants were matched with scores on x5, x6, x7, x8, and *g*) post-matching point estimate was produced to emulate that of a MZ difference score point estimate. Sixth, a non shared environment (*e*), a shared environment (*c*), and genetic factors (*a*; where participants were matched with scores on x1, x2, x3, x4, x5, x6, x7, x8, and *g*) post-matching point estimate was produced to emulate that of a perfect post-GAPSM point estimate. Generally, the perfect post-GAPSM point estimate was equal to that of the true point estimate.

Finally, 21 iterations of post-GAPSM point estimates were produced to identify the various situations where the post-GAPSM point estimates would more closely approach the true point estimate than the MZ difference score point estimate and the unconfounded propensity score point estimate.[50] Six of the iterations varied the degree to which participants were matched on the non shared environment (*e*). Specifically, post-GAPSM point estimates were produced when the participants were matched on genetics (*a*) and one, two, or three of the predictors for the non shared environment (*e*; i.e., x1, x2, x3, and x4) and post-GAPSM point estimates were produced when the participants were matched on genetics (*a*), the shared environment (*c*), and one, two, or three of the predictors for the non shared environment (*e*; i.e., x1, x2, x3, and x4).

[50] All the post-GAPSM point estimates were produced when participants were matched on genetics (*g*) because of the theoretical expectation that a polygenic score would measure all the variation in a treatment condition predicted by an individual's genome.

Furthermore, six of the iterations varied the degree to which participants were matched on the shared environment (*c*). To specify, post-GAPSM point estimates were produced when the participants were matched on genetics (*a*) and one, two, or three of the predictors for the shared environment (*c*; i.e., x5, x6, x7, and x8) and post-GAPSM point estimates were produced when the participants were matched on genetics (*a*), the non shared environment (*e*), and one, two, or three of the predictors for the shared environment (*c*; i.e., x5, x6, x7, and x8). The final nine iterations varied the degree to which participants were matched on the non shared environment (*e*) and the degree to which participants were matched on the shared environment (*c*). To provide a description, post-GAPSM point estimates were produced when the participants were matched on genetics (*a*) and one, two, or three of the predictors for the non shared environment (*e*; i.e., x1, x2, x3, and x4) and one, two, or three of the predictors for the shared environment (*c*; i.e., x5, x6, x7, and x8). Overall, the 21 iterations of post-GAPSM point estimates were designed to provide a substantive number of comparisons between the GAPSM methodology and the PSM approach, and the GAPSM methodology and the MZ difference score approach.

CHAPTER 6: RESULTS

6.1. Study 1: Results: Exploring the Existence of Social and Genetic Self-Selection

6.1.1 Descriptive statistics full sample

Table 6.1 presents the descriptive statistics for the variables of interest within the full restricted use Add Health Data set. As illustrated, the average age of the Wave I participants was 16.15, 38 percent of the sample was of Non-White and was 49 percent of the sample was male. Additionally, parental respondents indicated an average household income of 46 thousand dollars a year (Parent Income: $\bar{X}$ = 45.73; SD = 51.62), 62 percent said that they were employed full time, and 43 percent said that they completed at least some college. The participants on average reported a low level of maternal ($\bar{X}$ = -.01; SD = 1.28) and paternal conflict ($\bar{X}$ = -.01; SD = 1.38; Min, Max: Maternal Conflict: -1.35, 7.49; Paternal Conflict: -1.37, 6.12), and a high level of school attachment ($\bar{X}$ = .01; SD = 1.08; Min, Max: -4.88, 2.06). The descriptive statistics indicated that the majority of the sample had high levels of social support ($\bar{X}$ = 7.98; SD = 1.19), had approximately one close friend use drugs at the baseline ($\bar{X}$ = .85; SD = .89), had engaged in low levels of delinquent behaviors at the baseline ($\bar{X}$ = .55; SD = .70), and had reported using low levels drugs at the baseline ($\bar{X}$ = .53; SD = .47).

In terms of the treatment conditions, the average score on the Peabody Picture Vocabulary Test at Wave III was approximately 98.48, and the majority of the sample had completed at least 1 year of college prior to the interview (educational attainment: $\bar{X}$ = .54). At Wave IV, the participants on average reported low levels of delinquent activity ($\bar{X}$ = .07; SD = .25) and low levels of drug use on the standardized scale ($\bar{X}$ = .01; SD = 1.25; min, max: -1.60, 3.45).

Table 6.1: Descriptive statistics for the full sample.

	N	$\bar{X}$ (%)	SD	Min, Max
Dependent Variables (Wave IV)				
Delinquency	15,638	.07	.25	.00, 5.60
Drug Use	15,522	.01	1.25	-1.60, 3.45
Treatment Conditions (Wave III)				
Intelligence	14,652	98.48	17.09	7, 122
Educational Attainment	15,183	54%	.50	0, 1
Covariates (Wave I)				
Age	20,728	16.15	1.74	12, 21
Non-White	20,704	38%	.49	0, 1
Male	20,743	49%	.50	0, 1
Parent Income	15,351	45.73	51.62	0, 999
Parent Employment Status	17,609	62%	.49	0, 1
Parent Education	17,527	43%	.49	0, 1
Maternal Conflict	19,386	-.01	1.28	-1.35, 7.49
Paternal Conflict	14,403	-.01	1.38	-1.37, 6.12
School Attachment	20,279	.01	1.08	-4.88, 2.06
Social Support	20,092	7.98	1.19	2, 10
Peer Drug Use	20,028	.85	.89	0, 3
Baseline Delinquency	20,410	.55	.70	0, 6
Baseline Drug Use	20,221	.53	.47	0, 2

Notes: Drug Use, Maternal Conflict, Paternal Conflict, and School attachment were standardized to account for the differences in coding schemes between items.

Table 6.2 provides the correlation matrix for the variables of interest within the full restricted use Add Health Data set. As illustrated, delinquency (Wave IV) was associated with educational attainment (Wave III; $r = -.06$, $p < .05$), but not intelligence (Wave III; $r = -.01$, $p > .05$). Furthermore, drug use (Wave IV) was positively associated with intelligence (Wave III; $r = .10$, $p < .05$) and negatively associated with educational attainment (Wave III; $r = -.07$, $p < .05$). Additionally, intelligence was associated with seven of the covariates (Non-White: $r = -.29$; Male: $r = .03$; Parental Income: $r = .17$; Parent Employment Status: $r = .02$; Parent Education: $r = .20$; School attachment: $r = .03$; Baseline Drug Use: $r = .04$, all associations were $p < .05$). Furthermore, educational attainment was associated with all of the Wave I covariates (Age: $r = .02$; Non-White: $r = -.04$; Male: $r = -.07$; Parental Income: $r = .21$; Parent Employment Status: $r = .06$; Parent Education: $r = .28$; Maternal Conflict: $r = -.05$; Paternal Conflict: $r = -.05$; School attachment: $r = .16$; Social Support: $r = .09$; Peer Drug Use: $r = -.17$; Baseline Delinquency: $r = -.10$; Baseline Drug Use: $r = -.14$, all associations were $p < .05$).

Table 6.2. Correlation matrix for the Add Health sample.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Dependent Variables (Wave IV)																
Delinquency (1)	--															
Drug Use (2)	.28*	--														
Treatment Conditions (Wave III)																
Intelligence (3)	-.01	.10*	--													
Educational Attainment (4)	-.06*	-.07*	.30*	--												
Covariates (Wave I)																
Age (5)	-.07*	-.09*	-.01	.02*	--											
Non-White (6)	.03*	-.14*	-.29*	-.04*	.04*	--										
Male (7)	.13*	.16*	.03*	-.07*	.05*	-.01	--									
Parent Income (8)	-.02*	.04*	.17*	.21*	-.01	.13*	-.01	--								
Parent Employment Status (9)	.01	.03*	.02*	.06*	.01	.06*	-.01	.03*	--							
Parent Education (10)	.01	.04*	.20*	.28*	-.05*	-.02*	.01	.23*	.15*	--						
Maternal Conflict (11)	.03*	.05*	.01	-.05*	.12*	-.02*	-.07*	-.01	-.01	-.02*	--					
Paternal Conflict (12)	.02*	.06*	-.01	-.05*	.16*	.05*	-.09*	-.02*	.01	-.03*	.45*	--				
School Attachment (13)	-.10*	-.15*	.03*	.16*	-.05*	.01	-.05*	.04*	-.01	.01	-.28*	-.30*	--			
Social Support (14)	-.06*	-.11*	-.01	.09*	-.17*	-.01*	-.01	.03*	-.02*	.01	-.58*	-.56*	.46*	--		
Peer Drug Use (15)	.09*	.23*	-.01	-.17*	.25*	-.10*	.05*	-.01	.02*	-.04*	.16*	.19*	-.29*	-.28*	--	
Baseline Delinquency (16)	.18*	.23*	.01	-.10*	.01	.01	.12*	-.01	.02*	.01	.22*	.21*	-.36*	-.32*	.41*	--
Baseline Drug Use	.11*	.32*	.04*	-.14*	.22*	-.13*	.03*	.01	.04*	-.02*	.21*	.22*	-.32*	-.31*	.61*	.50*

Notes: Pairwise deletion was used to remove cases with missing values on the specified variables. The majority of the correlation coefficients represent Pearson correlations. The correlation coefficients for educational attainment, non-white, male, parent employment status, and parent education point biserial coefficient.
*$p < .05$

6.1.2. Descriptive statistics MZ/DZ subsamples

Table 6.3 presents the descriptive statistics for the variables of interest within the MZ and same sex DZ twins. As illustrated, the average age of the Wave I participants was 16.08, 37 percent of the sample was of Non-White and was 52 percent of the sample was male. Additionally, parental respondents indicated an average household income of 47 thousand dollars a year (Parent Income: $\bar{X}$ = 46.76; SD = 50.22), 66 percent said that they were employed full time, and 49 percent said that they completed at least some college. The participants on average reported a low level of maternal ($\bar{X}$ = -.08; SD = 1.19) and paternal conflict ($\bar{X}$ = -.11; SD = 1.25; Min, Max: Maternal Conflict: -1.35, 6.36; Paternal Conflict: -1.37, 6.12), and a high level of school attachment ($\bar{X}$ = .02; SD = 1.06; Min, Max: -3.89, 2.06). The descriptive statistics indicated that the majority of the sample had high levels of social support ($\bar{X}$ = 8.07; SD = 1.15), had approximately one close friend use drugs at the baseline ($\bar{X}$ = .85; SD = .90), had engaged in low levels of delinquent behaviors at the baseline ($\bar{X}$ = .53; SD = .68), and had reported using low levels drugs at the baseline ($\bar{X}$ = .51; SD = .45).

In terms of the treatment conditions, the average score on the Peabody Picture Vocabulary Test at Wave III was approximately 97.12, and the majority of the sample had completed at least 1 year of college prior to the interview (educational attainment: $\bar{X}$ = .53). At Wave IV, the participants on average reported low levels of delinquent activity ($\bar{X}$ = .06; SD = .20) and low levels of drug use on the standardized scale ($\bar{X}$ = -.03; SD = 1.30; min, max: -1.60, 3.45).

Table 6.3: Descriptive statistics for the MZ and same sex DZ twins subsample.

	N	$\bar{X}$ (%)	SD	Min,Max
Dependent Variables (Wave IV)				
Delinquency	888	.06	.20	.00,2.80
Drug Use	876	-.03	1.30	-1.60,3.45
Treatment Conditions (Wave III)				
Intelligence	839	97.12	15.79	7,122
Educational Attainment	871	53%	.50	0,1
Covariates (Wave I)				
Age	1,060	16.08	1.63	12,20
Non-White	1,060	37%	.48	0,1
Male	1,060	52%	.50	0,1
Parent Income	803	46.76	50.22	0,800
Parent Employment Status	912	66%	.48	0,1
Parent Education	904	49%	.50	0,1
Maternal Conflict	980	-.08	1.19	-1.35,6.36
Paternal Conflict	738	-.11	1.25	-1.37,6.12
School Attachment	1,040	.02	1.06	-3.89,2.06
Social Support	1,038	8.07	1.15	3,10
Peer Drug Use	1,023	.85	.90	0,3
Baseline Delinquency	1,047	.53	.68	0,6
Baseline Drug Use	1,038	.51	.45	0,2

Notes: Drug Use, Maternal Conflict, Paternal Conflict, and School attachment were standardized to account for the differences in coding schemes between items.

Table 6.4 presents the descriptive statistics for the variables of interest within the MZ and different sex DZ twins. As illustrated, the average age of the Wave I participants was 16.07, 38 percent of the sample was of Non-White and was 51 percent of the sample was male. Additionally, parental respondents indicated an average household income of 46 thousand dollars a year (Parent Income: $\bar{X}$ = 45.90; SD = 48.16), 64 percent said that they were employed full time, and 49 percent said that they completed at least some college. The participants on average reported a low level of maternal ($\bar{X}$ = -.09; SD = 1.17) and paternal conflict ($\bar{X}$ = -.08; SD = 1.31; Min, Max: Maternal Conflict: -1.35, 6.36; Paternal Conflict: -1.37, 6.12), and a high level of school attachment ($\bar{X}$ = .01; SD =1.08; Min, Max: -4.30, 2.06). The descriptive statistics indicated that the majority of the sample had high levels of social support ($\bar{X}$ = 8.06; SD =1.15), had approximately one close friend use drugs at the baseline ($\bar{X}$ = .82; SD =.88), had engaged in low levels of delinquent behaviors at the baseline ($\bar{X}$ = .51; SD =.67), and had reported using low levels drugs at the baseline ($\bar{X}$ = .49; SD =.45).

In terms of the treatment conditions, the average score on the Peabody Picture Vocabulary Test at Wave III was approximately 97.08, and the majority of the sample had completed at least 1 year of college prior to the interview (educational attainment: $\bar{X}$ = .53). At Wave IV, the participants on average reported low levels of delinquent activity ($\bar{X}$ = .06; SD = .20) and low levels of drug use on the standardized scale ($\bar{X}$ = -.02; SD = 1.27; min, max: -1.60, 3.45).

Table 6.4: Descriptive statistics for the MZ and different sex DZ twins subsample.

	N	$\bar{X}$ (%)	SD	Min,Max
Dependent Variables (Wave IV)				
Delinquency	1,229	.06	.20	.00,2.80
Drug Use	1,216	-.02	1.27	-1.60,3.45
Treatment Conditions (Wave III)				
Intelligence	1,144	97.08	15.79	7,122
Educational Attainment	1,184	53%	.50	0,1
Covariates (Wave I)				
Age	1,461	16.07	1.62	12,20
Non-White	1,460	38%	.49	0,1
Male	1,461	51%	.50	0,1
Parent Income	1,117	45.90	48.16	0,800
Parent Employment Status	1,262	64%	.48	0,1
Parent Education	1,258	49%	.50	0,1
Maternal Conflict	1,361	-.09	1.17	-1.35,6.36
Paternal Conflict	1,007	-.08	1.31	-1.37,6.12
School Attachment	1,431	.01	1.08	-4.30,2.06
Social Support	1,426	8.06	1.15	3,10
Peer Drug Use	1,409	.82	.88	0,3
Baseline Delinquency	1,443	.51	.67	0,6
Baseline Drug Use	1,431	.49	.45	0,2

Notes: Drug Use, Maternal Conflict, Paternal Conflict, and School attachment were standardized to account for the differences in coding schemes between items.

Table 6.5 provides the correlation matrix for the variables of interest within the MZ and same sex DZ twins subsample. As illustrated, delinquency (Wave IV) was associated with educational attainment (Wave III; $r = -.08$, $p < .05$), but not intelligence (Wave III; $r = .01$, $p > .05$). Furthermore, drug use (Wave IV) was positively associated with intelligence (Wave III; $r = .11$, $p < .05$) and not associated with educational attainment (Wave III; $r = -.03$, $p > .05$). Additionally, intelligence was associated with five of the covariates (Non-White: $r = -.32$; Male: $r = .10$; Parental Income: $r = .22$; Parent Education: $r = .23$; School attachment: $r = .08$; all associations were $p < .05$). Furthermore, educational attainment was associated with eight of the Wave I covariates (Age: $r = .07$; Non-White: $r = -.14$; Parental Income: $r = .26$; Parent Education: $r = .25$; School attachment: $r = .11$; Peer Drug Use: $r = -.08$; Baseline Delinquency: $r = -.08$; Baseline Drug Use: $r = -.09$, all associations were $p < .05$).

Table 6.5. Correlation matrix for the MZ and same sex DZ twins subsample.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Dependent Variables (Wave IV)																
Delinquency (1)	--															
Drug Use (2)	.29*	--														
Treatment Conditions (Wave III)																
Intelligence (3)	.01	.11*	--													
Educational Attainment (4)	-.08*	-.03	.40*	--												
Covariates (Wave I)																
Age (5)	-.06	-.09*	.01	.07*	--											
Non-White (6)	.07*	-.09*	-.32*	-.14*	.08*	--										
Male (7)	.12*	.21*	.10*	-.04	.05	-.04	--									
Parent Income (8)	-.03	-.01	.22*	.26*	.03	-.19*	-.07	--								
Parent Employment Status (9)	-.04	.02	.04	.01	.02	-.05	-.02	-.02	--							
Parent Education (10)	-.02	.07*	.23*	.25*	-.01	-.08*	.09*	.23*	.13*	--						
Maternal Conflict (11)	-.03	.05	-.01	-.01	.09*	.01	-.04	-.03	.05	.02	--					
Paternal Conflict (12)	.01	.01	-.03	-.04	.04	.05	-.07	.01	.05	-.08*	.52*	--				
School Attachment (13)	-.08*	-.15*	.08*	.11*	-.01	-.01	-.04	.04	-.01	-.01	-.22*	-.27*	--			
Social Support (14)	-.03	-.10*	.02	.03	-.10*	.01	-.02	.03	-.04	-.01	-.52*	-.55*	.45*	--		
Peer Drug Use (15)	.01	.21*	-.04	-.08*	.23*	-.10*	.07*	.08*	.08*	.07*	.13*	.13*	-.23*	-.20*	--	
Baseline Delinquency (16)	.11*	.24*	.01	-.08*	-.07*	.01	.13*	-.06	-.02	.11*	.18*	.16*	-.33*	-.26*	.40*	--
Baseline Drug Use	.07*	.30*	.03	-.09*	.22*	-.08*	.10*	.02	.06	.06	.19*	.18*	-.27*	-.27*	.59*	.46*

Notes: Pairwise deletion was used to remove cases with missing values on the specified variables. The majority of the correlation coefficients represent Pearson correlations. The correlation coefficients for educational attainment, non-white, male, parent employment status, and parent education point biserial coefficient.
*$p < .05$

Table 6.6 provides the correlation matrix for the variables of interest within the MZ and different sex DZ twins subsample. As illustrated, delinquency (Wave IV) was associated with educational attainment (Wave III; r = -.10, p < .05), but not intelligence (Wave III; r = -.01, p > .05). Furthermore, drug use (Wave IV) was positively associated with intelligence (Wave III; r = .10, p < .05) and negatively associated with educational attainment (Wave III; r = -.07, p < .05). Additionally, intelligence was associated with four of the covariates (Non-White: r = -.34; Male: r = .07; Parental Income: r = .24; Parent Education: r = .25; all associations were p < .05). Furthermore, educational attainment was associated with eight of the Wave I covariates (Non-White: r = -.15; Male: r = -.08; Parental Income: r = .26; Parent Education: r = .26; School attachment: r = .13; Peer Drug Use: r = -.10; Baseline Delinquency: r = -.12; Baseline Drug Use: r = -.11, all associations were p < .05).

Table 6.6. Correlation matrix for the MZ and different sex DZ twin subsample.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Dependent Variables (Wave IV)																
Delinquency (1)	--															
Drug Use (2)	.29*	--														
Treatment Conditions (Wave III)																
Intelligence (3)	-.01	.10*	--													
Educational Attainment (4)	-.10*	-.07*	.40*	--												
Covariates (Wave I)																
Age (5)	-.05	-.05	.06	.06	--											
Non-White (6)	.05	-.10*	-.34*	-.15*	.04	--										
Male (7)	.12*	.20*	.07*	-.08*	.04	-.03	--									
Parent Income (8)	-.04	-.01	.24*	.26*	.03	-.23*	-.04	--								
Parent Employment Status (9)	-.02	.03	.02	.02	.02	-.02	-.02	-.01	--							
Parent Education (10)	-.02	.08*	.25*	.26*	-.02	-.13*	.07*	.28*	.13*	--						
Maternal Conflict (11)	-.02	.04	-.01	-.03	.09	-.01	-.04	-.03	.03	.03	--					
Paternal Conflict (12)	-.01	.06	-.01	-.06	.10	.07*	-.08*	.01	.03	-.02	.48*	--				
School Attachment (13)	-.08*	-.16*	.05	.13*	.01	.01	-.05	.04	-.03	-.06*	-.25*	-.28*	--			
Social Support (14)	-.04	-.11*	-.01	.05	-.13*	.01	-.03	.03	-.04	-.03	-.53*	-.57*	.48*	--		
Peer Drug Use (15)	.02	.21*	-.04	-.10*	.25*	-.10*	.07*	.06*	.07*	.03	.15*	.17*	-.22*	-.22*	--	
Baseline Delinquency (16)	.15*	.23*	.01	-.12*	-.04	-.01	.12*	-.06*	-.01	.07*	.18*	.18*	-.31*	-.28*	.41*	--
Baseline Drug Use	.10*	.28*	.04	-.11*	.25*	-.10*	.08*	.01	.05	.04	.21*	.20*	-.25*	-.28*	.61*	.47*

Notes: Pairwise deletion was used to remove cases with missing values on the specified variables. The majority of the correlation coefficients represent Pearson correlations. The correlation coefficients for educational attainment, non-white, male, parent employment status, and parent education point biserial coefficient.
*$p < .05$

6.1.3. Baseline multivariate regression models (DV: Antisocial Behavior)

6.1.3.1. Delinquency and Drug Use Regressed on Educational Attainment

Table 6.7 presents the results of the bivariate and multivariate regressions models of delinquency (Wave IV) and drug use (Wave IV) on educational attainment (Wave III) and the covariates. Consistent with the correlation matrix, educational attainment (Wave III) has a statistically significant negative bivariate association with both delinquency (Wave IV; b = -.029, SE = .004, β = -.061, 95%CI = -.038, -.021, $p < .05$) and drug use (Wave IV; b = -.175, SE = .022, β = -.070, 95%CI = -.218, -.132, $p < .05$). Nevertheless, when the covariates are introduced into the model, the associations between educational attainment (Wave III) and delinquency (Wave IV; b = -.011, SE = .006, β = -.024, 95%CI = -.023, .001, $p > .05$), and educational attainment (Wave III) and drug use (Wave IV; b = -.055, SE = .031, β = -.022, 95%CI = -.116, -.007, $p > .05$) were attenuated. The multivariate associations between educational attainment (Wave III) and delinquency (Wave IV), and educational attainment (Wave III) and drug use (Wave IV) did not approach statistical significance.

Table 6.7: Predicting delinquency (Wave IV) and drug use (Wave IV) with educational attainment (Wave III) and covariates.

	DV: Delinquency (Wave IV)				DV: Delinquency (Wave IV)				DV: Drug Use (Wave IV)				DV: Drug Use (Wave IV)			
	b	SE	β	95%CI	b	SE	β	95%CI	b	SE	β	95%CI	b	SE	β	95%CI
Treatment Conditions (Wave III)																
Educational Attainment	-.029*	.004	-.061	-.038,-.021	-.011	.006	-.024	-.023,.001	-.175*	.022	-.070	-.218,-.132	-.055	.031	-.022	-.116,.007
Covariates (Wave I)																
Age	--	--	--	--	-.012*	.001	-.087	-.015,-.008	--	--	--	--	-.142*	.009	-.196	-.159,-.124
Non-White	--	--	--	--	.012	.006	.023	-.001,.024	--	--	--	--	-.232*	.032	-.084	-.295,-.168
Male	--	--	--	--	.055*	.006	.121	.044,.066	--	--	--	--	.375*	.029	.152	.318,.431
Parent Income	--	--	--	--	-.001	.001	-.024	-.001,.001	--	--	--	--	.001*	.001	.036	.001,.001
Parent Employment Status	--	--	--	--	-.013*	.006	-.027	-.024,-.001	--	--	--	--	.040	.029	.016	-.017,.098
Parent Education	--	--	--	--	.016*	.006	.035	.004,.027	--	--	--	--	.071*	.030	.029	.012,.131
Maternal Conflict	--	--	--	--	.003	.003	.017	-.003,.009	--	--	--	--	-.016	.015	-.015	-.044,.013
Paternal Conflict	--	--	--	--	-.001	.003	-.002	-.005,.005	--	--	--	--	.003	.013	.003	-.023,.028
School Attachment	--	--	--	--	-.008*	.003	-.035	-.014,-.001	--	--	--	--	-.046*	.016	-.038	-.077,-.014
Social Support	--	--	--	--	.004	.004	.020	-.003,.011	--	--	--	--	-.044*	.019	-.040	-.082,-.007
Peer Drug Use	--	--	--	--	.012*	.004	.044	.004,.020	--	--	--	--	.090*	.022	.061	.047,.133
Baseline Delinquency	--	--	--	--	.045*	.005	.130	.035,.055	--	--	--	--	--	--	--	--
Baseline Drug Use	--	--	--	--	--	--	--	--	--	--	--	--	.728*	.041	.268	.648,.808
R2	.004*				.052*				.005*				.158*			
N	12,979				6,468				12,898				6,468			

*$p < .05$

6.1.3.2. Delinquency and Drug Use Regressed on Intelligence[51]

Table 6.8 presents the results of the multivariate regression models of delinquency (Wave IV) on intelligence (Wave III) and the covariates (Wave I). Indicated by contemporary scholarship (e.g., Mears and Cochran, 2013), the association between intelligence and antisocial behavior could be linear or curvilinear. Consistent with this knowledge, three separate multivariate regression models were estimated and presented in Table 6.8. Model 1 of Table 6.8 was specified with the assumption that a linear association between intelligence (Wave III) and delinquency (Wave IV) existed, Model 2 of Table 6.8 was specified with the assumption that a quadratic association (one curve) between intelligence (Wave III) and delinquency (Wave IV) existed, and Model 3 of Table 6.8 was specified with the assumption that a cubic association (two curves) between intelligence (Wave III) and delinquency (Wave IV) existed. The results presented in Table 6.8 suggest that intelligence (Wave III) was not associated delinquency at Wave IV. These results were consistent across the linear (b = -.001, SE = .001, β = -.001, 95%CI = -.001, .001, p > .05), quadratic (intelligence: b = .001, SE = .001, β = .016, 95%CI = -.001, .001, p > .05; intelligence^2: b = -.001, SE = .001, β = -.017, 95%CI = -.001, .001, p > .05), and cubic specifications of the association between intelligence and delinquency (intelligence: b = -.002, SE = .001, β = -.120, 95%CI = -.007, .003, p > .05; intelligence^2: b = .001, SE = .001, β = .387, 95%CI = -.001, .001, p > .05; intelligence^3: b = -.001, SE = .001, β = -.274, 95%CI = -.001, .001, p > .05).

[51] Due to the review of the correlation coefficients between the treatment conditions and the dependent variables, the results of the bivariate regression models are provided in Appendix C. The results of bivariate regression models (the standardized point estimate: β) are equal to that of the bivariate correlation coefficient.

Table 6.8: Predicting delinquency (Wave IV) with intelligence and covariates.

DV: Delinquency	Model 1				Model 2				Model 3			
(Wave IV)	b	SE	β	95%CI	b	SE	β	95%CI	b	SE	β	95%CI
Treatment Conditions (Wave III)												
Intelligence	-.001	.001	-.001	-.001,.001	.001	.001	.016	-.001,.001	-.002	.001	-.120	-.007,.003
Intelligence^2	--	--	--	--	-.001	.001	-.017	-.001,.001	.001	.001	.387	-.001,.001
Intelligence^3	--	--	--	--	--	--	--	--	-.001	.001	-.274	-.001,.001
Covariates (Wave I)												
Age	-.012*	.002	-.092	-.016,-.009	-.012*	.002	-.092	-.016,-.009	-.012*	.002	-.092	-.016,-.009
Non-White	.011	.007	.020	-.003,.024	.011	.007	.020	-.003,.024	.011	.007	.020	-.003,.024
Male	.056*	.006	.121	.044,.067	.056*	.006	.121	.044,.067	.056*	.006	.121	.044,.067
Parent Income	-.001*	.001	-.027	-.001,-.001	-.001*	.001	-.027	-.001,-.001	-.001*	.001	-.027	-.001,-.001
Parent Employment Status	-.013*	.006	-.028	-.025,-.002	-.013*	.006	-.028	-.025,-.002	-.013*	.006	-.028	-.025,-.002
Parent Education	.014*	.006	.031	.003,.026	.014*	.006	.031	.003,.026	.014*	.006	.031	.003,.026
Maternal Conflict	.003	.003	.017	-.003,.009	.003	.003	.017	-.003,.009	.003	.003	.017	-.003,.009
Paternal Conflict	-.001	.003	-.002	-.005,.005	-.001	.003	-.002	-.005,.005	-.001	.003	-.002	-.005,.005
School Attachment	-.007*	.003	-.034	-.014,-.001	-.007*	.003	-.034	-.014,-.001	-.008*	.003	-.034	-.014,-.001
Social Support	.003	.004	.015	-.004,.011	.003	.004	.015	-.004,.011	.003	.004	.015	-.004,.011
Peer Drug Use	.014*	.004	.049	.006,.021	.014*	.004	.049	.006,.021	.013*	.004	.049	.006,.021
Baseline Delinquency	.045*	.005	.127	.035,.055	.045*	.005	.127	.035,.055	.045*	.005	.127	.035,.055
R^2			.05*				.05*				.05*	
N			6,269				6,269				6,269	

*$p < .05$

Similarly, Table 6.9 presents the results of the multivariate regression models of drug use (Wave IV) on the linear, quadratic, and cubic specification of intelligence (Wave III, respectively) and the covariates (Wave I). The results of Model 3 in Table 6.9 suggest the existence of a cubic association (i.e., two curves) between intelligence (Wave III) and drug use at Wave IV. To specify, intelligence had a negative association with drug use at Wave IV (b = -.050, SE = .013, β = -.594, 95%CI = -.077, -.024, $p > .05$), intelligence^2 had a positive association with drug use at Wave IV (b = .001, SE = .001, β = 1.619, 95%CI = .001, .001, $p > .05$), and intelligence^3 had a negative association with drug use at Wave IV (b = -.001, SE = .001, β = -1.008, 95%CI = -.001, -.001, $p > .05$). All three specification of intelligence (linear, quadratic, and cubic) were statistically associated with drug use (Wave IV). These results suggest that two curves and three distinct slopes exist in the association between intelligence (Wave III) and drug use at Wave IV.

Table 6.9: Predicting drug use (Wave IV) with intelligence and covariates.

DV: Drug Use	Model 1				Model 2				Model 3			
(Wave IV)	b	SE	β	95%CI	b	SE	β	95%CI	b	SE	β	95%CI
Treatment Conditions (Wave III)												
Intelligence	.003*	.001	.033	.001,.005	-.008	.001	-.094	-.016,.001	-.050*	.013	-.594	-.077,-.024
Intelligence^2	--	--	--	--	.001*	.001	.132	.001,.001	.001*	.001	1.619	.001,.001
Intelligence^3	--	--	--	--	--	--	--	--	-.001*	.001	-1.008	-.001,-.001
Covariates (Wave I)												
Age	-.148*	.009	-.204	-.166,-.130	-.150*	.009	-.207	-.168,-.132	-.149*	.009	-.205	-.166,-.131
Non-White	-.211*	.034	-.075	-.279,-.144	-.208*	.034	-.074	-.276,-.141	-.200*	.034	-.071	-.267,-.132
Male	.372*	.029	.150	.314,.429	.369*	.029	.149	.312,.427	.367*	.029	.148	.309,.424
Parent Income	.001*	.001	.030	.001,.001	.001*	.001	.028	.001,.001	.001*	.001	.028	.001,.001
Parent Employment Status	.049	.030	.019	-.010,.107	.051	.030	.020	-.008,.110	.048	.030	.019	-.011,.107
Parent Education	.044	.030	.018	-.015,.104	.036	.031	.015	-.024,.096	.035	.030	.014	-.035,.095
Maternal Conflict	-.015	.015	-.015	-.044,.015	-.014	.015	-.014	-.044,.015	-.014	.015	-.013	-.043,.016
Paternal Conflict	.005	.013	.005	-.022,.031	.004	.013	.005	-.022,.031	.004	.013	.005	-.022,.031
School Attachment	-.050*	.017	-.042	-.082,-.017	-.050*	.017	-.042	-.082,-.017	-.052*	.017	-.043	-.084,-.019
Social Support	-.040*	.020	-.036	-.079,-.002	-.040*	.020	-.035	-.077,-.001	-.038*	.020	-.034	-.077,-.001
Peer Drug Use	.092*	.022	.063	.049,.136	.094*	.022	.064	.051,.138	.094*	.022	.064	.050,1.37
Baseline Drug Use	.736*	.041	.270	.656,.817	.740*	.041	.272	.660,.822	.732*	.041	.269	.651,.813
R2	.160*				.160*				.160*			
N	6,233				6,233				6,233			

Notes: The large β in model 3 correspond to the extremely small standard errors associated with the quadratic and cubic specification of intelligence.
*$p < .05$

6.1.4. Exploration of social self-selection (DV: Treatment Conditions)

Table 6.10 presents the results of the multivariate regressions models of the treatment conditions (i.e., intelligence and educational attainment at Wave III) on the covariates at Wave I. Concerning educational attainment (Wave III), the findings suggest that enrollment in college was predicted by eight of the covariates. Age (b = .17, SE = .02, OR = 1.19, OR95%CI = 1.15, 1.23, p > .05), parent income (b = .02, SE = .01, OR = 1.02, OR95%CI = 1.01, 1.02, p > .05), parent education (b = .89, SE = .06, OR = 2.43, OR95%CI = 2.18, 2.72, p > .05), school attachment (b = .24, SE = .03, OR = 1.27, OR95%CI = 1.20, 1.35, p > .05), and baseline delinquency (b = .13, SE = .05, OR = 1.14, OR95%CI = 1.03, 1.25, p > .05) were positively associated with educational attainment (Wave III), while male (b = -.38, SE = .05, OR = .68, OR95%CI = .61, .76, p > .05), peer drug use (b = -.32, SE = .04, OR = .73, OR95%CI = .67, .78, p > .05), and baseline drug use (b = -.41, SE = .08, OR = .67, OR95%CI = .57, .78, p > .05) were negatively associated with educational attainment (Wave III).

Regarding intelligence (Wave III), The findings suggest that levels of intelligence (Wave III) were predicted by seven of the covariates. Age (b = .38, SE = .11, β = .04, 95%CI = .17, .59, p > .05), parent income (b = .03, SE = .01, β = .10, 95%CI = .02, .03, p > .05), parent education (b = 5.40, SE = .36, β = .17, 95%CI = 4.44, 5.84, p > .05), and school attachment (b = .79, SE = .20, β = .05, 95%CI = .40, 1.18, p > .05) at Wave I had a positive association with intelligence at Wave III, while non-white (b = -9.26, SE = .39, β = -.27, 95%CI = -10.03, -8.50, p > .05), social support (b = -1.17, SE = .23, β = -.08, 95%CI = -1.61, -.72, p > .05), and peer drug use (b = -1.28, SE = .26, β = -.07, 95%CI = -1.80, -.77, p > .05) had negative associations with the participants' level of intelligence (Wave III).

Table 6.10: Predicting intelligence and educational attainment with the covariates.

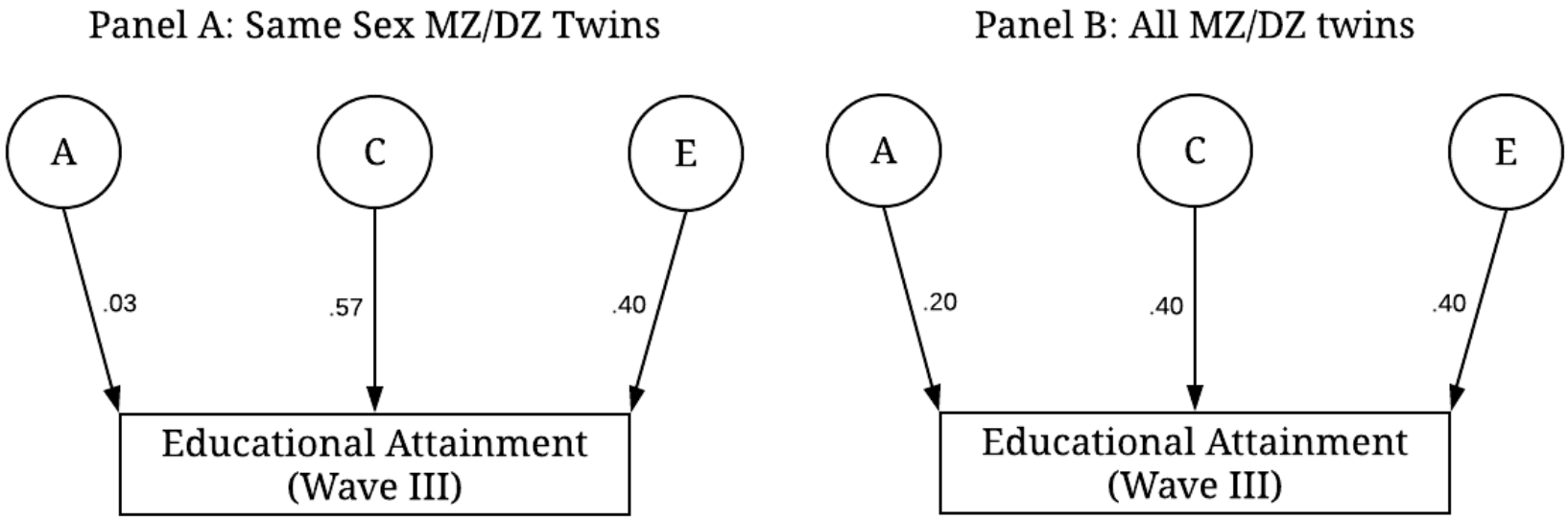
	DV: Intelligence (Wave III)				DV: Educational Attainment (Wave III)			
	b	SE	β	95%CI	b	SE	OR	OR 95%CI
Covariates (Wave I)								
Age	.38*	.11	.04	.17,.59	.17*	.02	1.19	1.15,1.23
Non-White	-9.26*	.39	-.27	-10.03,-8.50	-.10	.06	.91	.81,1.02
Male	.45	.35	.01	-.23,1.14	-.38*	.05	.68	.61,.76
Parent Income	.03*	.01	.10	.02,.03	.02*	.01	1.02	1.01,1.02
Parent Employment Status	-.43	.35	-.01	-1.12,.27	-.03	.05	.97	.87,1.08
Parent Education	5.40*	.36	.17	4.44,5.84	.89*	.06	2.43	2.18,2.72
Maternal Conflict	-.25	.18	-.02	-.60,.10	-.02	.03	.98	.93,1.03
Paternal Conflict	.08	.16	.01	-.23,.39	-.01	.02	.99	.95,1.04
School Attachment	.79*	.20	.05	.40,1.18	.24*	.03	1.27	1.20,1.35
Social Support	-1.17*	.23	-.08	-1.61,-.72	-.05	.03	.96	.89,1.02
Peer Drug Use	-1.28*	.26	-.07	-1.80,-.77	-.32*	.04	.73	.67,.78
Baseline Delinquency	.41	.33	.02	-.23,1.04	.13*	.05	1.14	1.03,1.25
Baseline Drug Use	.13	.52	.01	-.88,1.15	-.41*	.08	.67	.57,.78
R^2	.12*				.22*			
N	7,126				7,356			

Notes: R^2 for Educational Attainment (Wave III) model represents *Nagelkerke* R^2
*$p < .05$

6.1.5. Exploration of genetic self-selection (DV: Treatment Conditions)

Figure 6.1 provides the results of the ACE decomposition model for educational attainment (Wave III) with the same sex MZ/DZ subsample and the different sex MZ/DZ subsample. The results of the same sex MZ/DZ subsample suggests that 34 percent of the variance in educational attainment (Wave III) is accounted for by genetic factors (*a*), 31 percent of the variance in educational attainment (Wave III) is accounted for by the shared environment (*c*), and 36 percent of the variance in educational attainment (Wave III) is accounted for by the non shared environment (*e*). The results of the different sex MZ/DZ subsample suggests that 63 percent of the variance in educational attainment (Wave III) is accounted for by genetic factors (*a*), 0 percent of the variance in educational attainment (Wave III) is accounted for by the shared environment (*c*), and 37 percent of the variance in educational attainment (Wave III) is accounted for by the non shared environment (*e*).

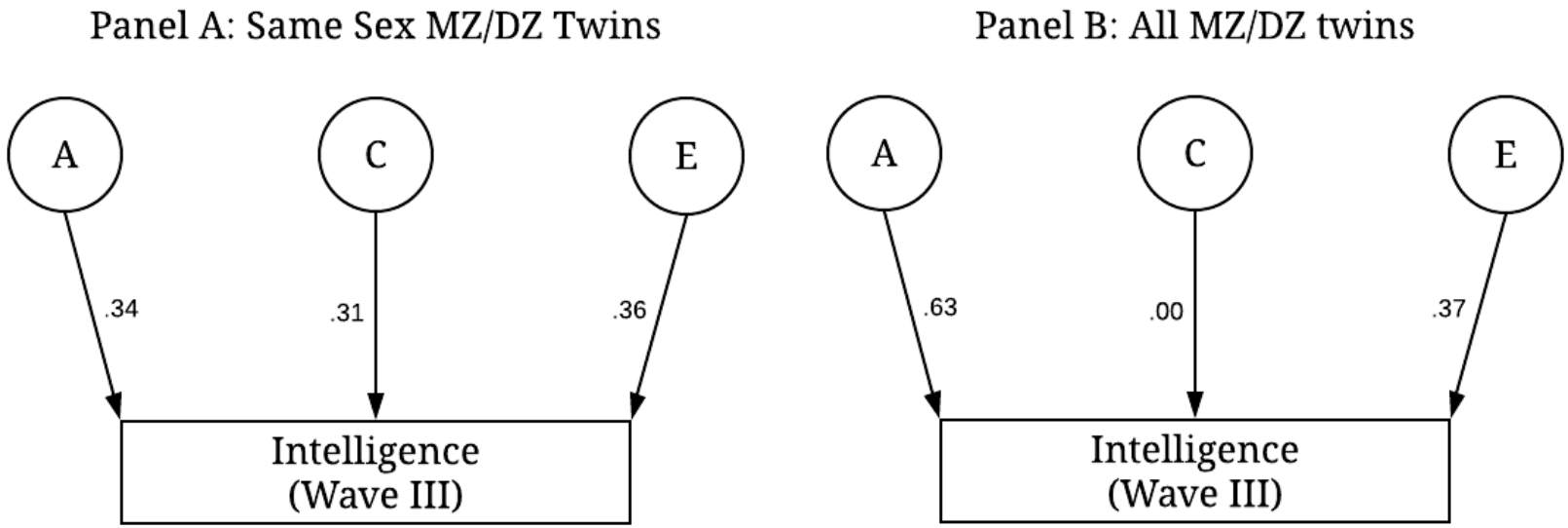
Figure 6.1: ACE Model Predicting Educational Attainment at Wave III



Notes: "A" represents the variance in intelligence predicted by the additive genetic component. "C" represents the variance in intelligence predicted by the shared environment. "E" represents the variance in intelligence predicted by the non shared environment.

Figure 6.2 provides the results of the ACE decomposition model for intelligence (Wave III) with the same sex MZ/DZ subsample and the different sex MZ/DZ subsample. The results of the same sex MZ/DZ subsample suggests that 34 percent of the variance in intelligence (Wave III) is accounted for by genetic factors (*a*), 31 percent of the variance in intelligence (Wave III) is accounted for by the shared environment (*c*), and 36 percent of the variance in intelligence (Wave III) is accounted for by the non shared environment (*e*). The results of the different sex MZ/DZ subsample suggests that 63 percent of the variance in intelligence (Wave III) is accounted for by genetic factors (*a*), 0 percent of the variance in intelligence (Wave III) is accounted for by the shared environment (*c*), and 37 percent of the variance in intelligence (Wave III) is accounted for by the non shared environment (*e*).

Figure 6.2: ACE Model Predicting Intelligence at Wave III



Notes: "A" represents the variance in intelligence predicted by the additive genetic component. "C" represents the variance in intelligence predicted by the shared environment. "E" represents the variance in intelligence predicted by the non shared environment.

6.2. Study 2: Adjusting for Social Self-Selection with Propensity Score Matching

6.2.1. Post-matching balancing statistics for educational attainment

Due to the consistency regarding balance between the treatment (i.e., educational attainment = 1) and control groups (i.e., educational attainment = 0), post-matching descriptive statistics, and the post-matching bivariate associations between antisocial behavior and educational attainment, only the results pertaining to the cases matched with nearest neighbor matching with a caliper of .005 are reviewed. The post-matching statistics associated cases matched with nearest neighbor matching with a caliper of .05, .01, .001, and .0001 are presented in Appendix E and provide findings similar to that reviewed in the primary text.

Table 6.11 provides the post-matching covariate balance between the treatment and control case on educational attainment (caliper = .005). Evident by the results, the assumption of balance for the 13 covariates was satisfied when matching treatment and control participants with nearest neighbor matching ant a caliper equal to .005. While not statistically significant, the largest mean differences on percent bias were observed for age (Control: $\bar{X}$ = 15.920, Treatment: $\bar{X}$ = 15.862, % Bias = -3.503, $t_{df = 4332}$ = -1.108, $p < .05$), peer drug use (Control: $\bar{X}$ = .785, Treatment: $\bar{X}$ = .812, % Bias = 3.340, $t_{df = 4332}$ = 1.034, $p < .05$), and baseline delinquency (Control: $\bar{X}$ = .517, Treatment: $\bar{X}$ = .533, % Bias = 2.739, $t_{df = 4332}$ = .855, $p < .05$).

Table 6.11: Balance statistics for participants matched with nearest neighbor matching (caliper = .005) on educational attainment (Wave III).

DV: Educational Attainment (Wave III)	Did Not Complete One Year of College (c)	Completed One Year of College (t)	% Bias	*t-value*
	$\bar{X}$	$\bar{X}$		
Covariates (Wave I)				
Age	15.920	15.862	-3.503	-1.108
Non-White	.288	.292	.932	.301
Male	.505	.513	1.484	.486
Parent Income	45.280	46.734	2.331	1.387
Parent Employment Status	.604	.599	-1.049	-.341
Parent Education	.389	.383	-1.216	-.406
Maternal Conflict	-.055	-.081	-2.194	-.710
Paternal Conflict	-.022	-.056	-2.566	-.824
School Attachment	.056	.054	-.159	-.050
Social Support	8.077	8.079	.250	.079
Peer Drug Use	.785	.812	3.340	1.034
Baseline Delinquency	.517	.533	2.739	.855
Baseline Drug Use	.510	.512	.455	.147
N	2,167	2,167		

Notes: Caliper for the nearest neighbor matching was set at $p < .005$. (c) signifies the control cases and (t) signifies the treatment cases.
** $p < .05$*

Table 6.12 presents the descriptive statistics for the pre-matching sample, the post-matching subsample, and a mean difference comparison (i.e., independent samples t-test) between the two samples on the dependents variables, educational attainment, and the 13 covariates. The mean values for the post-matching subsample differed from the mean value for the pre- matching sample on seven of the 16 variables. Specifically, it was observed that the post-matching subsample had lower average scores on educational attainment (Post-matching: $\bar{X}$ = .500, Pre-matching: $\bar{X}$ = .620, $t_{\text{df} = 9147}$ = -12.246, $p < .05$), parent income (Post-matching: $\bar{X}$ = 46.007, Pre-matching: $\bar{X}$ = 55.084, $t_{\text{df} = 10404}$ = -10.340, $p < .05$), and parent education (Post-matching: $\bar{X}$ = .386, Pre-matching: $\bar{X}$ = .480, $t_{\text{df} = 9299}$ = -9.662, $p < .05$) than the pre-matching sample. Furthermore, the post-matching subsample had higher average scores on non-White (Post-matching: $\bar{X}$ = .290, Pre-matching: $\bar{X}$ = .266, $t_{\text{df} = 9168}$ = 2.662, $p < .05$), male (Post-matching: $\bar{X}$ = .509, Pre-matching: $\bar{X}$ = .467, $t_{\text{df} = 9311}$ = 4.219, $p < .05$), peer drug use (Post-matching: $\bar{X}$ = .798, Pre-matching: $\bar{X}$ = .759, $t_{\text{df} = 9311}$ = 2.306, $p < .05$), and baseline drug use (Post-matching: $\bar{X}$ = .511, Pre-matching: $\bar{X}$ = .490, $t_{\text{df} = 9404}$ = 2.392, $p < .05$) than the pre-matching sample

Table 6.12: Descriptive statistics for the matched sample (caliper = .005).

	Pre-matching Sample	Post-matching Sample	*t-value*
	$\bar{X}$	$\bar{X}$	
Dependent Variables (Wave IV)			
Delinquency	.065	.071	1.292
Drug Use	.029	.056	1.037
Treatment Conditions (Wave III)			
Educational Attainment	.620	.500	-12.246*
Covariates (Wave I)			
Age	15.918	15.891	-.802
Non-White	.266	.290	2.662*
Male	.467	.509	4.219*
Parent Income	55.084	46.007	-10.340*
Parent Employment Status	.611	.602	-.956
Parent Education	.480	.386	-9.662*
Maternal Conflict	-.069	-.068	.049
Paternal Conflict	-.056	-.039	.625
School Attachment	.093	.055	-1.89
Social Support	8.089	8.078	-.485
Peer Drug Use	.759	.798	2.306*
Baseline Delinquency	.521	.525	.292
Baseline Drug Use	.490	.511	2.392*
N	6,202	4,334	

Notes: Caliper for the nearest neighbor matching was set at $p < .005$. Pre-matching sample designates the cases remaining after listwise deletion for the model. The sample size for delinquency (Wave IV) was 3,742 and the sample size for Drug use (Wave IV) was 3,728 on the matched sample.
* $p < .05$

6.2.2. Post-matching balancing statistics for intelligence

In review, due to the continuous nature of Intelligence (Wave III) generalized propensity score matching (GPS) was used to create a subsample of cases in which the bivariate association between antisocial behavior and educational attainment could be estimated unconfounded by the 13 covariates. Table 6.13 presents the post-matching balancing results of the GPS matches. The $\bar{X}\Delta$ represent the mean difference between the specified percentile with a specified GPS score and the cases outside of the specified percentile with the same GPS score. All of the mean differences were than averaged to create an average mean difference, which is then used to estimate an adjusted *t*-statistic. As such, the assessment of balance for each covariate comparison for the post-matching sample must be improved beyond statistical significance (i.e., *t-value* must be below 1.96). Due to the difficulties of satisfying this criterion, scholars often evaluate the

balance for each covariate comparison as the percent reduction in bias from the pre-matching sample to the post-matching sample (Bia and Mattei, 2008; Mears and Cochran, 2013).

Table 6.13 presents the balancing statistics for the pre-matching sample and the post-matching sample derived from the generalized propensity score matching. As suggested by the overall evaluation of balance, the matches derived from the GPS analysis had substantive deviations from balance for the 13 covariates beyond the $p < .01$ level. Notably, although the overall evaluation of balance suggested substantive deviations from balance, the percent reduction in bias analyses suggested that the post-matching sample consistently reduced bias across the majority of the comparisons. For a full interpretation of the results please see Appendix G.

Table 6.13. Balancing statistics for the generalized propensity score match.

DV: Intelligence (Range per Percentile)	Pre-matching		Post- matching		% Reduction in Bias
	$\bar{X}\Delta$	*t-value*	$\bar{X}\Delta$	*t-value*	
10th Percentile (7,79)					
Age	-.147	-2.989*	-.079	-2.334*	-46.353
Non-White	-.358	-28.131*	-.022	-5.093*	-93.858
Male	.042	3.122*	.004	.412	-90.588
Parent Income	17.240	10.057*	3.943	3.340*	-77.129
Parent Employment Status	.058	3.847*	.006	.640	-89.646
Parent Education	.201	14.847*	.043	4.422*	-78.622
Maternal Conflict	.022	.623	-.021	-.867	-5.628
Paternal Conflict	-.029	-.608	-.049	-1.826	71.305
School Attachment	.027	.885	-.018	-.881	-32.771
Social Support	-.081	-2.226*	-.003	-.133	-96.291
Peer Drug Use	.094	3.935*	.043	2.639*	-54.316
Baseline Delinquency	.027	1.429	.010	.767	-63.451
Baseline Drug Use	.111	9.032*	.028	3.122*	-74.761
20th Percentile (80,87)					
Age	.129	2.787*	.025	.916	-80.680
Non-White	-.196	-15.114*	.004	1.269	-97.961
Male	.062	4.829*	.011	1.356	-82.373
Parent Income	14.068	12.251*	3.243	3.416*	-76.948
Parent Employment Status	.018	1.261	.003	.362	-82.933
Parent Education	.174	13.352*	.042	5.517*	-75.926
Maternal Conflict	.094	2.737*	.014	.711	-85.078
Paternal Conflict	.070	1.578	.034	1.577	-51.500
School Attachment	.054	1.868	.017	1.088	-68.611
Social Support	-.082	-2.570*	-.002	-.140	-97.570
Peer Drug Use	-.045	-1.962*	-.006	-.474	-86.791
Baseline Delinquency	.036	2.063*	.009	.838	-74.844
Baseline Drug Use	.022	1.871	.003	.423	-86.255
30th Percentile (88,91)					
Age	-.069	-1.502	.005	.209	-92.745
Non-White	-.103	-7.701*	-.002	-.709	-98.050
Male	.002	.182	.003	.460	23.381
Parent Income	8.407	7.344*	2.145	2.509*	-74.487
Parent Employment Status	-.013	-.932	-.003	-.415	-77.040
Parent Education	.085	6.057*	.008	1.261	-90.608
Maternal Conflict	-.035	-.940	-.022	-1.218	-36.871
Paternal Conflict	-.027	-.576	-.033	-1.677	23.425
School Attachment	.118	3.849*	.033	2.199*	-72.074
Social Support	.053	1.603	.039	2.403*	-26.184
Peer Drug Use	-.145	-5.905*	-.039	-3.236*	-73.131
Baseline Delinquency	-.021	-1.161	-.007	-.784	-67.433
Baseline Drug Use	-.053	-4.224*	-.025	-3.763*	-52.443
40th Percentile (92,96)					
Age	.035	.601	.059	2.150*	70.127
Non-White	-.040	-2.589*	.010	2.533*	-74.977
Male	-.013	-.844	0.00	.554	-69.825
Parent Income	9.081	8.052*	2.833	3.014*	-68.804
Parent Employment Status	.003	.173	-.001	-.016	-64.868
Parent Education	.099	6.109*	.017	2.509*	-82.781
Maternal Conflict	.060	1.461	.034	1.707	-43.345
Paternal Conflict	.006	.127	-.020	-.918	209.265
School Attachment	.062	1.813	.002	.108	-96.787

Table 6.13. Balancing statistics for the generalized propensity score match.

DV: Intelligence (Range per Percentile)	Pre-matching		Post- matching		% Reduction in Bias
	$\bar{X}\Delta$	*t-value*	$\bar{X}\Delta$	*t-value*	
Social Support	.011	.284	-.012	-.678	10.436
Peer Drug Use	-.042	-1.449	-.011	-.844	-73.759
Baseline Delinquency	-.019	-.825	.001	.121	-94.820
Baseline Drug Use	-.023	-1.591	-.012	-1.569	-48.848
50th Percentile (97,103)					
Age	.188	4.391*	.087	4.259*	-53.812
Non-White	.073	6.362*	.004	.864	-94.519
Male	-.009	-.771	.004	.741	-57.885
Parent Income	2.690	2.081*	.957	1.380	-64.429
Parent Employment Status	-.014	-1.126	-.003	-.490	-78.946
Parent Education	.047	3.655*	.010	1.869	-78.804
Maternal Conflict	.001	.041	.007	.451	427.321
Paternal Conflict	-.016	-.400	.010	.638	-38.065
School Attachment	-.009	-.321	-.010	-.813	17.196
Social Support	.065	2.195*	.011	.829	-83.187
Peer Drug Use	-.060	-2.663*	-.026	-2.605*	-56.729
Baseline Delinquency	-.039	-2.245*	-.016	-2.042*	-59.350
Baseline Drug Use	-.052	-4.420*	-.021	-3.762*	-59.891
60th Percentile (104,106)					
Age	.122	2.781*	.018	.870	-85.203
Non-White	.142	12.592*	.015	3.674*	-89.438
Male	-.057	-4.423*	-.011	-1.887	-80.654
Parent Income	-2.774	-2.558*	.324	.474	-88.321
Parent Employment Status	-.015	-1.121	-.003	-.538	-79.796
Parent Education	-.054	-3.961*	-.003	-.633	-94.483
Maternal Conflict	-.024	-.724	-.014	-.970	-42.399
Paternal Conflict	.044	1.103	-.001	-.004	-97.706
School Attachment	-.063	-2.374*	-.007	-.589	-88.890
Social Support	.006	.188	-.007	-.554	23.982
Peer Drug Use	.018	.802	.005	.501	-72.419
Baseline Delinquency	-.016	-.897	-.006	-.712	-62.595
Baseline Drug Use	-.031	-2.473*	-.006	-1.106	-80.407
70th Percentile (107,108)					
Age	.083	1.730	.018	.786	-78.300
Non-White	.122	9.900*	.012	2.612*	-90.182
Male	-.011	-.803	-.006	-.989	-46.161
Parent Income	-6.774	-4.115*	-1.254	-1.701	-81.487
Parent Employment Status	-.009	-.653	.003	.395	-68.127
Parent Education	-.086	-5.773*	-.002	-.407	-97.678
Maternal Conflict	-.040	-1.096	-.005	-.332	-87.583
Paternal Conflict	.044	1.027	.028	1.606	-36.386
School Attachment	-.036	-1.232	.002	.164	-94.479
Social Support	-.023	-.701	-.011	-.783	-52.076
Peer Drug Use	.059	2.461*	-.001	.002	-98.310
Baseline Delinquency	-.015	-.801	-.013	-1.543	-15.142
Baseline Drug Use	.019	1.525	.005	.782	-74.330
80th Percentile (109,111)					
Age	-.031	-.511	-.027	-.967	-11.620
Non-White	.079	5.018*	.001	.235	-98.730
Male	.013	.741	.008	.979	-36.580
Parent Income	-5.477	-3.494*	-.567	-.650	-89.648
Parent Employment Status	-.054	-3.132*	-.018	-2.285*	-66.465
Parent Education	-.119	-6.548*	-.009	-1.441	-92.416
Maternal Conflict	-.035	-.778	.006	.290	-82.839
Paternal Conflict	-.006	-.110	-.004	-.189	-32.539

Table 6.13. Balancing statistics for the generalized propensity score match.

DV: Intelligence (Range per Percentile)	Pre-matching		Post- matching		% Reduction in Bias
	$\bar{X}\Delta$	*t-value*	$\bar{X}\Delta$	*t-value*	
School Attachment	-.074	-2.060*	-.011	-.645	-85.202
Social Support	-.006	-.164	-.011	-.618	70.226
Peer Drug Use	.007	.219	.008	.570	16.903
Baseline Delinquency	.004	.192	.011	1.051	153.450
Baseline Drug Use	-.025	-1.540	-.004	-.553	-83.987
90th Percentile (112,116)					
Age	-.182	-4.157*	-.079	-3.619*	-56.500
Non-White	.102	8.560*	-.006	-1.210	-94.107
Male	.013	.994	.009	1.368	-30.538
Parent Income	-11.567	-5.863*	-3.169	-5.093*	-72.604
Parent Employment Status	.004	.294	.001	.181	-74.975
Parent Education	-.150	-10.853*	-.023	-4.118*	-84.682
Maternal Conflict	.002	.053	-.014	-.923	690.203
Paternal Conflict	-.087	-2.074*	-.045	-2.620*	-48.251
School Attachment	-.056	-2.030*	.011	.840	-80.325
Social Support	.007	.248	.014	1.007	93.960
Peer Drug Use	.040	1.741	.001	.019	-97.516
Baseline Delinquency	.012	.684	.006	.696	-50.481
Baseline Drug Use	-.001	-.073	.010	1.739	977.470
99th Percentile (117,122)					
Age	-.214	-4.286*	-.040	-1.590	-81.272
Non-White	.204	17.262*	.011	1.715	-94.604
Male	-.045	-3.040*	-.006	-.857	-86.681
Parent Income	-23.696	-9.430*	-2.995	-5.079*	-87.361
Parent Employment Status	.015	.991	.015	2.070*	-1.091
Parent Education	-.230	-15.348*	-.013	-2.188*	-94.347
Maternal Conflict	-.056	-1.431	-.003	-.142	-94.610
Paternal Conflict	.001	.031	-.004	-.216	199.531
School Attachment	-.034	-1.106	.001	.072	-97.037
Social Support	.043	1.359	.022	1.388	-48.609
Peer Drug Use	.103	4.119*	.023	1.825	-77.714
Baseline Delinquency	.043	2.3088*	.013	1.399	-69.800
Baseline Drug Use	.045	3.196	.019	2.787*	-57.916

* $p < .05$

6.2.3. Post-matching bivariate associations between the dependent variables and the treatment conditions.

Table 6.14 provides the bivariate regression associations between educational attainment (Wave III) and delinquency (Wave IV), and educational attainment (Wave III) and drug use (Wave IV) estimated on the post-matching subsample.[52] The findings of the bivariate OLS regression model suggested that the association between educational attainment (Wave III) and delinquency (Wave IV) has a -.005 slope (b = -.005, SE = .007, β = -.011, 95%CI = -.020, .010), which was not statistically different from the null hypothesis (i.e., a slope of 0). Concerning drug use, the bivariate OLS regression model suggested that the association between educational attainment (Wave III) and drug use (Wave IV) has a -.090 slope (b = -.090, SE = .041, β = -.036, 95%CI = -.171, -.009), which was statistically significant at the $p < .05$ level.

Table 6.15 provides the bivariate associations between intelligence (Wave III) and delinquency (Wave IV) estimated on the post-GPS matching subsample. The results of Model 1 (specifying a linear association between intelligence and delinquency) suggested that the association between intelligence (Wave III) and delinquency (Wave IV) has a -.001 slope (b = -.001, SE = .001, 95%CI = -.001, .001), which was not statistically different from the null hypothesis. Similarly, the results of the quadratic specification suggested that the association between intelligence (Wave III) and delinquency (Wave IV) has a .001 slope (b = .001, SE = .001, 95%CI = -.002, .002) and the association between intelligence^2 (Wave III) and delinquency (Wave IV) has a -.001 slope (b = -.001, SE = .001, 95%CI = -.001, .001), which were not statistically different from the null hypothesis. The results of the cubic specification indicated that the association between intelligence (Wave III) and delinquency (Wave IV) has a -.007 slope (b = -.007, SE = .003, 95%CI = -.013, .001), the association between intelligence^2 (Wave

[52] Treatment and control cases were matched using nearest neighbor matching with a caliper of .005.

III) and delinquency (Wave IV) has a .001 slope (b = .001, SE = .001, 95%CI = .001, .001), and the association between intelligence^3 (Wave III) and delinquency (Wave IV) has a -.001 slope (b = -.001, SE = .001, 95%CI = -.001, -.001). The point estimates pertaining to the associations between intelligence^2 (Wave III) and delinquency (Wave IV), and intelligence^3 (Wave III) and delinquency (Wave IV) were statistically significant at the $p < .05$ level.

Table 6.16 provides the bivariate associations between intelligence (Wave III) and drug use (Wave IV) estimated on the post-GPS matching subsample. The results of Model 1 (specifying a linear association between intelligence and drug use) suggested that the association between intelligence (Wave III) and drug use (Wave IV) has a .003 slope (b = .003, SE = .001, 95%CI = .001, .005), which was statistically significant at the $p < .05$ level. The results of the quadratic specification suggested that the association between intelligence (Wave III) and drug use (Wave IV) has a -.011 slope (b = -.011, SE = .006, 95%CI = -.022, .001) and the association between intelligence^2 (Wave III) and drug use (Wave IV) has a .001 slope (b = .001, SE = .001, 95%CI = .001, .001). For Model 2, the bivariate association between intelligence (Wave III) and drug use (Wave IV) was not statistically significant, while the bivariate association between intelligence^2 (Wave III) and drug use (Wave IV) was statistically significant at the $p < .05$ level. The results of the cubic specification indicated that the association between intelligence (Wave III) and drug use (Wave IV) has a -.084 slope (b = -.084, SE = .018, 95%CI = -.120, -.048), the association between intelligence^2 (Wave III) and drug use (Wave IV) has a .001 slope (b = .001, SE = .001, 95%CI = .001, .002), and the association between intelligence^3 (Wave III) and drug use (Wave IV) has a -.001 slope (b = -.001, SE = .001, 95%CI = -.001, -.001). The point estimates pertaining to the associations between intelligence (Wave III) and drug use (Wave IV), intelligence^2 (Wave III) and drug use (Wave IV), and intelligence^3 (Wave III) and drug use (Wave IV) were statistically significant at the $p < .05$ level.

Table 6.14: Predicting delinquency (Wave IV) and drug use (Wave IV) with educational attainment (Wave III) after nearest neighbor matching.

	DV: Delinquency (Wave IV)				DV: Drug Use (Wave IV)			
	b	SE	β	95%CI	b	SE	β	95%CI
Treatment Conditions (Wave III)								
Educational Attainment	-.005	.007	-.011	-.020,.010	-.090*	.041	-.036	-.171,-.009
R^2		-.001				.001		
N		3,742				3,728		

Notes: Caliper for the nearest neighbor matching was set at $p = .005$.
*$p < .05$

Table 6.15: Predicting delinquency (Wave IV) with intelligence (Wave III) post generalized propensity score matching.

DV: Delinquency	Model 1			Model 2			Model 3		
(Wave IV)	b	SE	95%CI	b	SE	95%CI	b	SE	95%CI
Treatment Conditions (Wave III)									
Intelligence	-.001	.001	-.001,.001	.001	.001	-.002,.002	-.007	.003	-.013,.001
Intelligence^2	--	--	--	-.001	.001	-.001,.001	.001*	.001	.001,.001
Intelligence^3	--	--	--	--	--	--	-.001*	.001	-.001,-.001
R^2		.001			.001			.001	
N		6,230			6,230			6,230	

Notes: The balancing analysis indicated that the generalized propensity score matching achieved balance at the $p < .01$ level.
*$p < .05$

Table 6.16: Predicting drug use (Wave IV) with intelligence post generalized propensity score matching.

DV: Drug use	Model 1			Model 2			Model 3		
(Wave IV)	b	SE	95%CI	b	SE	95%CI	b	SE	95%CI
Treatment Conditions (Wave III)									
Intelligence	.003*	.001	.001,.005	-.011	.006	-.022,.001	-.084*	.018	-.120,-.048
Intelligence^2	--	--	--	.001*	.001	.001,.001	.001*	.001	.001,.002
Intelligence^3	--	--	--	--	--	--	-.001*	.001	-.001,.-.001
R^2		.005*			.007*			.009*	
N		6,213			6,213			6,213	

Notes: The balancing analysis indicated that the generalized propensity score matching achieved balance at the $p < .01$ level.
*$p < .05$

6.3 Study 3: Adjusting for Genetic Self-Selection with MZ Difference Scores

6.3.1. Descriptive statistics and cross-twin correlations: MZ twin subsample.

Table 6.17 provides the descriptive statistics for the MZ twin subsample and a cross twin mean difference comparison. To reiterate, only identical twins (MZ twins) were included in the subsample and the two twins within a twin pair were randomly assigned the designation *Twin 1* or *Twin 2*. As indicated by the cross twin mean difference comparison, the average score for Twin 1 cases across the 17 variables of interest did not significantly differ from the average score for Twin 2 cases. Furthermore, while the largest mean difference (i.e., maternal conflict) did approach statistical significance (*t-value* = 1.90), the next largest mean difference had a *t-value* of -1.28.

Table 6.17: Descriptive statistics for the MZ Twin Sample.

	Twin 1				Twin 2				*t-value*
	N	$\bar{X}$	SD	Min,Max	N	$\bar{X}$	SD	Min,Max	
Dependent Variables (Wave IV)									
Delinquency	235	.03	.12	.00,1.20	240	.05	.17	.00,1.40	-1.28
Drug Use	235	-.14	1.17	-1.60,3.45	237	-.24	1.20	-1.60,3.45	.96
Treatment Conditions (Wave III)									
Intelligence	230	96.29	15.84	8,122	222	97.55	16.88	7,122	-.81
Educational Attainment	238	.53	.50	0,1	228	.54	.50	0,1	-.13
Covariates (Wave I)									
Maternal Conflict	257	.02	1.17	-1.35,4.23	255	-.17	1.06	-1.35,4.18	1.90
Paternal Conflict	199	-.13	1.20	-1.37,4.41	200	-.22	1.20	-1.37,4.72	.79
School Attachment	276	.04	1.10	-3.56,2.06	278	.07	1.12	-3.89,2.06	-.31
Social Support	275	8.06	1.20	3.75,10.00	278	8.12	1.17	3.75,10.00	-.62
Peer Drug Use	275	.88	.90	0,3	273	.82	.91	0,3	.72
Baseline Delinquency	280	.56	.74	.00,4.86	280	.49	.72	.00,6.00	1.12
Baseline Drug Use	276	.49	.45	.00,1.67	277	.48	.43	.00,2.00	.05

Notes: Drug Use, Maternal Conflict, Paternal Conflict, and School attachment were standardized to account for the differences in coding schemes between items. No between twin differences were expected for Age, Non-White, Male, Parent Income, Parent Employment Status, and Parent Education because these mechanisms are encompassed within the shared environment.

*$p < .05$

Table 6.18 presents the cross-twin correlations (i.e., the correlation between Twin 1 and Twin 2) on the 17 variables of interest. In reference to the dependent variables, it can be observed that the scores for Twin 1 and the scores for Twin 2 were correlated at .324 (95%CI = .199, .439, $p < .05$) for delinquency (Wave IV) and .445 (95%CI = .330, .547, $p < .05$) for drug use (Wave IV). For the treatment conditions of interest, it can be observed that the scores for Twin 1 and the scores for Twin 2 were correlated at .704 (95%CI = .629, .767, $p < .05$) for intelligence (Wave III) and .602 (95%CI = .510, .681, $p < .05$) for educational attainment (Wave III). As indicated by the cross-twin correlations for the covariates, the scores for Twin 1 and the scores for Twin 2 were perfectly correlated on five of the 13 covariates. Specifically, the scores for Twin 1 and Twin 2 were perfectly correlated for age ($r = 1.00$, 95%CI = 1.00, 1.00, $p < .05$), male ($r = 1.00$, 95%CI = 1.00, 1.00, $p < .05$), parent income ($r = 1.00$, 95%CI = 1.00, 1.00, $p < .05$), parent employment status ($r = 1.00$, 95%CI = 1.00, 1.00, $p < .05$), and parent education ($r = 1.00$, 95%CI = 1.00, 1.00, $p < .05$). Due to the perfect cross-twin correlations on the specified variables, a MZ difference score could not be created of age, male, parent income, parent employment status, and parent education.

The next highest cross-twin correlation was observed for non-White ($r = .893$, 95%CI = .867, .914, $p < .05$). A MZ difference score was not created for non-White, since the disjunction between the scores for Twin 1 and the scores for Twin 2 on non-White were likely due to measurement error, since it is unlikely for identical twins to differ on race. The cross-twin correlations concerning maternal conflict ($r = .473$, 95%CI = .371, .564, $p < .05$), paternal conflict ($r = .554$, 95%CI = .448, .645, $p < .05$), school attachment ($r = .475$, 95%CI = .377, .561, $p < .05$), social support ($r = .503$, 95%CI = .409, .587, $p < .05$), peer drug use ($r = .658$, 95%CI = .585, .721, $p < .05$), baseline delinquency ($r = .450$, 95%CI = .351, .539, $p < .05$), and

baseline drug use (r = .629, 95%CI = .552, .696, p < .05) permitted the creation of MZ difference scores for the specified covariates.

Table 6.18. Cross-MZ twin correlations.

	N_{pairs}	r	95%CI
Dependent Variables (Wave IV)			
Delinquency	235	.324*	.199,.439
Drug Use	235	.445*	.330,.547
Treatment Conditions (Wave III)			
Intelligence	222	.704*	.629,.767
Educational Attainment	228	.602*	.510,.681
Covariates (Wave I)			
Maternal Conflict	255	.473*	.371,.564
Paternal Conflict	199	.554*	.448,.645
School Attachment	276	.475*	.377,.561
Social Support	275	.503*	.409,.587
Peer Drug Use	273	.658*	.585,.721
Baseline Delinquency	280	.450*	.351,.539
Baseline Drug Use	276	.629*	.552,.696

Notes: Drug Use, Maternal Conflict, Paternal Conflict, and School attachment were standardized to account for the differences in coding schemes between items. No between twin differences were expected for Age, Non-White, Male, Parent Income, Parent Employment Status, and Parent Education because these mechanisms are encompassed within the shared environment.
*p < .05

6.3.2. Descriptive statistics for MZ difference scores.

Table 6.19 provides the descriptive statistics for the MZ difference scores on the covariates where the scores for Twin 1 were not perfectly correlated with the scores for Twin 2.[53] Concerning the dependent variables, it can be observed that the average difference between Twin 1 and Twin 2 on delinquency (Wave IV) was approximately -.004 (min, max: -1.200, 1.200), and the average difference between Twin 1 and Twin 2 on drug use (Wave IV) was approximately .163 (min, max: -4.002, 4.002). In reference to the treatment conditions, it can be observed that the average difference between Twin 1 and Twin 2 on intelligence (Wave III) was approximately -.855 (min, max: -52, 99), and the average difference on educational attainment (Wave III) was .005 (min, max: -1, 1). Regarding the covariates, it can be observed that average difference between Twin 1 and Twin 2 was .170 for maternal conflict (min, max: -3.115, 5.576), for paternal conflict (min, max: -4.417, 3.183), -.027 for school attachment (min, max: -4.085, 3.697), -.065 for social support (min, max: -3.750, 4.000), .053 for peer drug use (min, max: -2.667, 2.667), .070 for baseline delinquency (min, max: -5.571, 4.286), and .005 for baseline drug use (min, max: -1.333, 1.667).

[53] An MZ difference score was not created for non-White, since the disjunction between the scores for Twin 1 and the scores for Twin 2 were likely a function of measurement error.

Table 6.19: Descriptive statistics for the MZ difference scores.

	N_{pairs} (N_{In})	$\bar{X}$	SD	Min,Max
Dependent Variables (Wave IV)				
Delinquency	215 (430)	-.004	.172	-1.200,1.200
Drug Use	212 (424)	.163	1.221	-4.002,4.002
Treatment Conditions (Wave III)				
Intelligence	209 (418)	-.885	12.833	-52,99
Educational Attainment	217 (434)	-.005	.445	-1,1
Covariates (Wave I)				
Maternal Conflict	250 (500)	.170	1.142	-3.115,5.576
Paternal Conflict	193 (386)	.071	1.121	-4.417,3.183
School Attachment	274 (548)	-.027	1.138	-4.085,3.697
Social Support	273 (546)	-.065	1.182	-3.750,4.000
Peer Drug Use	268 (536)	.053	.747	-2.667,2.667
Baseline Delinquency	279 (558)	.070	.770	-5.571,4.286
Baseline Drug Use	272 (544)	-.005	.380	-1.333,1.667

Notes: Difference scores were only created for the variables in which the correlation between the MZ twins was below 1.00. "N_{In}" = Number of individuals.

6.3.3. Multivariate models of the MZ difference scores for the dependent variables on the MZ difference scores for the treatment variables

Table 6.20 presents the bivariate and multivariate OLS regression models, where the MZ difference score for dependent variables (Wave IV) were regressed on the MZ difference score for educational attainment (Wave III) and the covariates. The results of the bivariate model where the MZ difference score for delinquency (Wave IV) was regressed on the MZ difference score for educational attainment (Wave III) suggested that the association between educational attainment (Wave III) and delinquency (Wave IV) has a -.016 slope (b = -.016, SE = .034, β = -.040, 95%CI = -.085, .053, p > .05). Consistent with the bivariate results, the multivariate analysis suggested that the association between educational attainment (Wave III) and delinquency (Wave IV) has a .001 slope (b = .001, SE = .019, β = .001, 95%CI = -.037, .038, p > .05). Neither the bivariate or multivariate associations reached statistical significance. In reference to drug use (Wave IV), the bivariate model suggested that the association between educational attainment (Wave III) and drug use (Wave IV) has a -.021 slope (b = -.021, SE = .186, β = -.008, 95%CI = -.388, .345, p > .05). Similarly, the multivariate analysis suggested that the association between educational attainment (Wave III) and delinquency (Wave IV) has a -.100 slope (b = -.100, SE = .302, β = -.032, 95%CI = -.699, .499, p > .05). Again, the bivariate and the multivariate associations between educational attainment (Wave III) and drug use (Wave IV) did not reach statistical significance.

Table 6.20: Predicting the MZ difference scores for delinquency (Wave IV) and drug use (Wave IV) with the MZ difference scores for educational attainment (Wave III) and covariates.

	DV: Delinquency (Wave IV)				DV: Delinquency (Wave IV)				DV: Drug Use (Wave IV)				DV: Drug Use (Wave IV)			
	b	SE	β	95%CI	b	SE	β	95%CI	b	SE	β	95%CI	b	SE	β	95%CI
Treatment Conditions (Wave III)																
Educational Attainment	-.016	.034	-.040	-.085,.053	.001	.019	.001	-.037,.038	-.021	.186	-.008	-.388,.345	-.100	.302	-.032	-.699,.499
Covariates (Wave I)																
Maternal Conflict	--	--	--	--	-.005	.007	-.035	-.020,.009	--	--	--	--	.014	.106	.012	-.196,.224
Paternal Conflict	--	--	--	--	.009	.019	.060	-.029,.046	--	--	--	--	-.196	.121	-.162	-.437,.044
School Attachment	--	--	--	--	.020	.014	.120	-.008,.047	--	--	--	--	-.155	.147	-.117	-.447,.136
Social Support	--	--	--	--	-.003	.009	-.023	-.022,.016	--	--	--	--	-.136	.125	-.122	-.383,.111
Peer Drug Use	--	--	--	--	-.013	.012	-.053	-.037,.011	--	--	--	--	.113	.240	.058	-.364,.589
Baseline Delinquency	--	--	--	--	.055*	.027	.180	.001,.109	--	--	--	--	--	--	--	--
Baseline Drug Use	--	--	--	--	--	--	--	--	--	--	--	--	.843*	.389	.239	.072,1.615
R^2	-.004				-.033				-.005				.053			
N_{pairs} (N_{In})	187 (374)				111 (222)				184 (368)				101 (202)			

Notes: Difference scores were only created for the variables in which the correlation between the MZ twins was below 1.00. "N_{In}" = Number of individuals.
*$p < .05$

Table 6.21 provides the results of the multivariate OLS regression models, where the MZ difference score for delinquency (Wave IV) was regressed on the MZ difference score for intelligence (Wave III) and the covariates. The results of Model 1 (specifying a linear association between intelligence and delinquency) suggested that the association between intelligence (Wave III) and delinquency (Wave IV) has a -.002 slope (b = -.002, SE = .002, β = -.107, 95%CI = -.0056 .003, p > .05). Similarly, the results of the quadratic specification suggested that the association between intelligence (Wave III) and delinquency (Wave IV) has a -.002 slope (b = -.002, SE = .004, β = -.127, 95%CI = -.010, .006, p > .05) and the association between intelligence^2 (Wave III) and delinquency (Wave IV) has a -.001 slope (b = -.001, SE = .001, β = -.039, 95%CI = -.001, .001, p > .05). The results of the cubic specification indicated that the association between intelligence (Wave III) and delinquency (Wave IV) has a -.002 slope (b = -.002, SE = .002, β = -.127, 95%CI = -.007, .003, p > .05), the association between intelligence^2 (Wave III) and delinquency (Wave IV) has a -.001 slope (b = -.001, SE = .001, β = -.039, 95%CI = -.001, .001, p > .05), and the association between intelligence^3 (Wave III) and delinquency (Wave IV) has a .001 slope (b = .001, SE = .001, β = .001, 95%CI = -.001, .001, p > .05). None of the observed associations presented in Table 6.21 reached statistical significance.

Table 6.21: Predicting the MZ difference score for delinquency (Wave IV) with the MZ difference scores for intelligence (Wave III) and covariates.

DV: Delinquency	Model 1				Model 2				Model 3			
(Wave IV)	b	SE	β	95%CI	b	SE	β	95%CI	b	SE	β	95%CI
Treatment Conditions (Wave III)												
Intelligence	-.002	.002	-.107	-.006,.003	-.002	.004	-.127	-.010,.006	-.002	.002	-.127	-.007,.003
Intelligence^2	--	--	--	--	-.001	.001	-.039	-.001,.001	-.001	.001	-.039	-.001,.001
Intelligence^3	--	--	--	--	--	--	--	--	.001	.001	.001	-.001,.001
Covariates (Wave I)												
Maternal Conflict	-.004	.006	-.026	-.016,.009	-.005	.010	-.031	-.025,.016	-.005	.006	-.031	-.016,.007
Paternal Conflict	.007	.019	.049	-.031,.045	.008	.024	.056	-.039,.056	.008	.023	.056	-.037,.053
School Attachment	.023	.016	.130	-.009,.054	.023	.015	.133	-.008,.054	.023	.015	.133	-.008,.054
Social Support	-.001	.011	-.009	-.022,.020	-.001	.010	-.010	-.021,.018	-.001	.010	-.010	-.021,.018
Peer Drug Use	-.008	.013	-.032	-.033,.017	-.009	.012	-.035	-.033,.016	-.009	.013	-.035	-.033,.016
Baseline Delinquency	.055	.029	.178	-.002,.111	.053	.031	.173	-.008,.114	.053	.030	.173	-.006,.112
R^2			-.023				-.032				-.043	
N_{pairs} (N_{In})			106 (212)				106 (212)				106 (212)	

Notes: Difference scores were only created for the variables in which the correlation between the MZ twins was below 1.00. "N_{In}" = Number of individuals.
*$p < .05$

Table 6.22 provides the results of the multivariate OLS regression models, where the MZ difference score for drug use (Wave IV) was regressed on the MZ difference score for intelligence (Wave III) and the covariates. The results of Model 1 (specifying a linear association between intelligence and drug use) suggested that the association between intelligence (Wave III) and drug use (Wave IV) has a -.021 slope (b = -.021, SE = .012, β = -.169, 95%CI = -.045, .003, p > .05). Similarly, the results of the quadratic specification suggested that the association between intelligence (Wave III) and drug use (Wave IV) has a -.019 slope (b = -.019, SE = .014, β = -.152, 95%CI = -.047, .009, p > .05) and the association between intelligence^2 (Wave III) and drug use (Wave IV) has a .001 slope (b = .001, SE = .001, β = .035, 95%CI = -.001, .001, p > .05). The results of the cubic specification indicated that the association between intelligence (Wave III) and drug use (Wave IV) has a -.009 slope (b = -.009, SE = .015, β = -.076, 95%CI = -.040, .021, p > .05), the association between intelligence^2 (Wave III) and drug use (Wave IV) has a -.001 slope (b = -.001, SE = .001, β = -.247, 95%CI = -.003, .001, p > .05), and the association between intelligence^3 (Wave III) and drug use (Wave IV) has a -.001 slope (b = -.001, SE = .001, β = -.369, 95%CI = -.001, .001, p > .05). None of the observed associations presented in Table 6.22 reached statistical significance.

Table 6.22: Predicting the MZ difference score for drug use (Wave IV) with the MZ difference scores for intelligence (Wave III) and covariates.

DV: Drug Use	Model 1				Model 2				Model 3			
(Wave IV)	b	SE	β	95%CI	b	SE	β	95%CI	b	SE	β	95%CI
Treatment Conditions (Wave III)												
Intelligence	-.021	.014	-.169	-.049,.007	-.019	.019	-.152	-.056,.019	-.009	.019	-.076	-.047,.028
Intelligence^2	--	--	--	--	.001	.001	.035	-.001,.001	-.001	.001	-.247	-.004,.002
Intelligence^3	--	--	--	--	--	--	--	--	-.001	.001	-.369	-.001,.001
Covariates (Wave I)												
Maternal Conflict	.048	.101	.041	-.153,.249	.054	.108	.046	-.159,.268	.098	.106	.083	-.112,.307
Paternal Conflict	-.205	.123	-.167	-.450,.039	-.214	.139	-.175	-.490,.061	-.230	.132	-.187	-.492,.033
School Attachment	-.149	.150	-.108	-.448,.150	-.153	.152	-.112	-.457,.148	-.169	.153	-.123	-.473,.135
Social Support	-.100	.128	-.089	-.354,.155	-.097	.128	-.087	-.351,.156	-.099	.125	-.088	-.348,.151
Peer Drug Use	.182	.250	.093	-.315,.678	.191	.262	.098	-.329,.712	.232	.263	.119	-.292,.755
Baseline Drug Use	.778	.406	.214	-.028,1.585	.778	.409	.214	-.034,1.591	.778*	.405	.214	-.026,1.583
R^2			.077				.068				.079	
N_{pairs} (N_{In})			103 (206)				103 (206)				103 (206)	

Notes: Difference scores were only created for the variables in which the correlation between the MZ twins was below 1.00. "N_{In}" = Number of individuals.
*$p < .05$

6.4 Summary of Findings: Studies 1, 2, and 3[54]

6.4.1. Educational attainment and antisocial behavior

Table 6.23 provides the point estimates for the bivariate and multivariate associations between educational attainment (Wave III) and delinquency (Wave IV), and educational attainment (Wave III) and drug use (Wave IV) derived from Studies 1, 2, and 3. Regarding the association between educational attainment (Wave III) and delinquency (Wave IV), discrepancies existed between the baseline bivariate association and the baseline multivariate association, the post-matching bivariate associations, and the MZ difference score bivariate association and multivariate association. Specifically, while the baseline bivariate regression model of delinquency (Wave IV) on educational attainment (Wave III) suggested a statistically significant -.029 association (b = -.029, SE = .004, β = -.061, 95%CI = -.038, -.021, $p < .05$), the models adjusting for social self-selection (PSM Caliper = .005: b = -.005, SE = .007, β = -.011, 95%CI = -.020,.010, $p < .05$) and genetic self-selection (MZ difference with covariates: b = -.016, SE = .034, β = -.040, 95%CI = -.085, .053, $p > .05$) suggested that the slope of the association was not statistically significant and ranged between approximately -.016 (i.e., Bivariate MZ difference model) to .001 (i.e., Multivariate MZ difference model).

Concerning educational attainment (Wave III) and drug use (Wave IV), discrepancies existed between the baseline bivariate association and the baseline multivariate association, the post-matching bivariate associations, and the MZ difference score bivariate association and multivariate association. Specifically, while the baseline bivariate regression model of drug use (Wave IV) on educational attainment (Wave III) suggested a statistically significant -.75 association (b = -.75, SE = .022, β = -.070, 95%CI = -.218,-.132, $p < .05$), the models adjusting

[54] While comparisons are provided, the analytical samples vary in size and generalizability between the different estimation techniques.

for social self-selection (PSM Caliper = .005: b = -.090, SE = .041, β = -.036, 95%CI = -.171,-.009, $p < .05$) and genetic self-selection (MZ difference with covariates: b = -.021, SE = .186, β = -.008, 95%CI = -.388, .345, $p > .05$) suggested that the slope of the association was not statistically significant (the bivariate association was only statistically significant when the nearest neighbor caliper was set at .05 and .005) and ranged between approximately -.100 (i.e., Multivariate MZ difference model) to -.021 (i.e., Bivariate MZ difference model).

Table 6.23. Point estimates for educational attainment when predicting drug use (Wave IV).

	DV: Delinquency (Wave IV)				DV: Drug Use (Wave IV)			
	b	SE	β	95%CI	b	SE	β	95%CI
Baseline (Wave III)								
Educational Attainment	-.029*	.004	-.061	-.038,-.021	-.175*	.022	-.070	-.218,-.132
Educational Attainment with covariates	-.011	.006	-.024	-.023,.001	-.055	.031	-.022	-.116,.007
Post Matching (Wave III)								
Educational Attainment (caliper = .05)	-.009	.007	-.020	-.024,.005	-.094*	.040	-.038	-.173,-.016
Educational Attainment (caliper = .01)	-.005	.008	-.011	-.020,.010	-.074	.041	-.029	-.153,.006
Educational Attainment (caliper = .005)	-.005	.007	-.011	-.020,.010	-.090*	.041	-.036	-.171,-.009
Educational Attainment (caliper = .001)	-.005	.008	-.012	-.021,.011	-.054	.046	-.022	-.144,.035
Educational Attainment (caliper = .0001)	-.026	.017	-.051	-.060,.008	-.048	.087	-.019	-.218,.122
MZ Difference Models (Wave III)								
Educational Attainment	-.016	.034	-.040	-.085,.053	-.021	.186	-.008	-.388,.345
Educational Attainment with covariates	.001	.019	.001	-.037,.038	-.100	.302	-.032	-.699,.499

Notes: The sample size varies between each estimation technique and each point estimate. The covariates included in the baseline models and the PSM model are as followed: Age, Non-White, Male, Parent Income, Parent Employment Status, Parent Education, Maternal Conflict, Paternal Conflict, School Attachment, Social Support, Peer Drug use, and Baseline Delinquency or Drug Use (depending upon the dependent variable). The covariates included in the MZ difference model are as followed: Maternal Conflict, Paternal Conflict, School Attachment, Social Support, Peer Drug use, and Baseline Delinquency or Drug Use (depending upon the dependent variable).

*$p < .05$

6.4.2. Intelligence and antisocial behavior

Table 6.24 presents the point estimates for the bivariate and multivariate associations between intelligence (Wave III) and delinquency (Wave IV) derived from Studies 1, 2, and 3. Evident by the results, the linear, quadratic, and cubic point estimates for the baseline bivariate models, baseline multivariate models, post-matching models, and MZ difference models generally suggested that the slope of the association between intelligence (Wave III) and delinquency (Wave IV) was not statistically significant and ranged between -.007 (Model 3: post-GPS matching bivariate association) or .001 (Model 2: baseline bivariate model, baseline multivariate model, and post-GPS matching bivariate model). Furthermore, the slope of the association between intelligence^2 (Wave III) and delinquency (Wave IV) was not statistically significant and ranged between -.001 (Model 2: baseline bivariate model, baseline multivariate model, post-GPS matching bivariate model, and MZ difference score multivariate model; Model 3: MZ difference score bivariate model and MZ difference score multivariate model) or .001 (Model 2: MZ difference score bivariate model; Model 3: baseline bivariate model, baseline multivariate model, and post-GPS matching bivariate model). Similarly, the slope of the association between intelligence^3 (Wave III) and delinquency (Wave IV) was not statistically significant and ranged between -.001 (Model 3: baseline bivariate model, baseline multivariate model, and post-GPS matching bivariate model) or .001 (Model 3: MZ difference score bivariate model and MZ difference score multivariate model).

Table 6.24. Point estimates for intelligence when predicting delinquency (Wave IV).

DV: Delinquency	Model 1				Model 2				Model 3			
(Wave IV)	b	SE	β	95%CI	b	SE	β	95%CI	b	SE	β	95%CI
Baseline (Wave III)												
Intelligence	-.001	.001	-.003	-.001,.001	.001	.001	.053	-.001,.001	-.001	.001	-.081	-.005,.002
Intelligence^2	--	--	--	--	-.001	.001	-.057	-.001,.001	.001	.001	.335	-.001,.001
Intelligence^3	--	--	--	--	--	--	--	--	-.001	.001	-.264	-.001,.001
Intelligence with covariates	-.001	.001	-.001	-.001,.001	.001	.001	.016	-.001,.001	-.002	.001	-.120	-.007,.003
Intelligence^2 with covariates	--	--	--	--	-.001	.001	-.017	-.001,.001	.001	.001	.387	-.001,.001
Intelligence^3 with covariates	--	--	--	--	--	--	--	--	-.001	.001	-.274	-.001,.001
Post GPS Matching (Wave III)												
Intelligence	-.001	.001	--	-.001,.001	.001	.001	--	-.002,.002	-.007	.003	--	-.013,.001
Intelligence^2	--	--	--	--	-.001	.001	--	-.001,.001	.001*	.001	--	.001,.001
Intelligence^3	--	--	--	--	--	--	--	--	-.001	.001	--	-.001,-.001
MZ Difference Models (Wave III)												
Intelligence	-.001	.001	-.098	-.003,.001	-.001	.001	-.119	-.004,.001	-.003	.002	-.183	-.007,.002
Intelligence^2	--	--	--	--	.001	.001	.060	-.001,.001	-.001	.001	-.056	-.001,.001
Intelligence^3	--	--	--	--	--	--	--	--	.001	.001	.164	-.001,.001
Intelligence with covariates	-.002	.002	-.107	-.006,.003	-.002	.004	-.127	-.010,.006	-.002	.002	-.127	-.007,.003
Intelligence^2 with covariates	--	--	--	--	-.001	.001	-.039	-.001,.001	-.001	.001	-.039	-.001,.001
Intelligence^3 with covariates	--	--	--	--	--	--	--	--	.001	.001	.001	-.001,.001

Notes: The sample size varies between each estimation technique and each point estimate. The covariates included in the baseline models and the PSM model are as followed: Age, Non-White, Male, Parent Income, Parent Employment Status, Parent Education, Maternal Conflict, Paternal Conflict, School Attachment, Social Support, Peer Drug use, and Baseline Delinquency. The covariates included in the MZ difference model are as followed: Maternal Conflict, Paternal Conflict, School Attachment, Social Support, Peer Drug use, and Baseline Delinquency.
*$p < .05$

Table 6.25 presents the point estimates for the bivariate and multivariate associations between intelligence (Wave III) and drug use (Wave IV) derived from Studies 1, 2, and 3. Evident by the results, the linear, quadratic, and cubic point estimates for the baseline bivariate models, baseline multivariate models, post-matching models, and MZ difference models generally suggested that the slope of the association between intelligence (Wave III) and drug use (Wave IV) could be statistically significant and ranged between -.081 (Model 3: Baseline bivariate model) or .003 (Model 1: baseline multivariate model and post-GPS matching bivariate model). Furthermore, the slope of the association between intelligence^2 (Wave III) and drug use (Wave IV) could be statistically significant and ranged between -.001 (Model 2: baseline bivariate model, baseline multivariate model, post-GPS matching bivariate model, MZ difference score bivariate model, and MZ difference score multivariate model; Model 3 baseline bivariate model, baseline multivariate model, and post-GPS matching bivariate model) or .001 (Model 3: MZ difference score bivariate model, and MZ difference score multivariate model). Similarly, the slope of the association between intelligence^3 (Wave III) and drug use (Wave IV) could be statistically significant and was consistently estimated as -.001 across all specifications.

In context, these findings, although valuable, are limited when approximating the true counterfactual condition for the association between educational attainment and antisocial behavior, and intelligence and antisocial behavior. Given that PSM cannot adjust for genetic factors and MZ difference scores generally cannot adjust for non shared environmental factors an alternative methodological strategy should be employed to more closely approximate the true counterfactual condition. The subsequent subsection provides the results of the evaluation comparing the GAPSM methodology to an unconfounded PSM and a MZ difference score analysis.

Table 6.25. Point estimates for intelligence when predicting drug use (Wave IV).

DV: Drug Use	Model 1				Model 2				Model 3			
(Wave IV)	b	SE	β	95%CI	b	SE	β	95%CI	b	SE	β	95%CI
Baseline (Wave III)												
Intelligence	-.008*	.001	.100	.006,.009	-.001	.003	-.004	-.006,.005	-.081*	.009	-1.051	-.099,-.063
Intelligence^2	--	--	--	--	.001*	.001	.107	.001,.001	.001*	.001	3.163	.001,.002
Intelligence^3	--	--	--	--	--	--	--	--	-.001*	.001	-2.057	-.001,-.001
Intelligence with covariates	.003*	.001	.033	.001,.005	-.008	.001	-.094	-.016,.001	-.050*	.013	-.594	-.077,-.024
Intelligence^2 with covariates	--	--	--	--	.001*	.001	.132	.001,.001	.001*	.001	1.619	.001,.001
Intelligence^3 with covariates	--	--	--	--	--	--	--	--	-.001*	.001	-1.008	-.001,.001
Post GPS Matching (Wave III)												
Intelligence	.003*	.001	--	.001,.005	-.011	.006	--	-.022,.001	-.084*	.018	--	-.120,-.048
Intelligence^2	--	--	--	--	.001*	.001	--	.001,.001	.001*	.001	--	.001,.002
Intelligence^3	--	--	--	--	--	--	--	--	-.001*	.001	--	-.001,.-.001
MZ Difference Models (Wave III)												
Intelligence	-.017*	.007	-.184	-.030,-.004	-.018*	.008	-.193	-.034,-.003	-.021	.013	-.223	-.046,.004
Intelligence^2	--	--	--	--	.001	.001	.026	-.001,.001	-.001	.001	-.025	-.001,.001
Intelligence^3	--	--	--	--	--	--	--	--	.001	.001	.072	-.001,.001
Intelligence with covariates	-.021	.014	-.169	-.049,.007	-.019	.019	-.152	-.056,.019	-.009	.019	-.076	-.047,.028
Intelligence^2 with covariates	--	--	--	--	.001	.001	.035	-.001,.001	-.001	.001	-.247	-.004,.002
Intelligence^3 with covariates	--	--	--	--	--	--	--	--	-.001	.001	-.369	-.001,.001

Notes: The sample size varies between each estimation technique and each point estimate. The covariates included in the baseline models and the PSM model are as followed: Age, Non-White, Male, Parent Income, Parent Employment Status, Parent Education, Maternal Conflict, Paternal Conflict, School Attachment, Social Support, Peer Drug use, and Baseline Drug Use. The covariates included in the MZ difference model are as followed: Maternal Conflict, Paternal Conflict, School Attachment, Social Support, Peer Drug use, and Baseline Drug Use. The large β in the baseline model 3 correspond to the extremely small standard errors associated with the quadratic and cubic specification of intelligence.

*$p < .05$

6.5. Study 4: GAPSM Proof of Concept[55]

To identify the conditions in which the GAPSM methodology provides superior adjustments for social and genetic self-selection (i.e., point estimates closer to the true point estimate) when compared to point estimates derived from MZ difference score model and an unconfounded post-PSM model, a series of three figures were created. In each figure, five iterations of the post-GAPSM point estimates are compared to the point estimates derived from the MZ difference score approach and the unconfounded post-PSM approach. Figure 6.3 presents these comparisons for the first 13 specifications of the treatment condition and the point of equivalence, Figure 6.4 presents these comparisons for the second 13 specifications of the treatment condition and the point of equivalence, and Figure 6.5 presents these comparisons for the third 13 specifications of the treatment condition and the point of equivalence. In each figure, Panel A and Panel B compare the five iterations of the post-GAPSM point estimates to the point estimates derived from MZ difference score approach. Panel C and Panel D compare the five iterations of the post-GAPSM point estimates to the point estimates derived from unconfounded post-PSM approach.

The five iterations presented signify the most likely situations in which the GAPSM methodology can be implemented. Specifically, the black bars represent the iteration in which the participants are matched on all of the genetic variance (*a*), on 25 percent of the variance in the treatment condition predicted by the non shared environment (*e*), and on 25 percent of the variance in the treatment condition predicted by the shared environment (*c*; panel A and C in each figure). The dark gray bars represent the iteration in which the participants are matched on all of the genetic variance (*a*), on 50 percent of the variance in the treatment condition predicted by the non shared environment (*e*), and on 50 percent of the variance in the treatment condition

[55] The of 30-point estimates produced for each specification of the treatment condition (i.e., 40 different specifications, 120 point estimates in total) are presented in Appendix F.

predicted by the shared environment (*c*; panel A and C in each figure). The light gray bars represent the iteration in which the participants are matched on all of the genetic variance (*a*), on 75 percent of the variance in the treatment condition predicted by the non shared environment (*e*), and on 75 percent of the variance in the treatment condition predicted by the shared environment (*c*; panel A and C in each figure).

The red bars represent the iteration in which the participants are matched on all of the genetic variance (*a*), on 50 percent of the variance in the treatment condition predicted by the non shared environment (*e*), and on 75 percent of the variance in the treatment condition predicted by the shared environment (*c*; panel B and D in each figure). The blue bars represent the iteration in which the participants are matched on all of the genetic variance (*a*), on 75 percent of the variance in the treatment condition predicted by the non shared environment (*e*), and on 50 percent of the variance in the treatment condition predicted by the shared environment (*c*; panel B and D in each figure).

In an effort to dispel any confusion, when stating that the participants were matched on 50 percent of the variance in the treatment condition predicted by the non shared environment (*e*), this means that if the non shared environment (*e*) predicted 50 percent of the variance in the treatment condition participants were matched on 50 percent of said prediction (i.e., 25 percent of the variance in the treatment condition predicted by the non shared environment (*e*)). To reiterate, this was achieved through the creation of independent variables for the shared (*c*; i.e., x1, x2, x3, and x4) and non shared environment (*e;* i.e., x5, x6, x7, and x8). In the example provided above, two of the independent variables corresponding to the non shared environment (i.e., x5 and x6) would be used to match participants. In addition to the results of the five iterations presented in figures 6.3, 6.4, and 6.5, Appendix F provides the results for the remaining 16 iterations of post-GAPSM and allows the reader to compare the post-GAPSM

point estimates to the point estimates produced by an MZ difference score model and an unconfounded PSM model.

6.5.1. First round of thirteen specifications of the treatment condition

Figure 6.3 provides the point estimate comparisons between the five iterations of the GAPSM approach, the MZ difference score approach, and the unconfounded PSM approach for the first 13 specifications[56] of the treatment condition and the point of equivalence. The black dashed line in every panel signifies that the point estimate from the GAPSM approach is further (below the line), equivalent (equal), or closer (above the line) to the true point estimate than the MZ difference score approach or the unconfounded PSM approach.

In reference to the GAPSM/MZ difference comparison (Panel A and Panel B), it can be observed that the GAPSM approach created a point estimate equivalent to or closer to the true point estimate (1.00) for the majority of the five iterations across the first thirteen specifications. For the first iteration of the GAPSM approach (black bar; participants matched on 25 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively), it can be observed that the GAPSM point estimate was further from the true point estimate than the MZ difference approach across all of the specifications of the treatment condition. Evidence suggested that the second iteration of the GAPSM approach (dark gray bar; participants matched on 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) generally gets closer to the true point estimate than the MZ difference score approach when the genetic contribution is lower. Specifically, it can be observed that the second iteration of the GAPSM approach is closer to the true point estimate for the second (variance in *t* predicted by $a = .10$, $e = .43$, $c = .43$), third

[56] The amount of variance in the treatment condition predicted by genetic factors (*a*) increased from .05 to .65 in .05 increments, while the variance in the treatment condition predicted the non shared environment (*e*) and the shared environment (*c*) were set equal to each other.

(variance in *t* predicted by *a* = .15, *e* = .40, *c* = .40), fourth (variance in *t* predicted by *a* = .20, *e* = .38, *c* = .38), fifth (variance in *t* predicted by *a* = .25, *e* = .35, *c* = .35), sixth (variance in *t* predicted by *a* = .30, *e* = .33, *c* = .33), seventh (variance in *t* predicted by *a* = .35, *e* = .30, *c* = .30), eighth (variance in *t* predicted by *a* = .40, *e* = .28, *c* = .28), and ninth (variance in *t* predicted by *a* = .45, *e* = .25, *c* = .25) specification of the treatment condition.

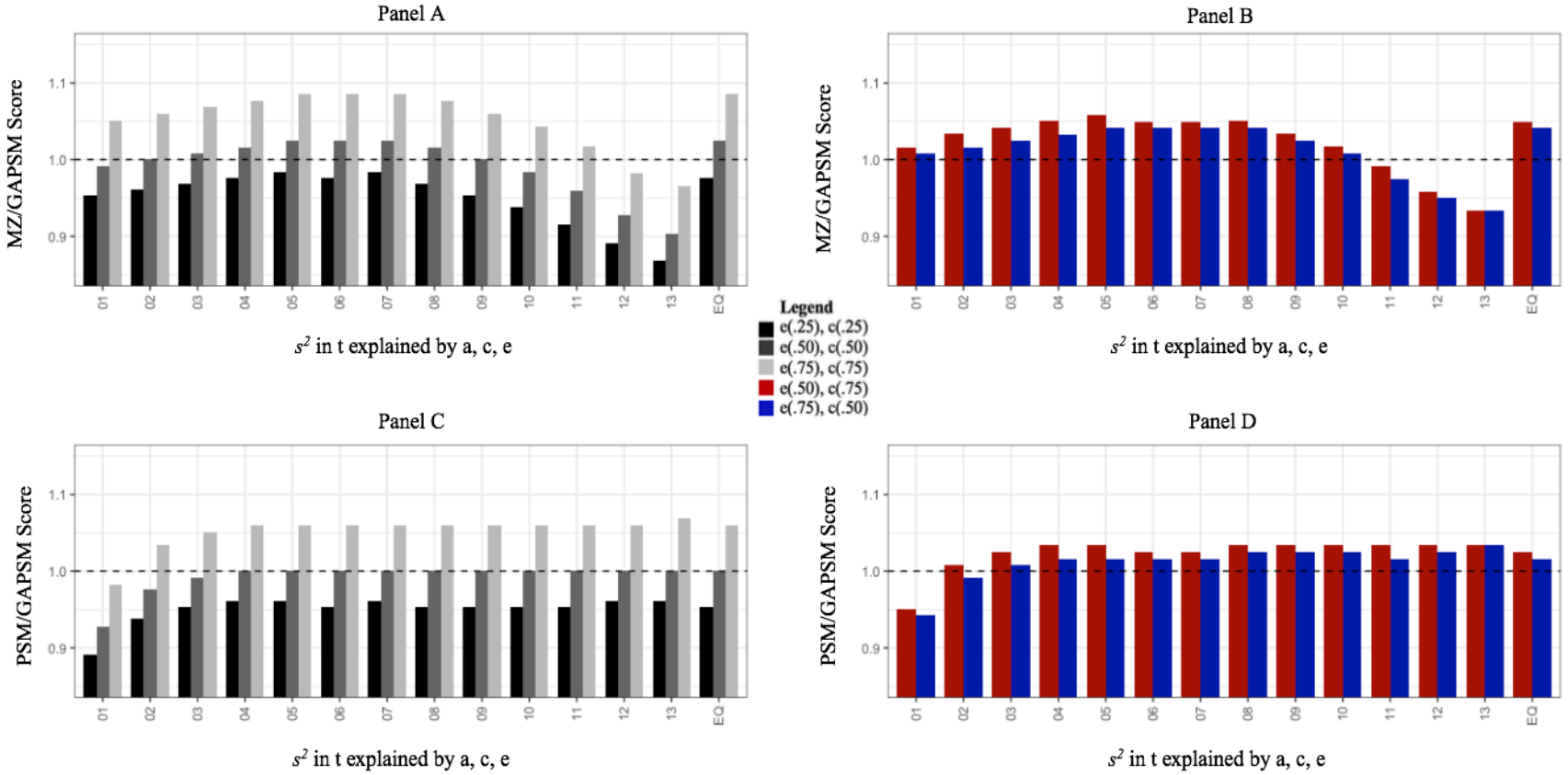
The same patterns can be observed for the third (light gray bar; participants matched on 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively), fourth (red bar; participants matched on 50 and 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively), and fifth iterations (blue bar; participants matched on 75 and 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively), with the additional evidence suggesting that these iterations perform better than the MZ difference score for the first (variance in *t* predicted by *a* = .05, *e* = .45, *c* = .45), tenth (variance in *t* predicted by *a* = .50, *e* = .23, *c* = .23), and eleventh (variance in *t* predicted by *a* = .55, *e* = .20, *c* = .20) specifications of the treatment condition.

In reference to the GAPSM/unconfounded PSM comparison (Panel C and Panel D), it can be observed that the GAPSM approach created a point estimate equivalent or closer to the true point estimate (1.00) for the majority of the five iterations across the first 13 specifications. For the first iteration of the GAPSM approach (black bar; participants matched on 25 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively), it can be observed that the GAPSM point estimate was further from the true point estimate than the unconfounded PSM approach across all of the specifications of the treatment condition. Evidence suggested that the second iteration of the GAPSM approach (dark gray bar; participants matched on 50 percent of the variance in the treatment condition

predicted by the non shared environment and shared environment, respectively) produced a point estimate generally equivalent to the point estimate produced by the unconfounded PSM approach for 11 of the 13 specifications of the treatment condition.

The third iteration (light gray bar; participants matched on 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) of the GAPSM approach generally produced a point estimate closer to that of the true point estimate than the unconfounded PSM approach across 12 of the 13 specifications of the treatment condition. Furthermore, this pattern of results was relatively consistent when comparing the fourth (red bar; participants matched on 50 and 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) and fifth (blue bar; participants matched on 75 and 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) iterations of the GAPSM approach to the point estimates derived from an unconfounded PSM approach. Distinct from the third and fourth iterations, the fifth iteration of the of the GAPSM approach (blue bar; participants matched on 75 and 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) was further from the true point estimate than the unconfounded PSM approach on the second specification (variance in *t* predicted by $a = .10$, $e = .43$, $c = .43$) of the treatment condition.

Figure 6.3: Comparing the GAPSM point estimate ($y \sim t$) to the MZ difference score and the unconfounded PSM point estimates ($c = e$).



Notes: The figure presents the first 13 specifications of the treatment condition (indicated by the number) and the point of equivalence. To reiterate, the first 13 iterations increase the variance in the treatment condition predicted by genetic factors (*a*) increased from .05 to .65 in .05 increments, while the variance in the treatment condition predicted the non shared environment (*e*) and the shared environment (*c*) were set equal to each other. Four percent (or .04) of the variance in the treatment condition was predicted by the error term (E). EQ represents the point of equivalence. *e* = Non shared environment; *c* = shared environment. For the legend, values in parentheses represents the percentage of the portion of the variance in the *treatment condition* explained by the non shared or shared environment that the participants were matched upon. The GAPSM accounted for all of the variance in the *treatment condition* explained by *a*. The black horizontal line represents a point estimate equivalent to the MZ difference point estimate or an unconfounded PSM point estimate for the specified condition. Bars above the black horizontal line indicate that the GAPSM point estimate was closer to the true point estimate (1.00) than the MZ difference point estimate or the PSM point estimate for the specified condition. Bars below the black horizontal line indicate that the GAPSM point estimate was further away from the true point estimate (1.00) than the MZ difference point estimate or the PSM point estimate for the specified condition

6.5.2. Second round of thirteen specifications of the treatment condition

Figure 6.4 provides the point estimate comparisons between the five iterations of the GAPSM approach, the MZ difference score approach, and the unconfounded PSM approach for the second 13 specifications[57] of the treatment condition and the point of equivalence. In reference to the GAPSM/MZ difference comparison (Panel A and Panel B), it can be observed that the GAPSM approach produced a point estimate equivalent or closer to the true point estimate (1.00) for the majority of the five iterations across the second thirteen specifications. For the first iteration of the GAPSM approach (black bar; participants matched on 25 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively), it can be observed that the GAPSM point estimate was generally equivalent to the point estimate produced by the MZ difference score approach.

Specifically, the results suggested that the GAPSM point estimate was equivalent to the point estimate produced by the MZ difference score approach on the fourteenth (variance in *t* predicted by *a* = .05, *e* = .68, *c* = .22), fifteenth (variance in *t* predicted by *a* = .10, *e* = .64, *c* = .21), sixteenth (variance in *t* predicted by *a* = .15, *e* = .60, *c* = .20), seventeenth (variance in *t* predicted by *a* = .20, *e* = .56, *c* = .19), eighteenth (variance in *t* predicted by *a* = .25, *e* = .53, *c* = .18), nineteenth (variance in *t* predicted by *a* = .30, *e* = .49, *c* = .16), and twentieth (variance in *t* predicted by *a* = .35, *e* = .45, *c* = .15) specification of the treatment condition. The evidence suggested that the second iteration (dark gray bar; participants matched on 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) of the GAPSM approach produces a point estimate closer to the true

[57] The amount of variance in the treatment condition predicted by genetic factors (*a*) increased from .05 to .65 in .05 increments, while the variance in the treatment condition predicted the non shared environment (*e*) was approximately three times that of the shared environment (*c*).

point estimate than the MZ difference score approach on 10 of the 13 specifications of the treatment condition.

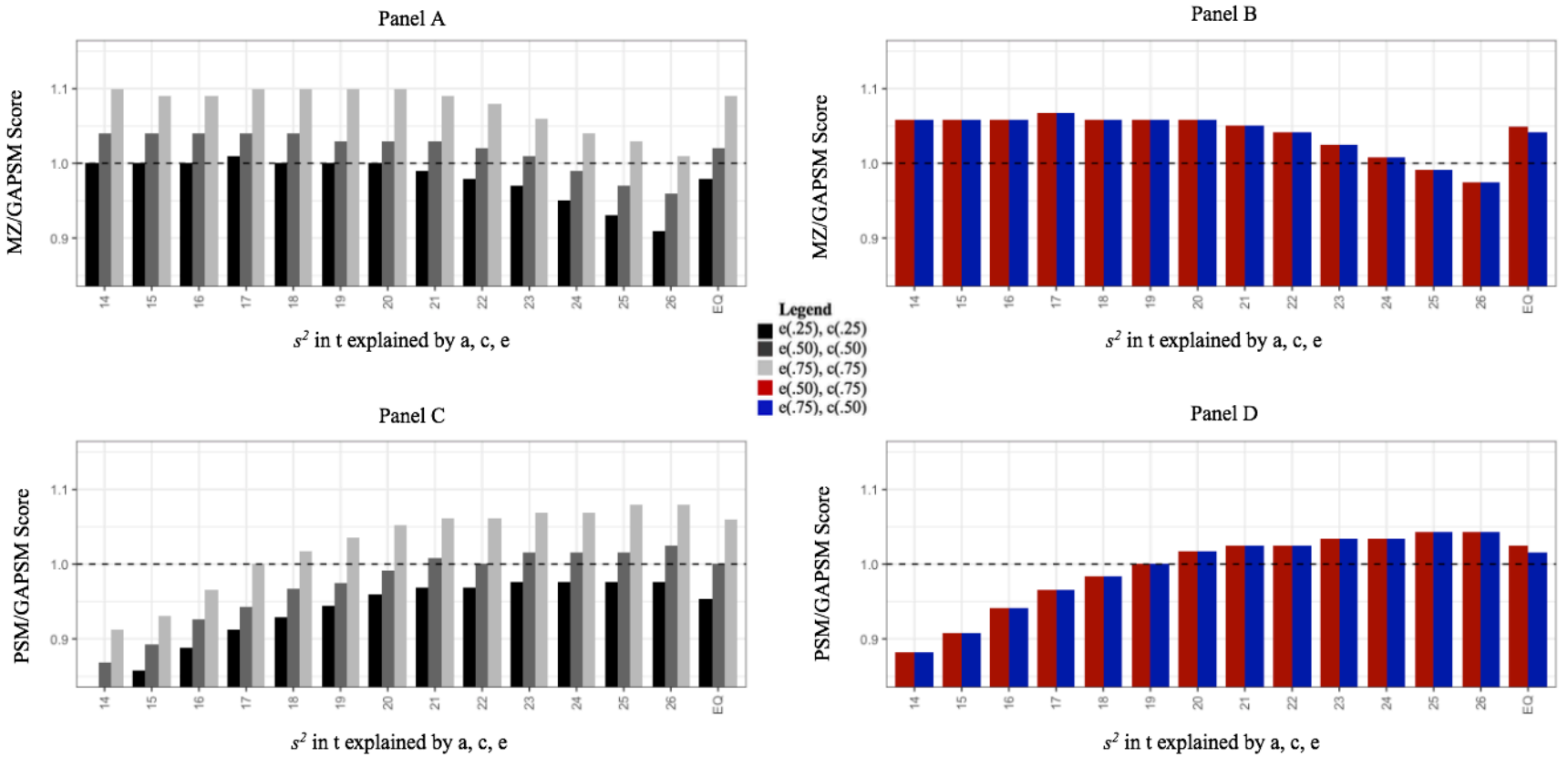
The second iteration only performs worse than the MZ difference score model when genetic factors (*a*) contribute approximately 55, 60, or 65 percent of the variance in the treatment condition. Similar to these findings, the fourth (red bar; participants matched on 50 and 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) and fifth (blue bar; participants matched on 75 and 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) iterations of the GAPSM approach produce point estimates closer to the true point estimate than the MZ difference score approach on 11 of the 13 specifications of the treatment condition. Specifically, unlike the second iteration, the fourth and fifth iterations produce point estimates closer to the true point estimate than the MZ difference score model when genetic factors (*a*) contribute approximately 55 percent of the variance in the treatment condition. The third iteration (light gray bar; participants matched on 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) of the GAPSM approach produced a point estimate closer to the true point estimate than the MZ difference score approach on all 13 specifications of the treatment condition.

In reference to the GAPSM/unconfounded PSM comparison (Panel C and Panel D), it can be observed that the GAPSM approach created a point estimate equivalent or closer to the true point estimate (1.00) for the majority of the five iterations across the second 13 specifications. For the first iteration of the GAPSM approach (black bar; participants matched on 25 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively), it can be observed that the GAPSM point estimate was further from the true point estimate than the unconfounded PSM approach across all of the

specifications of the treatment condition. Evidence suggested that the second iteration of the GAPSM approach (dark gray bar; participants matched on 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) produced a point estimate generally equivalent the point estimate produced by the unconfounded PSM model on 7 of the 13 specifications of the treatment condition.

The third iteration (light gray bar; participants matched on 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) of the GAPSM approach generally produced a point estimate closer to that of the true point estimate than the unconfounded PSM approach across 10 of the 13 specifications of the treatment condition. Furthermore, this pattern of results was relatively consistent when comparing the fourth (red bar; participants matched on 50 and 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) and fifth (blue bar; participants matched on 75 and 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) iterations of the GAPSM approach to the point estimates derived from an unconfounded PSM approach. Distinct from the third iteration, the fourth and fifth iterations of the GAPSM approach were further from the true point estimate than the unconfounded PSM approach when genetic factors predicted 20 and 25 percent of the variance in the treatment condition.

Figure 6.4: Comparing the GAPSM point estimate ($y \sim t$) to the MZ difference score and the unconfounded PSM point estimates ($e = 3*c$).



Notes: The figure presents the second 13 specifications of the treatment condition (indicated by the number) and the point of equivalence. To reiterate, the second 13 iterations increase the variance in the treatment condition predicted by genetic factors (*a*) increased from .05 to .65 in .05 increments, while the variance in the treatment condition predicted the non shared environment (*e*) is approximately three times that of the shared environment (*c*). Four percent (or .04) of the variance in the treatment condition was predicted by the error term (E). EQ represents the point of equivalence. *e* = Non shared environment; *c* = shared environment. For the legend, values in parentheses represents the percentage of the portion of the variance in the *treatment condition* explained by the non shared or shared environment that the participants were matched upon. The GAPSM accounted for all of the variance in the *treatment condition* explained by *a*. The black horizontal line represents a point estimate equivalent to the MZ difference point estimate or an unconfounded PSM point estimate for the specified condition. Bars above the black horizontal line indicate that the GAPSM point estimate was closer to the true point estimate (1.00) than the MZ difference point estimate or the PSM point estimate for the specified condition. Bars below the black horizontal line indicate that the GAPSM point estimate was further away from the true point estimate (1.00) than the MZ difference point estimate or the PSM point estimate for the specified condition.

6.5.3. Third round of thirteen specifications of the treatment condition

Figure 6.5 provides the point estimate comparisons between the five iterations of the GAPSM approach, the MZ difference score approach, and the unconfounded PSM approach for the final 13 specifications[58] of the treatment condition and the point of equivalence. In reference to the GAPSM/MZ difference comparison (Panel A and Panel B), it can be observed that the GAPSM approach created a point estimate equivalent or closer to the true point estimate (1.00) for the majority of the five iterations across the final thirteen specifications. For the first iteration of the GAPSM approach (black bar; participants matched on 25 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively), it can be observed that the GAPSM point estimate was further from the true point estimate than the MZ difference approach across all of the specifications of the treatment condition.

The second iteration (dark gray bars; participants matched on 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) of the of the GAPSM approach produced point estimates equivalent or closer to the true point estimate than that of the MZ difference score approach when genetic factors (*a*) accounted for approximately 30 (thirty-second specification), 35 (thirty-third specification), 40 (thirty-fourth specification), and 45 (thirty-fifth specification) percent of the variance. Similarly, the third (light gray bar; participants matched on 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively), fourth (red bar; participants matched on 50 and 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively), and fifth

[58] The amount of variance in the treatment condition predicted by genetic factors (*a*) increased from .05 to .65 in .05 increments, while the variance in the treatment condition predicted the shared environment (*c*) was approximately three times that of the shared environment (*e*).

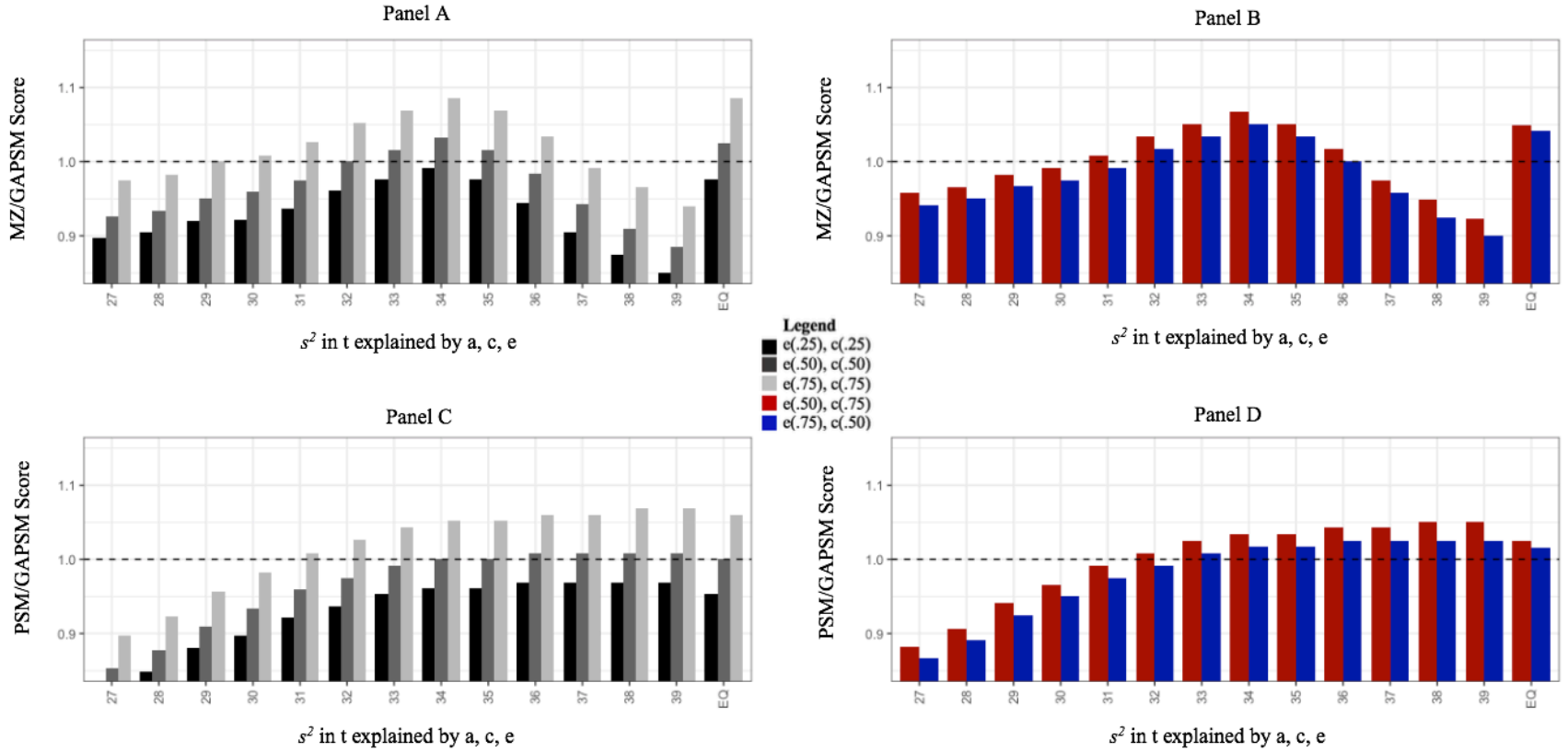
iterations (blue bar; participants matched on 75 and 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) of the GAPSM approach performed better than the MZ difference score approach on six to eight of the thirteen specifications of the treatment condition. Generally, the third, fourth, and fifth iterations of the GAPSM approach got closer to the true point estimate than the MZ difference score approach when both genetic factors (*a*) and the shared environment (*c*) contributed between approximately 75 percent and 84 percent of the variance in the treatment condition.

In reference to the GAPSM/unconfounded PSM comparison (Panel C and Panel D), it can be observed that the GAPSM approach created a point estimate equivalent or closer to the true point estimate (1.00) for the majority of the five iterations across the final 13 specifications. For the first iteration of the GAPSM approach (black bar; participants matched on 25 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively), it can be observed that the GAPSM point estimate was further from the true point estimate than the unconfounded PSM approach across all of the specifications of the treatment condition. Evidence suggested that the second iteration of the GAPSM approach (dark gray bar; participants matched on 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) produced a point estimate generally equivalent or closer to the true point estimate than the point estimate produced by the unconfounded PSM approach on 6 of the 13 specifications of the treatment condition.

The third iteration (light gray bar; participants matched on 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) of the GAPSM approach generally produced a point estimate closer to that of the true point estimate than the unconfounded PSM approach across 9 of the 13 specifications of the treatment condition. Furthermore, this pattern of results was relatively consistent when

comparing the fourth (red bar; participants matched on 50 and 75 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) and fifth (blue bar; participants matched on 75 and 50 percent of the variance in the treatment condition predicted by the non shared environment and shared environment, respectively) iterations of the GAPSM approach to the point estimates derived from an unconfounded PSM approach. Distinct from the third iteration, the fourth and fifth iterations of the GAPSM approach were further from the true point estimate than the unconfounded PSM approach when genetic factors predicted 25 percent of the variance in the treatment condition.

Figure 6.5: Comparing the GAPSM point estimate ($y \sim t$) to the MZ difference score and the unconfounded PSM point estimates ($c = 3*e$).



Notes: The figure presents the final 13 specifications of the treatment condition (indicated by the number) and the point of equivalence. To reiterate, the final 13 iterations increase the variance in the treatment condition predicted by genetic factors (*a*) increased from .05 to .65 in .05 increments, while the variance in the treatment condition predicted the shared environment (*c*) is approximately three times that of the non shared environment (*e*). Four percent (or .04) of the variance in the treatment condition was predicted by the error term (E). EQ represents the point of equivalence. *e* = Non shared environment; *c* = shared environment. For the legend, values in parentheses represents the percentage of the portion of the variance in the *treatment condition* explained by the non shared or shared environment that the participants were matched upon. The GAPSM accounted for all of the variance in the *treatment condition* explained by *a*. The black horizontal line represents a point estimate equivalent to the MZ difference point estimate or an unconfounded PSM point estimate for the specified condition. Bars above the black horizontal line indicate that the GAPSM point estimate was closer to the true point estimate (1.00) than the MZ difference point estimate or the PSM point estimate for the specified condition. Bars below the black horizontal line indicate that the GAPSM point estimate was further away from the true point estimate (1.00) than the MZ difference point estimate or the PSM point estimate for the specified condition.

CHAPTER 7: DISCUSSION

Contemporary social scientists often rely on various statistical techniques to produce unconfounded point estimates of the association between two variables. Propensity score matching (PSM) and MZ difference score modeling represent two of the foremost methodologies employed in such analyses. Although useful methodological techniques, limitations associated with each statistical technique hinder the ability to establish the best unbiased point estimates in certain conditions. *Genetically adjusted propensity score matching (GAPSM)* represents an alternative statistical technique designed to approach the true association between two variables when PSM and MZ difference score modeling cannot. The primary focus of the current study was to propose this alternative methodology for estimating causal associations within the social sciences.

The current chapter will focus on the findings associated with Study 4 (subsection 6.5. of Chapter 6). To reiterate, Study 4 evaluated the validity of the GAPSM approach by deriving post-matching point estimates and comparing them to the post-PSM point estimates and the point estimates produced by MZ difference score modeling. Nevertheless, to briefly summarize the findings associated with studies 1, 2 and 3, the evidence suggested that more stringent methodologies (PSM and MZ difference score analysis) were more likely to produce attenuated associations than the bivariate and multivariate regression models. Furthermore, the findings associated with the PSM and MZ difference score analyses had a higher probability of identifying a null association than the bivariate and multivariate regression models.

7.1. Findings

Three major findings associated with the simulation analyses should be highlighted. First, when the amount of variance in the treatment condition predicted by the non shared environment (*e*) and the shared environment (*c*) were equal to each other (i.e., the first 13 specifications) or

when the non shared environment (*e*) predicted three times the amount of variance of the shared environment (*c*; i.e., the second 13 specifications), four of the GAPSM iterations (i.e., participants matched on Iteration 2: *e* = .50, *c* = .50; Iteration 3: *e* = .75, *c* = .75; Iteration 4: *e* = .50, *c* = .75; Iteration 5: *e* = .75, *c* = .50) appeared to provide point estimates equal to or closer to the true point estimate (1.00) than the MZ difference score model for the majority of the specifications (generally about .05 to .1 closer to the true point estimate than the MZ difference score model). Furthermore, the evidence suggested that the GAPSM approach was generally superior (i.e., post-matching point estimate was closer to the true point estimate) to the MZ difference score model when a larger amount of variance in the treatment condition was predicted by the non shared environment (*e*). This finding is consistent with logic, given that MZ difference scores cannot adjust for the effects of the non shared environment (*e*) without the introduction of additional covariates (Plomin et al., 2013).

Second, when the amount of variance in the treatment condition predicted by the non shared environment (*e*) and the shared environment (*c*) was equal to each other (i.e., the first 13 specifications), when the non shared environment (*e*) predicted three times the amount of variance of the shared environment (*c*; i.e., the second 13 specifications), or when the shared environment (*c*) predicted three times the amount of variance of the non shared environment (*e;* i.e., the third 13 specifications), four of the GAPSM iterations (i.e., participants matched on Iteration 2: *e* = .50, *c* = .50; Iteration 3: *e* = .75, *c* = .75; Iteration 4: *e* = .50, *c* = .75; Iteration 5: *e* = .75, *c* = .50) appeared to provide point estimates equal to or closer to the true point estimate (1.00) than the unconfounded post-PSM point estimate for the majority of the specifications(generally about .05 to .1 closer to the true point estimate than the unconfounded post-PSM point estimate). Furthermore, the evidence suggested that the GAPSM approach was generally superior to the unconfounded post-PSM point estimate when a larger amount of

variance in the treatment condition was predicted by genetic factors (*a*). This finding is consistent with logic, given that unconfounded PSM cannot adjust for the effects of genetic factors (*a*) without the introduction of genetically sensitive covariates (Guo and Fraser, 2015).

Finally, only a limited number of specifications were observed where the first iteration of the GAPSM approach produced point estimates equal to or closer to the true point estimate than the point estimates produced by the MZ difference score model and post-PSM model. The findings suggested that the first iteration of the GAPSM approach ($e = 25$, $c = 25$) produced point estimates generally further from true point estimate than the point estimates derived by the MZ difference score model and post-PSM model. Overall, the findings of the simulation analysis suggest that GAPSM can provide less biased point estimates than an unconfounded PSM and MZ difference score model during various conditions.

7.2. Implications for Social Science Research

The three findings outlined above have broad methodological implications for conducting research within the social sciences. As alluded to in Chapter 1, contemporary social science research often relies on statistical techniques that cannot estimate the causal association between two or more concepts.[59] Social scientists, however, have the statistical tools to more readily estimate the causal association between concepts. As demonstrated by the results of simulation analysis, unconfounded PSM and quantitative genetic techniques (i.e., MZ difference scores) can be used to examine causal associations and approach the true point estimate more readily than bivariate analyses (see Appendix F). Furthermore, MZ difference score analysis, the gold standard in causal estimation, generally approaches the true point estimate more readily than bivariate analyses and an unconfounded PSM (see Appendix F).[60] Given these findings, various

[59] This results from the inability to satisfy all the assumptions associated with said statistical techniques.
[60] Generally, PSM analyses are confounded by unmeasured genetical and environmental influences.

methodologies are already present in the literature to estimate the effects of social/environmental factors (i.e., a treatment condition) on behavioral and health outcomes.

Although the simulation analysis posits favorable findings for the preexisting methodologies, many of the comparisons demonstrated that the GAPSM approach can outperform the preexisting methodologies when certain conditions exist. These findings provide substantive and practical implications concerning future social science research. Regarding the substantive implications, the findings highlight the advantages of studying the association between a treatment condition and an outcome of interest with GAPSM. Through the observation of both genetic and environmental factors, future scholars can approximate the true association between concepts without suffering the limitations associated with PSM and MZ difference score analyses (see Chapter 3). Specifically, GAPSM allows scholars to control for observed genetic factors (a potential limitation associated with PSM), while not relying on generally smaller twin based subsamples (a limitation associated with MZ difference score analyses). Furthermore, similar to PSM, GAPSM can be used across levels of measurement (i.e., nominal, ordinal, interval, or ratio) and is mathematically adaptable to be used in conjunction with various pre-GAPSM estimation techniques (e.g., cross-sectional, hieratical linear modeling, longitudinal modeling) and post-GAPSM evaluation models (e.g., *t*-tests, OLS regression, longitudinal modeling). The quintessential meaning is that GAPSM can potentially serve at the forefront of modern social science research or in a complimentary fashion to statistical methodologies currently employed by social scientists.

Given the substantive implications, two practical implications for social scientists exist to increase the likelihood of conducting GAPSM based research studies in the future. First, given the recent advancements in sociogenomics and the reductions in cost associated with mapping the human genome, social scientists can more readily collect molecular genetic information from

research participants. Although essential to estimating causal associations through the employment of the GAPSM approach, molecular genetic information can be used for additional purposes (e.g., study the link between genes and life outcomes; Dudbridge, 2013). In terms of the GAPSM approach, whole genome data can be used to limit the confounding effects of genetic factors on the estimation of causal associations without requiring a MZ twin sample.

Second, surveys implemented alongside genetic data collection should be expanded to certify that shared and non shared environmental factors that are empirically associated with a large number of treatment conditions are measured. The expansion of survey instruments could potentially increase the amount of variance in a treatment condition observed through the introduction of shared and non shared environmental factors during the estimation of GAPS. The expansion of survey instruments should be guided by prior theoretical and empirical literature identifying the shared and non shared environmental factors predictive of the specified treatment condition. Although pertinent to all social scientists, these practical implications are essential when collecting data designed to address research questions that involve the estimation of social causation at the individual level. The collection of genetic data and rich environmental data, and the use of the GAPSM approach can potentially increase our ability to isolate the causal effects of social conditions on behavioral outcomes at the individual level. Within criminology and criminal justice, the GAPSM approach can be particularly useful.

7.3. Implications for Criminology and Criminal Justice

The identification of conditions in which the post-GAPSM point estimate is equivalent or closer to the true point estimate than an unconfounded post-PSM point estimate and an MZ difference score point estimate has broad implications for criminology and criminal justice. As such, the current subsection will discuss the value of the GAPSM approach to three different

areas of criminological and criminal justice research: theory testing, policy evaluation, treatment evaluation.

7.3.1. GAPSM and testing criminological theories

Germane to empirical assessments of criminological theory is the ability of GAPSM to more accurately approximate the causal association between social concepts than preexisting methodologies (i.e., SSSMs). As such, through the employment of GAPSM scholars can continue to support or refute the claims of various criminological theories using genetically sensitive methodologies. While this has already begun through the employment of MZ difference score methodologies (see Beaver, Vaugh, and DeLisi, 2013; Pike et al., 1996; Nedelec et al., 2017), GAPSM can offer a reevaluation of various criminological hypotheses when employing samples derived from the genetically unrelated individuals (i.e., non-twins).

For instance, given the non-zero heritability estimate for key criminological constructs (e.g., self-control, learning, peer selection; Polderman et al., 2013), scholars can reevaluate the causal effects of these variables on subsequent antisocial behavior. Through the employment of the GAPSM approach, one such reevaluation could estimate the causal effects of self-control on subsequent antisocial behavior independent of the confounding genetic effects. As such, the results could illuminate if self-control has a causal effect (or provide a more accurate point estimate) on antisocial behavior or if common genetic and environmental variants generate the covariance between the two constructs. Similar research studies can be conducted to evaluate if learning and peers have a causal effect on antisocial behavior or if the hypothesized associations are spurious. Nevertheless, these example has yet to be evaluated using the GAPSM approach.

Although supporting and refuting the claims of criminological hypotheses is beneficial, the true value of the GAPSM approach when assessing criminological theories is the approximation of the true effect size between treatment conditions (i.e., predictors of antisocial

behavior) and antisocial behavior. Estimating the true effect size allows scholars to identify the factors that have the strongest influence on antisocial behavior and is important when employing large data sets where a number of statistically significant observations will likely appear. For instance, Pratt and Cullen (2000), using meta-analytical techniques, estimated that the true effect size between self-control and criminal behavior is approximately .20. Nevertheless, these findings were solely based on previous research only employing SSSMs and are likely biased considering that confounding genetic and shared environmental factors were not included in the analyses. These claims have been demonstrated by scholarship using twin based methodologies (see Beaver, Boutwell, and Barnes, 2014; Waller et al., 2018). This example, consistent with most empirical tests of criminological theories, highlights the conceivable possibility that a wide variety of factors contribute to both the theorized predictors of antisocial behavior and antisocial behavior itself (i.e., a self-selection effect). As demonstrated, the GAPSM approach can enhance our understanding of the true effect of the social processes on antisocial behavior beyond contemporary methodologies.

7.3.2. GAPSM and testing criminal justice policies

Similar to testing the theoretical hypotheses of various criminologists, GAPSM can be used to evaluate the effectiveness of criminal justice policies on behavioral outcomes at the individual level. Specifically, micro-level assessments of criminal justice policies can employ GAPSM to evaluate how policies influence a myriad of behavioral outcomes. For example, scholars often claim that solitary confinement, a prison based punishment for inmate misconduct, results in increases in negative psychological outcomes and negative behavioral outcomes (Arrigo and Bullock, 2008; Browne, Cambier, and Agha, 2011; Henderson, 2015; Lucas and Jones, 2017). While these claims have been evaluated with SSSMs and are generally unsubstantiated (see O'Keefe, Klebe, Stucker, Sturm, and Leggett, 2010; O'Keefe et al., 2013;

however see Haney, 2008, 2009), contemporary scholars generally suggest that we do not know the true effects of solitary confinement on negative psychological outcomes or negative behavioral outcomes (Gendreau and Lebreque, 2016). Specifically, the effects of solitary confinement on negative psychological outcomes and negative behavioral outcomes could be confounded by genetic and additional environmental factors. Nonetheless, scholars should be guided by theoretical reasoning to determine the analytical strategy.

While speculative, GAPSM is beneficial during assessments of the psychological harm because adverse psychological effects have a non-zero h^2 (Polderman et al., 2014). Thus, corrections scholars could employ GAPSM to approximate the true association between solitary confinement and the exacerbation of negative psychological status or subsequent antisocial behavior. Although cumbersome, the collection of genetic data could be more accessible than finding twins differentially exposed to criminal justice policies. As demonstrated by the example, the usefulness of the GAPSM approach for the evaluation of criminal justice policies is widespread and can potentially increase our understanding of how criminal justice policies effect individual-level outcomes.

7.3.3. GAPSM and testing treatment effects

In addition to testing theoretical perspectives and evaluating policy effects, GAPSM offers a methodological advancement for scholars interested in individual level treatment effects. As demonstrated by the simulation analysis, GAPSM can be particularly useful during the reassessment of the efficacy of treatment programs on subsequent antisocial behavior (given that approximately 50 percent of the variance in antisocial behavior is heritable; Polderman et al., 2015). Specifically, contemporary evaluations of individual level treatment effects often rely on PSM and post-matching analyses which do not account for genetic predispositions. As such, the observed post-PSM similarities and differences between the treatment and control cases could be

biased by confounding factors. To address the potential confounding effects of genetic predispositions on the efficacy of treatment programs scholars could employ the GAPSM approach. Through the employment of the GAPSM approach, scholars might be able to determine the true effectiveness of various treatment programs and the most effective methods for treating antisocial behavior.

The ability to determine the most effective methods for treating antisocial behavior is exceptionally important to the rehabilitation paradigm. Currently, numerous offender rehabilitation programs, ranging from religion based programs (e.g., Alcoholics Anonymous) to evidence based programs (e.g., Thinking for a Change), claim to effectively reduce antisocial behavior amongst offending populations. Nevertheless, given the stark differences between the numerous rehabilitation programs, some are likely to be more effective at reducing antisocial behavior than others (Lipsey and Cullen, 2007). Furthermore, these effects might vary by group (Lipsey and Cullen, 2007). The GAPSM approach can be used to address these rising concerns within the rehabilitation paradigm, without the reliance on twin based subsamples. Specifically, since exposure to treatment and recidivism likely have non-zero heritability estimates (Polderman et al., 2015), scholars can use the GAPSM approach to control for the observed genetic and environmental factors potentially confounding the association between the treatment condition and recidivism. Exposure to treatment could have a non-zero heritability estimate because previous pro- and anti-social behaviors have been shown to contribute to exposure to treatment during rehabilitation efforts (Lipsey and Cullen, 2007) and both pro- and anti-social behaviors have non-zero heritability estimates (Polderman et al., 2015).

In addition to the proper evaluation of rehabilitation programs, scholars can use the GAPSM approach to estimate the potential causal interaction between various phenotypes and exposure to rehabilitation programs when predicting recidivism.

Upon the completion of treatment evaluations with the GAPSM approach, correctional scholars can assure policy makers that the effectiveness of the specified rehabilitation program was independent of observed genetic and environmental factors.

7.4. Limitations

Although the current dissertation demonstrated the validity of the GAPSM approach four limitations associated with both the analyses and the GAPSM approach overall should be highlighted.

7.4.1. Simulation analyses

While simulation analyses represent the foremost technique for evaluating the validity of a new statistical methodology, these analyses have limitations. At the forefront of these limitations is the inability to estimate every specification of the treatment condition in which comparisons between the GAPSM approach, PSM, and MZ difference score analysis could have been conducted. Specifically, for the sake of brevity, the current evaluation of the GAPSM approach is reliant on comparisons across 40 specifications of a treatment condition. Although 40 specifications of a treatment condition highlight the circumstances in which GAPSM approach appears more stringent than an unconfounded PSM and MZ difference score analysis, an infinite number of specifications were left unexamined. Of those, the most important specifications left unexamined are the ones in which gene-environment interactions have occurred. As noted by scholars (e.g., Plomin et al., 2013), gene-environment interactions likely contribute to self-selection into a treatment condition and subsequent behavioral outcomes. Concerning the GAPSM approach, the current study provided no specifications of a treatment

condition that allowed for the comparison of the GAPSM approach, unconfounded PSM, and MZ difference score analysis when a gene-environment interaction had occurred. Future research should examine how gene-environment interactions influence the observed validity of the GAPSM approach when compared to unconfounded PSM and MZ difference score analysis.

7.4.2. Mathematical proof vs. empirical assessment

Although discussed within chapters three and seven, the current project focused primarily on providing theoretical and empirical evidence for the GAPSM approach rather than evaluating an empirical question with the GAPSM methodology. As such, the analyses are limited in demonstrating a true empirical assessment employing the GAPSM approach. While this could be completed with the Add Health, I favor the evaluation of a new methodology with a proof of concept study rather than a demonstration of an empirical assessment. The primary reason for this favoritism is the ability to fix the true point estimate and provide readers with various comparisons between the new methodology (i.e., GAPSM) and preexisting statistical techniques. Thus, as a proof of concept study the current project provides limited substantive findings and the results should only be interpreted as demonstrating the validity of the GAPSM approach. Future research should employ the GAPSM approach to evaluate meaningful research questions within the social sciences.

7.4.3. Mathematical proof vs. practical applications

Parallel to the preceding limitation, the current dissertation provided a limited discussion of applying the GAPSM approach to future empirical analyses. Again, while the simulation analysis provided readers with various comparisons, in practice the GAPSM approach should be used with caution. Specifically, consistent with concerns surrounding PSM, GAPSM is heavily reliant on the selection of appropriate comparison cases and the mis-selection of comparison cases could result in biased findings. Furthermore, the ability to demonstrate that the data satisfy

the assumptions associated with the GAPSM approach is quite cumbersome and requires scholars to understand the counterfactual logic, PSM, molecular genetics, and polygenic risk scores.

Generally, scholars should conduct a GAPSM analysis using a five-step approach. First, select theoretically and empirically supported comparison cases (e.g., individuals who received the treatment condition vs. individuals with no-exposure to the treatment condition). Second, estimate the polygenic risk score and the propensity score when predicting the treatment condition of interest (in one of the two manners described in Chapter 3). Third, visually and mathematically evaluate distributional structure of the polygenic risk score and propensity score between the treatment and control cases to determine common support. Generally, the visual evaluation can be conducted using overlapping histograms, while the mathematical evaluation can be estimated using interquartile ranges. Fourth, post-matching balance statistics for the polygenic risk score and the environmental covariates should be produced and evaluated. Balance can be assessed through overlapping histograms, interquartile ranges, and mean difference tests. Finally, post-matching evaluations of the causal association between the treatment condition and outcome of interest can be estimated using bivariate mean difference tests or bivariate regression models. Although these steps generally correspond to propensity score matching analyses, more cumbersome pre- and post-matching evaluations can occur if one is employing ordinal or continuous treatment conditions. Furthermore, readings external to the current dissertation could provide additional guidance on various visual and empirical methods that can be used to demonstrate that the data satisfy the assumptions associated with the GAPSM approach (Guo and Fraser, 2015).

7.4.4. Data requirements

The final limitation pertains not to the current study, but to the GAPSM approach overall. Evident by the simulation analyses, the GAPSM approach requires a substantive amount of information pertaining to participants' environmental conditions. For instance, data including, but not limited to, the participants' prenatal care, childhood conditions, peer networks, school environments, neighborhood conditions, and previous behavioral manifestations must be included within the estimation of the GAPS to ensure that enough variance in the treatment condition is predicted by environmental factors. Furthermore, this rich survey information must be supplemented by the collection of whole genome data. While we address some of the limitations of measuring the environment with the independent estimation of the GAPS (i.e., a Bayesian method of ensuring that an appropriate amount of variance is attributed to the environment), studies designed to employ the GAPSM approach must be heavily funded or rely on secondary data. As such, I encourage scholars currently collecting whole genome data to employ rich survey instruments and I encourage scholars currently collecting rich survey data to also collect whole genome data. Eventually, secondary data within the social sciences and social genomics should be able to encourage the use of the GAPSM approach when evaluating the effects of treatment conditions on various outcomes of interest. For instance, Wave 5 of the National Longitudinal Study of Adolescent to Adult Health should allow scholars to employ the GAPSM approach to evaluate various treatment conditions, given that both polygenic risk scores and rich survey data are available to academics.

7.5. Conclusion

Overall, GAPSM represents a potentially useful methodology for social scientists when examining the causal association between a treatment condition or environmental factor and an outcome of interest. To demonstrate this logically structured arguments were made throughout

the current dissertation. Chapter 1 offered an outline of the problem in criminal justice, whereas Chapter 2 and Chapter 3 provided the reader with a detailed understanding of the counterfactual framework and the preexisting methodologies used to address similar or the same problems, respectively. Chapter 3 also highlighted the limitations of the preexisting methodologies. At the conclusion of Chapter 3, the GAPSM approach was introduced to the reader and explained in detail. Chapter 4, Chapter 5, and Chapter 6 afforded a demonstration of the preexisting methodologies with a real dataset and an evaluation of the validity of the GAPSM approach when compared to the preexisting methodologies. Overall the findings of the evaluation demonstrate the conditions in which the GAPSM methodology approaches the true point estimate closer than post-PSM models and MZ difference score analyses. These findings support the postulation that the GAPSM methodology should be used in certain conditions, while unconfounded PSM and MZ difference score analyses should be used in other conditions. In closing, and perhaps most importantly, GAPSM represents another tool social scientists can use to conduct rigorous genetically sensitive examinations of the etiological influence of environmental factors on human behavior.

REFERENCES

Agerbo, E., Sullivan, P. F., Vilhjálmsson, B. J., Pedersen, C. B., Mors, O., Børglum, A. D., ... & Ripke, S. (2015). Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: a Danish population-based study and meta-analysis. *JAMA Psychiatry*, *72*, 635-641.

Agnew, R. (1992). Foundation for a general strain theory of crime and delinquency. *Criminology*, *30*, 47-88.

Agnew, R. (2006). General strain theory: Current status and directions for further research. In *Taking Stock: The Status of Criminological Theory* (pp. 101-123) F. T. Cullen, Wright, J., and Blevins, K. (Eds.), New Brunswick, NJ: Transaction Publishers.

Agrawal, A. F. (2001). Sexual selection and the maintenance of sexual reproduction. *Nature*, *411*, 692-695.

Akers, R. L. (2011). *Social learning and social structure: A general theory of crime and deviance*. Piscataway, NJ: Transaction Publishers.

Alcock, J. (2009). *Animal behavior: An evolutionary approach*. Sunderland, MA: Sinauer Associates.

Apel, R. J., & Sweeten, G. (2010). Propensity score matching in criminology and criminal justice. In A. R. Piquero & D. Weisburd (Eds.) *Handbook of quantitative criminology* (pp. 543-562). New York, NY: Springer.

Arrigo, B. A., & Bullock, J. L. (2008). The psychological effects of solitary confinement on prisoners in supermax units: Reviewing what we know and recommending what should change. *International Journal of Offender Therapy and Comparative Criminology*, *52*, 622-640.

Asbury, K., Dunn, J. F., Pike, A., & Plomin, R. (2003). Nonshared environmental influences on individual differences in early behavioral development: A monozygotic twin differences study. *Child Development*, *74*, 933-943.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*, 399-424.

Bäckman, O. (2017). High school dropout, resource attainment, and criminal convictions. *Journal of Research in Crime and Delinquency*, *54*, 715-749.

Bailey, J. M., Dunne, M. P., & Martin, N. G. (2000). Genetic and environmental influences on sexual orientation and its correlates in an Australian twin sample. *Journal of Personality and Social Psychology*, *78*, 524-536.

Barnes, J. C., Beaver, K. M., & Boutwell, B. B. (2011). Examining the genetic underpinnings to Moffitt's developmental taxonomy: A behavioral genetic analysis. *Criminology*, *49*, 923-954.

Barnes, J. C., Golden, K., Mancini, C., Boutwell, B. B., Beaver, K. M., & Diamond, B. (2014). Marriage and involvement in crime: A consideration of reciprocal effects in a nationally representative sample. *Justice Quarterly*, *31*, 229-256.

Barnes, J. C., Wright, J. P., Boutwell, B. B., Schwartz, J. A., Connolly, E. J., Nedelec, J. L., & Beaver, K. M. (2014). Demonstrating the validity of twin research in criminology. *Criminology*, *52*, 588-626.

Beaver, K. M. (2008). Nonshared environmental influences on adolescent delinquent involvement and adult criminal behavior. *Criminology*, *46*, 341-369.

Beaver, K. M. (2009). *Biosocial criminology: A primer*. Dubuque, IA: Kendall Hunt Publishing Company.

Beaver, K. M., Al-Ghamdi, M. S., Kobeisy, A. N., Alqurashi, F. H., Connolly, E. J., & Schwartz, J. A. (2016). The association between psychopathic personality traits and victimization and exposure to violence in a sample of Saudi Arabian youth. *Psychiatric Quarterly*, *87*(2), 217-228.

Beaver, K. M., Boutwell, B. B., & Barnes, J. C. (2014). Social support or biosocial support? A genetically informative analysis of social support and its relation to self-control. *Criminal Justice and Behavior*, *41*, 453-470.

Beaver, K. M., Eagle Schutt, J., Boutwell, B. B., Ratchford, M., Roberts, K., & Barnes, J. C. (2009). Genetic and environmental influences on levels of self-control and delinquent peer affiliation: Results from a longitudinal sample of adolescent twins. *Criminal Justice and Behavior*, *36*, 41-60.

Beaver, K. M., Vaughn, M. G., & DeLisi, M. (2013). Nonshared environmental effects on adulthood psychopathic personality traits: Results from a monozygotic twin difference scores analysis. *Psychiatric Quarterly*, *84*, 381-393.

Beaver, K. M., Wright, J. P., DeLisi, M., Walsh, A., Vaughn, M. G., Boisvert, D., & Vaske, J. (2007). A gene× gene interaction between DRD2 and DRD4 is associated with conduct disorder and antisocial behavior in males. *Behavioral and Brain Functions*, *3*, 30.

Bia, M., & Mattei, A. (2008). A Stata package for the estimation of the dose–response function through adjustment for the generalized propensity score. *The Stata Journal*, *8*, 354-373.

Blomberg, T. G., Bales, W. D., & Piquero, A. R. (2012). Is educational achievement a turning point for incarcerated delinquents across race and sex?. *Journal of Youth and Adolescence*, *41*, 202-216.

Boardman, J. D., & Fletcher, J. M. (2015). To cause or not to cause? That is the question, but identical twins might not have all of the answers. *Social Science & Medicine*, *127*, 198-200.

Boehnke, M. (1991). Allele frequency estimation from data on relatives. *American Journal of Human Genetics*, *48*, 22-25.

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics*, *3*, 872-882.

Braga, A. A., Piehl, A. M., & Hureau, D. (2009). Controlling violent offenders released to the community: An evaluation of the Boston Reentry Initiative. *Journal of Research in Crime and Delinquency*, *46*, 411-436.

Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social Forces*, *92*, 109-140.

Bridges, C. B. (1916). Non-disjunction as proof of the chromosome theory of heredity (concluded). *Genetics*, *1*, 107-163.

Browne, A., Cambier, A., & Agha, S. (2011). *Prisons Within Prisons: The Use of Segregation in the United States*. New York: Vera Institute of Justice.

Burt, S. A., McGue, M., & Iacono, W. G. (2009). Nonshared environmental mediation of the association between deviant peer affiliation and adolescent externalizing behaviors over time: Results from a cross-lagged monozygotic twin differences design. *Developmental Psychology*, *45*, 1752-1760.

Bush, W. S., & Moore, J. H. (2012). Genome-wide association studies. *PLoS Computational Biology*, *8*, e1002822.

Campbell, B. C., Leslie, P. W., Little, M. A., & Campbell, K. L. (2005). Pubertal timing, hormones, and body composition among adolescent Turkana males. *American Journal of Physical Anthropology*, *128*, 896-905.

Carey, G. (2003). *Human Genetics for the Social Sciences*. London, UK: Sage Publications.

Caspi, A., Moffitt, T. E., Morgan, J., Rutter, M., Taylor, A., Arseneault, L., ... & Polo-Tomas, M. (2004). Maternal expressed emotion predicts children's antisocial behavior problems: using monozygotic-twin differences to identify environmental effects on behavioral development. *Developmental Psychology*, *40*, 149-161.

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science*, *24*, 304-312.

Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S. J., & Park, J. H. (2013). Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nature Genetics*, *45*, 400-405.

Christian, J. C., Kang, K. W., & Norton Jr, J. J. (1974). Choice of an estimate of genetic variance from twin data. *American Journal of Human Genetics*, *26*, 154-161.

Cloward, R. A., & Lloyd, E. (1960). Ohlin. *Delinquency and Opportunity*: *A theory of delinquent gangs*. New York, NY: Free press.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*, 155-159.

Coyne, M. A., & Wright, J. P. (2014). The stability of self-control across childhood. *Personality and Individual Differences*, *69*, 144-149.

Crow, J. F. (1994). Advantages of sexual reproduction. *genesis*, *15*, 205-213.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favored races in the struggle for life*. London, UK: John Murray.

Darwin, C. (1868). *The variation of animals and plants under domestication*. London, UK: John Murray.

Davies, G., Marioni, R. E., Liewald, D. C., Hill, W. D., Hagenaars, S. P., Harris, S. E., ... & Cullen, B. (2016). Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N= 112 151). *Molecular Psychiatry*, *21*, 758-767.

Davies, G., Armstrong, N., Bis, J. C., Bressler, J., Chouraki, V., Giddaluru, S., ... & Van Der Lee, S. J. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N= 53 949). *Molecular Psychiatry*, *20*, 183-192.

Dawkins, R. (1976). *The selfish gene*. Oxford, UK: Oxford university press.

Dawson, E., Abecasis, G. R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D. M., ... & Carter, D. (2002). A first-generation linkage disequilibrium map of human chromosome 22. *Nature*, *418*, 544-548.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*, *84*, 151-161.

DeLisi, M., Barnes, J. C., Beaver, K. M., & Gibson, C. L. (2009). Delinquent gangs and adolescent victimization revisited: A propensity score matching approach. *Criminal Justice and Behavior*, *36*, 808-823.

Demmler, J. C., Hill, R. A., Rahman, M. A., Bandyopadhyay, A., Healy, M. A., Paranjothy, S., ... & Lyons, R. A. (2017). Educational Attainment at Age 10–11 Years Predicts Health Risk Behaviors and Injury Risk During Adolescence. *Journal of Adolescent Health*, *61*, 212-218.

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics*, *9*, 403-411.

Derks, E. M., Vorstman, J. A., Ripke, S., Kahn, R. S., Ophoff, R. A., & Schizophrenia Psychiatric Genomic Consortium. (2012). Investigation of the genetic association between quantitative measures of psychosis and schizophrenia: a polygenic risk score analysis. *PloS One*, *7*, e37852 1-7.

Devenish, B., Hooley, M., & Mellor, D. (2017). The pathways between socioeconomic status and adolescent outcomes: a systematic review. *American Journal of Community Psychology*, *59*, 219-238.

Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature*, *388*, 468-471.

Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, *95*(3), 932-945.

Dingemanse, N. J., Both, C., Drent, P. J., & Tinbergen, J. M. (2004). Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society B: Biological Sciences*, *271*, 847-852.

Dobzhansky, T. (1937) *Genetics and the Origin of Species.* New York, NY: Columbia University Press.

Domingue, B. W., Belsky, D. W., Fletcher, J. M., Conley, D., Boardman, J. D., and Harris, K. M. (2018). The social genome of friends and schoolmates in the national longitudinal study of adolescent to adult health. *PNAS, online first,* 1-6.

Domingue, B. W., Liu, H., Okbay, A., & Belsky, D. W. (2017). Genetic heterogeneity in depressive symptoms following the death of a spouse: Polygenic score analysis of the US Health and Retirement Study. *American Journal of Psychiatry, 174,* 963-970.

Draper, N. R., & Smith, H. (2014). *Applied regression analysis*. Hoboken, NJ: John Wiley & Sons.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS genetics*, *9*, e1003348, 1-17.

Duwe, G., & Clark, V. (2014). The effects of prison-based educational programming on recidivism and employment. *The Prison Journal*, *94*, 454-478.

Einstein A. (1916), *Relativity: The Special and General Theory* (Translation 1920), New York: H. Holt and Company

Elliott, D. S., & Voss, H. L. (1974). *Delinquency and dropout.* Bethesda, MD: National Institute of Mental Health (DHEW).

Eldredge, N. (1985). *Unfinished synthesis: biological hierarchies and modern evolutionary thought*. Oxford, UK: Oxford University Press.

Emlen, S. T., & Oring, L. W. (1977). Ecology, sexual selection, and the evolution of mating systems. *Science*, *197*, 215-223.

Endler, J. A. (1986). *Natural selection in the wild* (No. 21). Princeton, NJ: Princeton University Press.

Evans, D. M., Brion, M. J. A., Paternoster, L., Kemp, J. P., McMahon, G., Munafò, M., ... & Pourcain, B. S. (2013). Mining the human phenome using allelic scores that index biological intermediates. *PLoS Genetics*, *9*, e1003919, 1-15.

Evans, J. R., & Olson, D. L. (2001). *Introduction to simulation and risk analysis*. Upper Saddle River, NJ: Prentice Hall PTR.

Evans, D. M., Visscher, P. M., & Wray, N. R. (2009). Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Human Molecular Genetics*, *18*, 3525-3531.

Falconer, D. S. & McKay, T. F. (1996). *Introduction to Quantitative Genetics*. New York, NY: Prentice Hall.

Farrington, D. P. (2005). Childhood origins of antisocial behavior. *Clinical Psychology & Psychotherapy*, *12*, 177-190.

Felson, J. (2014). What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. *Social Science Research*, *43*,184–99.

Fisher, R. A. (1919). XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, *52*, 399-433.

Fisher, R. A. (1999). *The genetical theory of natural selection: a complete variorum edition*. Oxford, UK: Oxford University Press.

Floderus-Myrhed, B., Pedersen, N., & Rasmuson, I., (1980). Assessment of heritability for personality, based on a short-form of the Eysenck Personality Inventory: A study of 12,898 twin pairs. *Behavior Genetics, 10*, 153–62.

Forrest, W. (2014). Cohabitation, relationship quality, and desistance from crime. *Journal of Marriage and Family*, *76*, 539-556.

Fox, J. (2016). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage Publications, Inc.

French, L., Gray, C., Leonard, G., Perron, M., Pike, G. B., Richer, L., ... & Banaschewski, T. (2015). Early cannabis use, polygenic risk score for schizophrenia and brain maturation in adolescence. *JAMA Psychiatry*, *72*, 1002-1011.

Fujiwara, T., & Kawachi, I. (2008). Social capital and health: a study of adult twins in the US. *American Journal of Preventive Medicine*, *35*, 139-144.

Geary, D. C. (1998). *Male, female: The evolution of human sex differences*. Washington, DC: American Psychological Association.

Geary, D. C. (2000). Evolution and proximate expression of human paternal investment. *Psychological Bulletin*, *126*, 55-77.

Gendreau, P and Labrecue, R. M. (2016). The effects of administrative segregation: A lesson in knowledge cumulation. In (John D. Wooldredge and Paula Smith, eds.), *The Oxford Handbook of Prisons and Imprisonment* (Online). Oxford, U.K.: Oxford University Press.

Glazier, A. M., Nadeau, J. H., & Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science*, *298*, 2345-2349

Gluckman, P. D., & Hanson, M. A. (2006). Evolution, development and timing of puberty. *Trends in Endocrinology and Metabolism*, *17*, 7-12.

Gonzalez, J. M. R., Salas-Wright, C. P., Connell, N. M., Jetelina, K. K., Clipper, S. J., & Businelle, M. S. (2016). The long-term effects of school dropout and GED attainment on substance use disorders. *Drug & Alcohol Dependence*, *158*, 60-66.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.

Gray, A. C., Durkin, K. F., Call, J. T., Evans, H. J., & Melton, W. (2015). Differential association and marijuana use in a juvenile drug court sample. *Applied Psychology in Criminal Justice*, *11*, 1-8.

Guo, S., & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and Applications (Second edition)*. Thousand Oaks, CA: Sage Publishing Inc.

Hagan, J., & Parker, P. (1999). Rebellion beyond the classroom: A life-course capitalization theory of the intergenerational causes of delinquency. *Theoretical Criminology*, *3*, 259-285.

Hamshere, M. L., O'Donovan, M. C., Jones, I. R., Jones, L., Kirov, G., Green, E. K., ... & Gurling, H. (2011). Polygenic dissection of the bipolar phenotype. *The British Journal of Psychiatry*, *198*, 284-288.

Haney, C. (2008). A culture of harm: Taming the dynamics of cruelty in supermax prisons. *Criminal Justice and Behavior*, *35*, 956-984.

Haney, C. (2009). The social psychology of isolation: Why solitary confinement is psychologically harmful. *Prison Service Journal*, *181*, 12-20.

Harding, D. J. (2003). Counterfactual models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy. *American Journal of Sociology*, *109*, 676-719.

Harris, J. R. (2011). *The nurture assumption: Why children turn out the way they do*. New York, NY: Simon and Schuster.

Heckman, J. (1990a). Varieties of selection bias. *The American Economic Review*, *80*, 313-318.

Heckman, J. J. (1990b). Selection bias and self-selection. In: Durlauf S.N., Blume L.E. (Eds.) Microeconometrics. The New Palgrave Economics Collection (pp. 201-224). Palgrave Macmillan, London.

Henderson, G. (2015). Disciplinary Segregation: How the Punitive Solitary Confinement Policy in Federal Prisions Violates the Due Process Clause of the Fifth Amendment in Spite of Sandin V. Conner. *Marquette Law Review, 99*, 477.

Henly, S. J. (1993). Robustness of some estimators for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *46*, 313-338.

Hetherington, E. M., Reiss, D., & Plomin, R. (Eds.). (2013). *Separate social worlds of siblings: The impact of nonshared environment on development*. Abingdon, UK: Routledge.

Herrnstein, R. J., and Murray, C. (1994). *The bell curve: The reshaping of American life by differences in intelligence.* New York, NY: Free Press Paperbacks.

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics*, *4*, e1000008, 1-10.

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. In Gelman, A., & Meng, X., (Eds.) *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (pp. 73-84), Hoboken, NJ: John Wiley & Sons.

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, *6*, 95-108.

Hirschi, T., and Hindelang, M. J. (1977). Intelligence and delinquency: A revisionist review. *American Sociological Review, 42*, 571–587.

Holland, P. W., & Rubin, D. B. (1987). Causal inference in retrospective studies. *ETS Research Report Series*, *1987*, 203-231.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, *87*, 706-710.

Intravia, J., Wolff, K. T., Gibbs, B. R., & Piquero, A. R. (2017). Violent Attitudes and Antisocial Behavior: Examining the Code of the Street's Generalizability among a College Sample. *Deviant Behavior*, *38*, 957-974.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, *2*, e124, 1-6.

Isik, F., Li, B., & Frampton, J. (2003). Estimates of additive, dominance and epistatic genetic variances from a clonally replicated test of loblolly pine. *Forest Science*, *49*, 77-88.

Jackson, F. (1977). A causal theory of counterfactuals. *Australasian Journal of Philosophy*, *55*, 3-21.

Jang, K. L., Livesley, W. J., & Vemon, P. A. (1996). Heritability of the big five personality dimensions and their facets: a twin study. *Journal of Personality*, *64*, 577-592.

Jansen, P. R., Polderman, T. J., Bolhuis, K., Ende, J., Jaddoe, V. W., Verhulst, F. C., ... & Tiemeier, H. (2018). Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population. *Journal of Child Psychology and Psychiatry*, *59*, 39-47.

Joffe, M. M., & Rosenbaum, P. R. (1999). Invited commentary: propensity scores. *American journal of Epidemiology*, *150*, 327-333.

Johnson, W., McGue, M., Krueger, R. F., & Bouchard Jr, T. J. (2004). Marriage and personality: a genetic analysis. *Journal of Personality and Social Psychology*, *86*, 285-294.

Johnstone, R. A. (1994). Female preference for symmetrical males as a by-product of selection for mate recognition. *Nature*, *372*, 172-175.

Kaplan, D. 2009. *Structural Equation Modeling: Foundations and Extensions*. Thousand Oaks, CA: Sage.

Kim, R. H., & Clark, D. (2013). The effect of prison-based college education programs on recidivism: Propensity score matching approach. *Journal of Criminal Justice*, *41*, 196-204.

King, R. D., Massoglia, M., & Macmillan, R. (2007). The context of marriage on crime: Gender, the propensity to marry, and offending in early adulthood. *Criminology, 45*, 33-65.

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., ... & Beerli, P. (2001). The strength of phenotypic selection in natural populations. *The American Naturalist*, *157*, 245-261.

Kun, A., & Weiner, B. (1973). Necessary versus sufficient causal schemata for success and failure. *Journal of Research in Personality*, *7*, 197-207.

Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, *37*, 1210-1226.

Larzelere, R. E., & Patterson, G. R. (1990). Parental management: Mediator of the effect of socioeconomic status on early delinquency. *Criminology*, *28*, 301-324.

Laub, J. H., & Sampson, R. J. (1993). Turning points in the life course: Why change matters to the study of crime. *Criminology*, *31*, 301-325.

Laub, J. H., & Sampson, R. J. (2003). *Shared beginnings: Divergent Lives*. Cambridge, MA: Harvard University Press.

Lavrakas, P. J. (2012). *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publishing.

Levine, D. I., & Painter, G. (2003). The schooling costs of teenage out-of-wedlock childbearing: Analysis with a within-school propensity-score-matching estimator. *The Review of Economics and Statistics*, *85*, 884-900.

Lewis, D. (1974). Causation. *The journal of philosophy*, *70*, 556-567.

Lewis, D. (1979). Counterfactual Dependence and Time's Arrow. *Nous*, *13*, 455-476.

Lewis, P. W. A., & McKenzie, E. (2017). *Simulation Methodology for Statisticians, Operations Analysts, and Engineers*. Boca Raton, FL: CRC Press.

Lilly, J. R., Cullen, F. T., & Ball, R. A. (2010). *Criminological theory: Context and consequences*. Thousand Oaks, CA: Sage Publishing

Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Science.*, *3*, 297-320.

Littvay, L. (2012). Do heritability estimates of political phenotypes suffer from an equal environment assumption violation? Evidence from an empirical study. *Twin Research and Human Genetics, 15*, 6–14.

Loehlin, J. C. (1992). *Sage series on individual differences and development, Vol. 2. Genes and environment in personality development.* Thousand Oaks, CA: Sage Publications.

Lori, A., Avramopoulos, D., Wang, A., Goes, F. S., Mulle, J., Powers, A., ... & Nestadt, G. (2017). 917-Genome-Wide Association Study (GWAS) of Toxoplasma Gondii Infection and Evaluation of Schizophrenia Risk by Using a Polygenic Risk Score (PRS). *Biological Psychiatry*, *81*, S371.

Lucas, J. W., & Jones, M. A. (2017). An analysis of the deterrent effects of disciplinary segregation on institutional rule violation rates. *Criminal Justice Policy Review*, 0887403417699930.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits* (Vol. 1). Sunderland, MA: Sinauer Associates.

Makarios, M., Cullen, F. T., & Piquero, A. R. (2017). Adolescent criminal behavior, population heterogeneity, and cumulative disadvantage: untangling the relationship between adolescent delinquency and negative outcomes in emerging adulthood. *Crime & Delinquency*, *63*, 683-707.

Marée, A. F., Keulen, W., Boucher, C. A., & De Boer, R. J. (2000). Estimating relative fitness in viral competition experiments. *Journal of Virology*, *74*, 11067-11072.

Mason, D. A., & Frick, P. J. (1994). The heritability of antisocial behavior: A meta-analysis of twin and adoption studies. *Journal of Psychopathology and Behavioral Assessment*, *16*, 301-323.

Mayr, E. (1972). Sexual selection and natural selection. In Campbell, Bernard G. (Ed) *Sexual selection and the descent of man: The darwinian pivot* (pp. 87-104). New Brunswick, NJ: Transaction Publishers.

Mayr, E. (1991). *One long argument: Charles Darwin and the genesis of modern evolutionary thought* (Vol. 2). Cambridge, MA: Harvard University Press.

Mayr, E. (November 24, 2009). Darwin's Influence on Modern Thought. *Scientific America.* Retrieved from: https://www.scientificamerican.com/article/darwins-influence-on-modern-thought/

Mayr, E., & Provine, W. B. (Eds.). (1998). *The evolutionary synthesis: perspectives on the unification of biology*. Cambridge, MA: Harvard University Press.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*, *9*, 356-369.

McCarty, M., Macleod, C. M., & Avery, O. T. (1946). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *Journal of Experimental Medicine*, *83*, 97-104.

McCormick, M. P., O'Connor, E. E., Cappella, E., & McClowry, S. G. (2013). Teacher–child relationships and academic achievement: A multilevel propensity score model approach. *Journal of School Psychology*, *51*, 611-624.

McCutcheon, A. L. (1987). *Latent class analysis* (No. 64). Thousand Oaks, CA: Sage Publishing

McGue, M., Osler, M., & Christensen, K. (2010). Causal inference and observational research: The utility of twins. *Perspectives on Psychological Science*, *5*, 546-556.

Mears, D. P., & Cochran, J. C. (2013). What is the effect of intelligence on offending?. *Criminal Justice and Behavior*, *40*, 1280-1300.

Mears, D. P., Cochran, J. C., & Beaver, K. M. (2013). Self-control theory and nonlinear effects on offending. *Journal of Quantitative Criminology*, *29*, 447-476.

Mears, D. P., Cochran, J. C., Siennick, S. E., & Bales, W. D. (2012). Prison visitation and recidivism. *Justice Quarterly*, *29*, 888-918.

Melguizo, T., Kienzl, G. S., & Alfonso, M. (2011). Comparing the educational attainment of community college transfer students and four-year college rising juniors using propensity score matching methods. *The Journal of Higher Education*, *82*, 265-291.

Mendel, G. (1866). *Experiments in plant hybridization (English Translation)* Edited by J.H. Bennett (1965). Edinburgh, Scotland: Oliver and Boyd Publishers.

Merton, R. K. (1938). Social structure and anomie. *American sociological review*, *3*, 672-682.

Merton, R. K. (1968). *Social theory and social structure*. New York, NY: Simon and Schuster.

Middeldorp, C. M., De Moor, M. H. M., McGrath, L. M., Gordon, S. D., Blackwood, D. H., Costa, P. T., ... & Tanaka, T. (2011). The genetic association between personality and major depression or bipolar disorder. A polygenic score analysis using genome-wide association data. *Translational Psychiatry*, *1*, e50, 1-8.

Miller, H. V., Barnes, J. C., & Beaver, K. M. (2011). Self-control and health outcomes in a nationally representative sample. *American journal of health behavior*, *35*, 15-27.

Miller, W. E. (1999). Temporal Order and Causal Inference. *Political Analysis*, *8*, 119-146.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychological Review*, *100*, 674-701.

Moffitt, T. E. (1997). Adolescence-limited and life-course-persistent offending: A complementary pair of developmental theories. *Developmental Theories of Crime and Delinquency*, *7*, 11-54.

Moffitt, T. E. (2006). *A review of research on the taxonomy of life-course persistent versus adolescence-limited antisocial behavior* (Vol. 15, p. 277). F. T. Cullen, J. P. Wright, & K. R. Blevins (Eds.). *Taking Stock: The Status of Criminological Theory.* New Brunswick, NJ: Transaction Publishers

Moffitt, T. E. (2007). A review of research on the taxonomy of life-course-persistent versus adolescent-limited antisocial behavior. In D.J. Flannery, A. T. Vazsonyi, and I. D. Waldman (Eds.), *The Cambridge Handbook of Violent Behavior and Aggression* (pp. 44-76). New York, NY: Cambridge University Press.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821). Hoboken, NJ: John Wiley & Sons.

Morgan, P. L., Frisco, M. L., Farkas, G., & Hibel, J. (2010). A propensity score matching analysis of the effects of special education services. *The Journal of Special Education*, *43*, 236-254.

Morgan, T. H. (1915). *The mechanism of Mendelian heredity*. New York, NY: Henry Holt and Company.

Morgan, T. H. (1917). The theory of the gene. *The American Naturalist*, *51*, 513-544.

Morgan, S. L., & Winship, C. (2015). Counterfactuals and causal inference: Methods and principles for social research (2nd Ed.). New York: Cambridge University Press.

Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S., & Cheung, V. G. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature*, *430*, 743-747.

Mukherjee, S. (2017). *The gene: An intimate history*. New York, NY: Simon and Schuster.

Neale, M. C. C. L., & Cardon, L. R. (2013). *Methodology for genetic studies of twins and families* (Vol. 67). Berlin, Germany: Springer Science & Business Media.

Nedelec, J. L., & Beaver, K. M. (2014). Physical attractiveness as a phenotypic marker of health: an assessment using a nationally representative sample of American adults. *Evolution and Human Behavior*, *35*, 456-463.

Nedelec, J. L., Park, I., & Silver, I. A. (2016). The effect of the maturity gap on delinquency and drug use over the life course: A genetically sensitive longitudinal design. *Journal of Criminal Justice*, *47*, 84-99.

Nedelec, J. L., Richardson, G., & Silver, I. A. (2017). Religiousness, Spirituality, and Substance Use: A Genetically Sensitive Examination and Critique. *Journal of Drug Issues*, *47*, 340-355.

Nedelec, J. L., & Silver, I. A. (2018). Challenging assumptions: A genetically sensitive assessment of the criminogenic effect of contact with the criminal justice system. *Journal of Contemporary Criminal Justice*, *35,* 69-86.

Nedelec, J. L., Schwartz, J. A., Connolly, E. J., & Beaver, K. M. (2012). Exploring the association between IQ and differential life outcomes: results from a longitudinal sample of monozygotic twins. *Temas em Psicologia*, *20*, 31-44.

Newton, I. (1687). *Mathematical principles of natural philosophy*. London.

Nye, F. I., Short Jr, J. F., & Olson, V. J. (1958). Socioeconomic status and delinquent behavior. *American Journal of Sociology*, *63*, 381-389.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., ... & Oskarsson, S. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*, 539-542.

O'Keefe, M. L., Klebe, K. J., Metzner, J., Dvoskin, J., Fellner, J., & Stucker, A. (2013). A longitudinal study of administrative segregation. *Journal of the American Academy of Psychiatry and the Law Online*, *41*, 49-60.

O'Keefe, M. L., Klebe, K. J., Stucker, A., Sturm, K., & Leggett, W. (2010). *One year longitudinal study of the psychological effects of administrative segregation*. Colorado Springs: Colorado Department of Corrections, Office of Planning and Analysis.

Orr, H. A. (2007). Absolute fitness, relative fitness, and utility. *Evolution*, *61*, 2997-3000.

Orr, H. A. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, *10*, 531-539.

Palla, L., & Dudbridge, F. (2015). A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *The American Journal of Human Genetics*, *97*, 250-259.

Paul, S. M. (1980). Sibling resemblance in mental ability: A review. *Behavior Genetics*, *10*, 277–290.

Pederson, N. L., McClearn, G. E., Plomin, R., and Friberg. L. (1985). Separated fraternal twins: Resemblance for cognitive abilities. *Behavior Genetics*, 15, 407–419.

Pike, A., Reiss, D., Hetherington, E. M., & Plomin, R. (1996). Using MZ differences in the search for nonshared environmental effects. *Journal of Child Psychology and Psychiatry*, *37*, 695-704.

Pinker, S. (2003). *The blank slate: The modern denial of human nature*. London, UK: Penguin.

Piotrowska, P. J., Stride, C. B., Croft, S. E., & Rowe, R. (2015). Socioeconomic status and antisocial behaviour among children and adolescents: A systematic review and meta-analysis. *Clinical Psychology Review*, *35*, 47-55.

Piquero, A. R., Farrington, D. P., & Blumstein, A. (2007). *Key issues in criminal career research: New analyses of the Cambridge Study in Delinquent Development*. Cambridge, UK: Cambridge University Press.

Plomin, R. (2011). Commentary: Why are children in the same family so different? Nonshared environment three decades later. *International Journal of Epidemiology*, *40*, 582–592.

Plomin, R. (2013). Commentary: Missing heritability, polygenic scores, and gene–environment correlation. *Journal of Child Psychology and Psychiatry*, *54*, 1147-1149.

Plomin, R., and Bergeman, C. S. (1991). The nature of nurture: Genetic influence on "environmental" measures. *Behavioral and Brain Sciences, 14*, 373–386.

Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*, 702-709.

Plomin, R., Chipuer, H. M., & Neiderhiser, J. M. (1994). Behavior genetic evidence for the importance of nonshared environment. In E. M. Hetherington, D. Reiss, & R. Plomin (Eds.), *Separate social worlds of siblings: The impact of nonshared environment on development*. Hillsdale, NJ: Erlbaum.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderheiser, J. (2013). *Behavioral genetics*. Basingstoke, UK: Palgrave Macmillan.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, *84*, 309-322.

Pompoco, A., Wooldredge, J., Lugo, M., Sullivan, C., & Latessa, E. J. (2017). Reducing inmate misconduct and prison returns with facility education programs. *Criminology & Public Policy*, *16*, 515-547.

Power, R. A., Steinberg, S., Bjornsdottir, G., Rietveld, C. A., Abdellaoui, A., Nivard, M. M., ... & Cesarini, D. (2015). Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nature Neuroscience*, *18*, 953-955.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*, 904-909.

Puig-Butille, J. A., Gimenez-Xavier, P., Visconti, A., Nsengimana, J., Garcia-García, F., Tell-Marti, G., ... & Dopazo, J. (2017). Genomic expression differences between cutaneous cells from red hair color individuals and black hair color individuals based on bioinformatic analysis. *Oncotarget*, *8*, 11589-11599.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559-575.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'donovan, M. C., Sullivan, P. F., ... & O'Dushlaine, C. T. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*, 748-752.

Qvarnström, A., & Price, T. D. (2001). Maternal effects, paternal effects and sexual selection. *Trends in Ecology & Evolution*, *16*, 95-100.

Raine, A. (2013). *The anatomy of violence: The biological roots of crime*. New York, NY: Vintage.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., ... & Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature*, *411*, 199-204.

Restivo, E., & Lanier, M. M. (2015). Measuring the contextual effects and mitigating factors of labeling theory. *Justice Quarterly*, *32*, 116-141.

Rice, J. C., Cloninger, R., and Reich, T. (1980). Analysis of behavioral traits in the presence of cultural transmission and assortative mating: Applications to IQ and SES. *Behavior Genetics 10*, 73–92.

Ridley, M. (1994). *The red queen: Sex and the evolution of human nature*. London, UK: Penguin Publishers.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., ... & Albrecht, E. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, *340*, 1467-1471.

Rijsdijk, F. V., & Sham, P. C. (2002). Analytic approaches to twin data using structural equation models. *Briefings in Bioinformatics*, *3*, 119-133.

Risch, N., Herrell, R., Lehner, T., Liang, K. Y., Eaves, L., Hoh, J., ... & Merikangas, K. R. (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA*, *301*, 2462-2471.

Roach, J. C., Glusman, G., Smit, A. F., Huff, C. D., Hubley, R., Shannon, P. T., ... & Shendure, J. (2010). Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science*, *328*, 636-639.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, *121*, 133-148.

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

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association*, *79*, 516-524.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*, 33-38.

Rowe, D. C. (1983). Biometrical genetic models of self-reported delinquent behavior: A twin study. *Behavior Genetics*, *13*, 473-489.

Rowe, D. C., & Plomin, R. (1981). The importance of nonshared (E_1) environmental influences in behavioral development. *Developmental Psychology*, *17*, 517-531.

Rovine, M. J. (1994). Estimating nonshared environment using sibling discrepancy scores. In Hetherington, M., Reiss, D., and Plomin, R (Eds. *Separate social worlds of siblings: The impact of nonshared environment on development* (pp. 33-61), Hillsdale, NJ: Lawrence Erlbaum Associates.

Rushton, J. P. (1989). Genetic similarity, human altruism, and group selection. *Behavioral and Brain Sciences*, *1*2, 503-518.

Rushton, J. P., Walker, D. W., Neale, M. C., Nias, D. K. B., and Eysenck, H, J. (1986). Altruism and aggression: The heritability of individual differences. *Journal of Personality and Social Psychology*, *50*, 1192–1198.

Sampson, R. J., & Laub, J. H. (1995). *Crime in the making: Pathways and turning points through life*. Cambridge, MA: Harvard University Press.

Sampson, R. J., Laub, J. H., & Wimer, C. (2006). Does marriage reduce crime? A counterfactual approach to within-individual causal effects. *Criminology*, *44*, 465-508.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, *277*, 918-924.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., ... & Smith, M. (1977). Nucleotide sequence of bacteriophage φX174 DNA. *Nature*, *265*, 687-695.

Scarr, S., Scarf, E., and Weinberg, R. A. (1980). Perceived and actual similarities in biological and adoptive families: Does perceived similarity bias genetic estimates? *Behavior Genetics, 10*, 445–58.

Schluter, D. (1988). Estimating the form of natural selection on a quantitative trait. *Evolution*, *42*, 849-861.

Schwartz, J. A., and Beaver K. M. (2018). Intelligence and Offending: A Longitudinal Examination of the Differential Detection Hypothesis. *PsyArXiv preprint services*. Retrieved from https://psyarxiv.com/.

Seber, G. A., & Lee, A. J. (2012). *Linear regression analysis* (Vol. 936). Hoboken, NJ: John Wiley & Sons.

Selevan, S. G., Rice, D. C., Hogan, K. A., Euling, S. Y., Pfahles-Hutchens, A., & Bethel, J. (2003). Blood lead concentration and delayed puberty in girls. *New England Journal of Medicine*, *348*, 1527-1536.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth Cengage.

Shaw, C.R. & McKay, H.D. (1942). *Juvenile Delinquency and Urban Areas.* Chicago, IL: University of Chicago Press.

Silver, I. A., & Nedelec, J. L. (2018). The moderating effects of intelligence: An examination of how IQ influences the association between environmental factors and antisocial behavior. *Journal of Criminal Justice*, *54*, 62-75.

Simons, R. L., & Barr, A. B. (2014). Shifting perspectives: Cognitive changes mediate the impact of romantic relationships on desistance from crime. *Justice Quarterly*, *31*, 793-821.

Singleton, R., & Straits, B. C. (2010). *Approaches to Social Research* (*5th edition*). Oxford, UK: Oxford University Press.

Smith, D. A., & Paternoster, R. (1990). Formal processing and future delinquency: Deviance amplification as selection artifact. *Law and Society Review*, 1109-1131.

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge, MA: MIT press.

Strachan, T., & Read, A. (2011). *Human Molecular Genetics* (4th edition). Oxford, UK: Taylor and Francis Group, LLC.

Sulloway, F. J. (1982). Darwin and his finches: The evolution of a legend. *Journal of the History of Biology*, *15*, 1-53.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam, Netherlands: North-Holland Publishing Company.

Sutherland, EH, & Cressey, DR (1970) *Criminology*. Philadelphia, PA: JB Lippincott.

Thornberry, T. P., & Farnworth, M. (1982). Social correlates of criminal involvement: Further evidence on the relationship between social status and criminal behavior. *American Sociological Review*, *47,* 505-518.

Trivers, R. (1972). *Parental investment and sexual selection* (Vol. 136). Cambridge, MA: Biological Laboratories, Harvard University.

Trivers, R. L., & Willard, D. E. (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science*, *179*, 90-92.

Trotta, A., Iyegbe, C., Di Forti, M., Sham, P. C., Campbell, D. D., Cherny, S., ... & Fisher, H. (2017). 133. Interplay Between Schizophrenia Polygenic Risk Score and Childhood Adversity in First-Presentation Psychotic Disorder: A Pilot Study. *Schizophrenia Bulletin*, *43*, S72-S72.

Turkheimer, E., & Waldron, M. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin, 126*, 78-108.

Van Beneden, E. (1870). *Recherches sur la composition et la signification de l'œuf: basées sur l'étude de son mode de formation et des premiers phénomènes embryonnaires (mammifères, oiseaux, crustacés, vers)*. F. Hayez.

Vandenberg, S. G. (1966). Contributions of twin research to psychology. *Psychological Bulletin*, *66*, 327-352.

VanderWeele, T. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford, UK: Oxford University Press.

Viding, E., Fontaine, N. M., Oliver, B. R., & Plomin, R. (2009). Negative parental discipline, conduct problems and callous–unemotional traits: monozygotic twin differences study. *The British Journal of Psychiatry*, *195*, 414-419.

Vigen, T. (2015). *Spurious correlations*. New York, NY: Hachette Books.

Vitaro, F., Brendgen, M., & Arseneault, L. (2009). The discordant MZ-twin method: One step closer to the holy grail of causality. *International Journal of Behavioral Development*, *33*, 376-382.

Waldeyer, W. (1888). Karyokinesis and its Relation to the Process of Fertilization. *NEW SER*, *32*, 1-12.

Waller, R., Hyde, L. W., Klump, K. L., & Burt, S. A. (2018). Parenting Is an Environmental Predictor of Callous-Unemotional Traits and Aggression: A Monozygotic Twin Differences Study. *Journal of the American Academy of Child & Adolescent Psychiatry*, Online First, 1-9

Warr, M. (1998). Life-course transitions and desistance from crime. *Criminology*, *36*, 183-216.

Watson, J. D., Baker, T. A., Bell, S. P., & Gann, A. (2008). *Molecular biology of the gene* (7th edition). London, UK: Pearson & CSHL press

Watts, S. J. (2018). Adhd symptomatology and criminal behavior during adolescence: exploring the mediating role of school factors. *International Journal of Offender Therapy and Comparative Criminology*, *62*, 3-23.

Weismann, A. (1893). *The germ-plasm: a theory of heredity*. New York, NY: Charles Scribner's Sons.

Wilson, D. S. (1975). A theory of group selection. *Proceedings of the National Academy of Sciences*, *72*, 143-146.

Wright, J. P., & Beaver, K. M. (2005). Do parents matter in creating self-control in their children? A genetically informed test of Gottfredson and Hirschi's theory of low self-control. *Criminology*, *43*, 1169-1202.

Zacharias, L., & Wurtman, R. J. (1969). Age at menarche. *New England Journal of Medicine, 280*, 868-875.

APPENDICIES

Genetically adjusted propensity score matching: A proposal of a novel analytical tool to help close the gap between non-experimental designs and true experiments in the social sciences.

A dissertation submitted to the
Graduate School
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

In the School of Criminal Justice
of the College of Education, Criminal Justice, and Human Services

By

Ian A. Silver

Dissertation Committee: Joseph L Nedelec, Ph.D. (Chair)
J. C. Barnes, Ph.D.
Joshua C. Cochran, Ph.D.
George B. Richardson, Ph.D.

Appendix A: Quantitative Genetic Methodologies A Conceptual Background

In response to the substantive methodological and technological advancements in the latter half of the 20th century, various biologists have employed innovative statistical techniques to adjust for a variety of limitations associated with SSSMs (Lynch and Walsh, 1998). Though a delay was experienced, these innovative statistical techniques have been implemented in contemporary examinations of social behavior (e.g., Glazier, Nadeau, and Aitman, 2002; Lynch and Walsh, 1998; Okbay et al., 2016). In sociological scholarship, quantitative genetic methodologies highlight the important genetic predispositions commonly associated with a variety of TCs. Regarding self-selection, these quantitative genetic methodologies often rely on the complexity of evolution, genetic relatedness, and molecular genetics to statistically adjust for the unobserved and observed biological predispositions associated with a treatment condition (Lynch and Walsh, 1998; Plomin et al., 2013). As provided below, a comprehensive understanding of the principals of evolution, genetic relatedness, and molecular genetics simplifies the assumptions and mathematics associated with the various statistical techniques (presented in subsequent discussions).

A.1. Evolution: Beyond basic knowledge

Similar to Newton's (1687) theory of universal gravitation and Einstein's (1920) theory of relativity, Darwin's (1859) theory of evolution has had widespread impact on scholarship for multiple centuries (Sulloway, 1982; Mayr, 2009). As noted by historians (Mukherjee, 2017), Darwin's interest in biology dawned when he longed for an answer to his observation of systematic patterns within the animal kingdom, which was not satisfied by the simplistic answer provided by intelligent design. In an effort to generate a comprehensive understanding of the systematic patterns of organisms, Darwin (1859) generated a theoretical explanation indicating that the environment was primarily responsible for the emergence of divergent subspecies on

different continents. Even though contemporary scholarship often refers to Darwin's (1859) theoretical perspective under a single title, evolution, two theoretical hypotheses were introduced in *The Origin of Species.* These two theoretical hypotheses were that organisms inherit phenotypic mutations from previous generations (i.e., heredity) and that selective pressures influenced the selection of phenotypic mutations (i.e., natural selection).

A.1.1. Heredity

The first hypothesis proposed by Darwin (1859) was the concept that organisms inherit phenotypic mutations from previous generations, otherwise known as heredity. While addressed briefly, Darwin (1859) argued that a mechanism within an organism – single or multi-celled – was responsible for the intergenerational correspondence of phenotypic variation. To state simply, phenotypic mutations were inherited from prior generations, which in turn generated a distribution of phenotypes in the population. Darwin's (1859) scholarship generally circumvented discussions of heredity in an effort limit the theoretical assumptions associated with natural selection.

Although Darwin (1859) can be credited with the introduction of the heredity hypothesis, the foundation of heritability can be attributed to the work of Gregor Mendel. Mendel (1866) was made famous by the conclusions he drew from numerous observational and experimental assessments of pea plants.[61] The accumulation of evidence provided by Mendel's (1866) experiments led to his development of three theoretical assertions – laws – regarding the nature of heredity. His first hypothesis – the law of segregation – relied on observational studies in which two distinct pea plants – one with smooth and the other with wrinkled seeds – were mated. The offspring were then subjected to asexual reproduction, which yielded an interesting pattern of smooth and wrinkled seeds in the second generation of offspring. Mendel discovered that ¾ of

[61] Gregor Mendel's research did not become famous until after his death.

the seeds from the second generation were smooth, whereas ¼ of the seeds were wrinkled. From this Mendel (1866) hypothesized that every individual inherited two copies of the same trait, one of which was dominant (e.g., observed at a higher probability than the recessive copy) while the other was recessive (e.g., observed at a lower probability than the dominate copy).

The second hypothesis of heredity proposed by Mendel (1866) was the law of independent assortment. The law of independent assortment maintains that one trait cannot influence the probability of inheriting another trait. For example, in observation of Mendel's pea plants, the inheritance of smooth seeds and smooth leaves would occur independently of one another.[62] Mendel's third hypothesis on heredity was the law of dominance, which essentially argued that dominate phenotypes will always conceal the effects of recessive phenotypes. Regarding Mendel's pea plants, the breeding of one pea plant with smooth seeds and the other with wrinkled seeds will always result in the same distribution of phenotypes in subsequent generations (i.e., ¾ of the seeds from the second generation were smooth and ¼ of the seeds were wrinkled).

In response to Darwin's (1859) and Mendel's (1866) work, August Weismann (1893) published "*The germ-plasm: A Theory of Heredity",* which developed a theoretical explanation for the consistency between organisms.[63] Weismann (1893) theorized that cells –both single-celled and multi-celled organisms – were comprised of four hierarchical levels: Biophors, Determinants, Ids, and Idants. At the lowest level were biophors, which referred to the chemical molecules comprising the whole cell. Biophors determined the metabolic and growth rate of the cell. At the second lowest level were determinants, which highlighted by the name, referred to

[62] Notably, contemporary scholarship has illustrated that there are exceptions to Mendel's law of independent assortment (Reich et al., 2001). The primary exception occurs when genes are inherited together at a non-random probability, which is also termed linkage disequilibrium (Reich et al., 2001).
[63] Notably, due to the obscurity of Mendel's (1866) publication, Mendel had limited influence on Weismann's (1893) work. Nevertheless, while Weismann (1893) did not reference Mendel's scholarship, his work is a clear advancement upon Mendel's (1866) scholarship.

the substances that determined the phenotypic variation of a cell. This concept remained consistent when discussing multi-celled organisms. At the two highest levels – Ids and Idants – Weismann (1893) suggested that Ids were the aggregate of many determinants and Idants were the aggregate of many Ids. The conceptualization of Ids was partially consistent with genes and the conceptualization of Idants would be comparable to the contemporary concept of chromosomes (Morgan, 1915).

In addition to theorizing that the mechanism of heredity was contained in cells, Weismann (1893) offered important insight into sexual reproduction. Weismann (1893) theorized that sex cells only carry half of the Idants (i.e., chromosomes) of each parent and upon combination the number of Idants in a given zygote doubles. This hypothesis offered an explanation of how subsequent generations inherit traits from both parents. Furthermore, Weismann (1893) reinforced his claims by arguing that Ids from past generations are unequally allocated to future generations, resulting in an imbalanced representation of phenotypic variation from the "1st generation" to the "3rd generation." Although theoretically viable, Weismann's (1893) claims received limited empirical support until the realization that chromosomes were heritable (Van Beneden, 1870; Waldeyer, 1888; Morgan, 1915).

Further refinement of heredity came upon the heels of Mendelian inheritance during the 20th century. This movement – often referred to as the modern evolutionary synthesis – was brought about upon the publication of various scholarly pieces integrating the concepts put forth by Darwin (1859), Mendel (1866), and Weismann (1891; Churchill, 1980 Eldredge, 1985; Fisher, 1919; Mayr and Provine, 1998). For example, as outlined by Fisher (1919), Mendel's three laws of inheritance provided theoretical guidelines for empirically estimating the genetic variability and relatedness of lineages. In addition to operationalizing genetic variability, Fisher (1919) offered an empirical model illustrating how discrete genes – or mutations of the same

gene – could function in combination to account for phenotypic variation, later termed as additive genetic influence. Furthermore, these mathematical techniques provided direct support for synthetization of Darwin (1859), Mendel (1866), and Weismann's (1891) theoretical components regrading heredity. In addition to Fisher (1919), various 20th and 21st century scholars have provided insight into heredity (e.g., Mayr, 1991; Mayr and Provine, 1998) and will be discussed in subsection 3.2.2.

A.1.2. Natural Selection: Fitness and Survival

The second tenant of evolution put forth by Darwin (1859) was the concept of non-random selection (i.e., natural selection). Though the term *natural selection* might allude to selection through higher powers, Darwin's publications exemplified the belief that "natural selection" referred directly to environmental selection. To state differently, selection was not a function of intentions but rather resulted from variation in exposure to conditions that affected the survival and/or reproduction of an organism. As reasoned by Darwin (1859), natural selection was the primary mechanism influencing the variation at which phenotypic characteristics can be observed within a single species. For example, Darwin asserted that only natural selection could account of the variation associated with the beak size of Galapagos finches (i.e., Darwin's Finches).

Darwin (1859) further cemented his theory of natural selection by publishing his well-known book titled: *On the origin of species by means of natural selection*. Within this book Darwin detailed how exposure to divergent environmental factors potentially influenced the inheritance of phenotypic mutations in a species, which eventually produced in species phenotypic variation. As it should be noted, within the current context the term phenotypic variation refers to the distribution of phenotypes in a specified population. Phenotypic mutation refers to a phenotypic trait restricted to a small portion of the population. For example, the height

distribution of humans can be referred to as phenotypic variation, while one individual's height is a phenotypic mutation.

Definitions aside, Darwin's (1859) natural selection theory illustrated how the interaction between randomly emerging phenotypic mutations and environmental conditions can influence the phenotypic variation within a population. Due to irregularities in heritability – later recognized as random genetic mutations (e.g., Morgan, 1915; Morgan, 1917) – phenotypic mutations generally occur randomly throughout a population. As outlined by Darwin (1859) the existence of phenotypic mutations ensured the survival of a species by generating variation in the susceptibility to disease and mortality. Upon the introduction of phenotypic mutations into a species, two environmental mechanisms influence the perseverance of the trait: reproductive success and survivability. As outlined by Darwin (1859), reproductive success and survivability are the primary mechanisms of natural selection influencing phenotypic variation within a species. Consistent with contemporary scholarship (e.g., Dawkins, 1976; Ridley, 1994), Darwin (1859) outlined that reproductive success is the foremost factor guiding the natural selection of phenotypic mutations.

Although a variety of scholars have defined reproductive success (e.g., Agrawal, 2001; Crow, 1994; Ridley, 1994), Fisher (1915) provides a definitive understanding of the term. In his own words "the [reproductive] success of an animal in the struggle for existence is not measured only by the number of offspring which it produces and rears, but also by the probable [reproductive] success of these offspring" (Fisher, 1915, pg. 185). This can be further contextualized by a simple example; the reproductive success of a single Galapagos finch was conditional upon the ability of that finch to produce offspring and that finch's offspring to reproduce. In reference to natural selection, phenotypic mutations can drastically alter reproductive success. Alterations in reproductive success are generally captured in measures of

reproductive fitness (Dawkins, 1976; Lande and Arnold, 1983; Orr, 2007). Reproductive fitness – in contemporary scholarship – is often demarcated into two distinct measures: absolute reproductive fitness and relative reproductive fitness (Lande and Arnold, 1983; Orr, 2007; Orr, 2009; Wilson, 1975). Absolute reproductive fitness generally refers to the number of offspring in the subsequent generation possessing the specified phenotypic mutation (Carey, 2003). If the phenotypic mutation increases absolute reproductive fitness one would expect to observe an increased number of individuals in each subsequent generation possessing said phenotypic mutation (Carey, 2003; Lande and Arnold, 1983; Orr, 2007). Relative reproductive fitness is indicative of the number of offspring within the subsequent generation possessing the specified phenotypic mutation when compared to a reference phenotype (Lande and Arnold, 1983; Maree, et al., 2000; Orr, 2009; Wilson, 1975). For example, if the absolute reproductive fitness of fitches with long beaks was 100 (i.e., 100 offspring in the subsequent generation had long beaks) and the absolute reproductive fitness of fitches with wide beaks was 100 (i.e., 100 offspring in the subsequent generation had wide beaks), the relative reproductive fitness of the long beak finches would be .50 (i.e., 100/200). Furthermore, the phenotypic variation in the population would be .50, where half of the population would possess long beaks and half of the population would possess wide beaks. Notably, although the absolute reproductive fitness of a phenotypic mutation might increase, the relative reproductive fitness of a phenotypic mutation can increase, remain stable, or decrease depending upon the reference phenotype (Lande and Arnold, 1983; Wilson, 1975).

The reproductive fitness of a phenotypic mutation is dependent upon mating selection and environmental pressures (Alcock, 2009; Dingemanse et al., 2004; Emlen and Oring, 1977). For simplicity, the examples discussing mating selection will pertain primarily to mammalian species. Due to the substantive maternal investment within mammalian species, the female sex

generally dictates the reproductive fitness of phenotypic mutations in males (Geary, 2000; Qvarnstrom and Price, 2001; Trivers and Willard, 1973). Within humans, females generally select mates that possess phenotypic mutations indicative of increased paternal investment and increased security (Geary, 1998; Johnstone, 1994; Trivers, 1972). For example, evidence has indicated that females generally select mates with similar intelligence levels and financial security (Komer, 2015). Furthermore, environmental pressures noticeably effect the relative reproductive fitness of phenotypic mutations (Campbell et al., 2005; Gluckman and Hanson, 2006; Selevan, 2003; Zacharias and Wurtman, 1969). This was first discovered by Dobzhansky (1937) who identified that environmental temperature could vary the reproductive success of fruit flies. One genetic variation of fruit flies had higher reproductive success in warm climates and the other genetic variation of fruit flies had higher reproductive success in cold climates.

In combination with reproductive fitness, survivability is the second mechanism driving the natural selection of phenotypic mutations. Generally, survivability refers to the increased probability of reaching reproductive potential given the distinct phenotypic mutations inherited (Darwin, 1859). Note the distinction between the ability to reproduce and reproductive potential, where reproductive potential refers to the absolute capacity of offspring that can be generated if the environment allowed. Furthermore, as alluded to by the definition, natural selection as a function of survival depends primarily on reproduction (Darwin, 1859). If a phenotypic mutation results in increased fitness, the pressures of survivability will likely not influence the natural selection of the phenotypic mutation (Endler, 1986; Schulter, 1988; Fisher 1999). To state differently, when a mutation is selected for fitness purposes, it is not selected for survivability purposes (Darwin, 1859; Fisher 1999). While this principal is evident within a variety of species, the most common example is displayed in peacocks. The large plumage associated with the peacocks' tail has experienced selection pressures due to the increased fitness associated with the

phenotype, but directly harms the survivability of male peacocks in the wild. The order of selection factors – fitness before survival – remains valid unless the phenotypic mutation results in substantive increases in pre-pubescent mortality (Fisher 1999; Kingslover et al., 2001; Mayr, 1972). In humans, genetic diseases resulting in mortality preceding puberty rarely have high heritability estimates as a function of the reduced likelihood to replicate upon adulthood (Polderman et al., 2015). Furthermore, though within species phenotypic variation can attributed to selection by survivability, a variety of interspecies phenotypes (i.e., hunger, sleep and thirst) can be attributed to selection pressures associated with survival (Fisher 1999).

In summary, the two tenants put forth Darwin (1859) and reiterated by subsequent scholars have provided the foundational components for contemporary scholarship. Without the scholarship described in the current subsection, the fields of behavioral genetics and molecular genetics would likely not exist. As demonstrated by the subsequent subsections, scholars have employed the tenants put forth Darwin (1859) to conceptualize and operationalize heredity and natural selection at the both the phenotypic and the molecular levels. At the phenotypic level, scholars (e.g., Morgan, 1915; Plomin et al. 2013) have relied on Mendelian principals in an effort to theoretically and empirically establish the etiological structure of heredity and natural selection in contemporary species. At the molecular levels, Darwin (1859) and Mendel's (1868) principals provided a guiding light for identifying the mechanisms of heredity and natural selection. Overall, the historical establishment of evolution resulted in substantive advancements in human knowledge far beyond the predictions of Darwin (1859) and Mendel (1868).

A.2. Genetic relatedness: Two branches of the same tree

Due to the evolutionary history of multi-celled organisms, genetic overlap is common across approximately every species on earth. Within a species, substantive genetic overlap can be observed. Although most of the overlap corresponds to molecular consistencies (e.g., protein

production), the genetic correspondence dictates some phenotypic consistencies as well (e.g., thirst, hunger, and fatigue; Plomin et al. 2013). Indicative by the substantive genetic overlap within species, common ancestors with high reproductive fitness can account for the majority of the correspondence between individuals (Plomin et al. 2013). Whereas genetic overlap accounts for the similarities in a species, genetic variation accounts for the divergences between individuals of the same species. Whereas the amount of genetic variation in a species fluctuates, humans share approximately 99.5% of their genetic makeup. The other .5 percent – approximately 3,000 of the 20,000 protein coding genes – accounts for the phenotypic variation observed in humans (Mukherjee, 2017). With this knowledge and Mendel's (1866) laws of heritability, contemporary behavioral geneticists have generated a framework conceptualizing and operationalizing the genetic relatedness of two humans.

To preface our discussion of genetic relatedness, it is necessary to understand that cells – of the human species – are diploid, indicating that two copies of each chromosomes are housed within each cell (Carey, 2003). The chromosomes housed in each cell were inherited during reproduction, where the maternal and paternal reproductive cells donate 23 chromosomes respectively (Watson, Baker, Bell, and Gann, 2008). While human cells contain 46 chromosomes, reproductive cells employ meiosis to randomly select[64] the 23 chromosomes that will be transmitted to the subsequent generation (Watson et a., 2008). Upon interaction, the sperm cell and the egg cell recombined the genetic material, giving the offspring 46

[64] The above description was a simplistic rendering of a complex process. Specifically, during meiosis, sex cells employ a moderately random process to select individual genes – not chromosomes – to transmit to subsequent generations. The selection of individual genes alludes to the phenomena that genes – not chromosomes – are the mechanism of heredity (Watson et a., 2008; Dawkins, 1976). Notably, although the selection of individual genes implies that a random assortment of phenotypes from the grandparents are inherited by the offspring (e.g., Watson et a., 2008), recent scholarship has provided evidence suggesting that the selection of genes is not a completely random process. Linkage disequilibrium (LD) increases the probability of gene collections inherited by subsequent generations. LD increases this probability by non-randomly requiring independent genes to be inherited together (Reich et al., 2001). While LD is detrimental to some assumptions associated with molecular genetics, LD has little influence on establishing a conceptual framework for genetic relatedness (Reich et al., 2001).

chromosomes or two copies of each chromosomes (Watson et a., 2008). With this information, it can be assumed that genetic makeup of any offspring is inherited 50 percent from the mother and 50 percent from the father (Plomin et al., 2013)

Developed as a guide for kinship, various behavioral geneticists have adapted the table of consanguinity to illustrate the genetic relatedness between two individuals (Plomin et al., 2013). The table of consanguinity illustrates genetic relatedness from one individual within the family tree. In reference to the "person", the immediate family members (i.e., parents, full siblings, and children) are assumed to share approximately 50 percent of their genetic makeup.[65] Furthermore, it can be assumed a fraternal twin would share 50 percent of their genetic makeup – in reference to said person – and an identical twin would share 100 percent of their genetic makeup (Plomin et al., 2013). These assumptions are based on zygosity, where dizygotic twins (DZ; i.e., fraternal twins) have two egg cells and two sperm cells and monozygotic twins (MZ; i.e., identical twins) share one egg cell and one sperm cell (Carey, 2003).[66] Outside of the immediate family, it can be assumed that genetic relatedness substantially declines with prior generations and distance cousins. For example, each grandparent and great grandparent shares 25 and 12.5 percent of their genetic material with the reference person. In reference to the offspring of siblings, it can be assumed that a sibling's offspring shares 25 percent of its genetic material with the person and is halved each subsequent generation. This assumption is violated when discussing MZ twins (Plomin et al., 2013). Since it can be assumed that MZ twins share 100 percent of their genetic makeup, the offspring of one twin shares approximately 50 percent of its genetic material with

[65] As noted below, this assumption can be violated by assortative mating or the existence of an identical twin sibling.
[66] To provide an overview, MZ twins result from the meiosis of single zygote. Simply, meiosis is the uncommon division of a cell, where identical pairs of chromosomes are produced and split from one another to result in the development of two identical zygotes. Following the creation of the two identical zygotes, each one develops naturally into a pair of identical twins.

the other twin (Plomin et al., 2013). In the rare case that MZ twin pairs mate, their offspring would be considered genetic siblings (Plomin et al., 2013)

When adapted to behavioral genetics, the table of consanguinity provides a detailed framework outlining the expression of a single phenotype between family members. For instance, it can be assumed that the probability of immediate family members (i.e., individuals who share 50 percent of their genetic makeup) possessing a single phenotype would be .50 (Plomin et al., 2013). Again, this assumption is violated when discussing MZ twins, where the probability of expressing a phenotype would be 1.00 (Plomin et al., 2013). It can be speculated that deviations from .50 – where the expression between immediate family members on a single phenotype is higher or lower – would indicate higher or lower genetic similarity (Plomin et al., 2013). This simplistic explanation assumes that the phenotype was resulted strictly from genetic inheritance and was not influenced by environmental factors.

As speculated, two types of environmental factors can influence the probability of phenotypic expression between family members: the shared environment and the non shared environment. As alluded to by the designations, the shared environment refers to the environmental conditions that are mutual between the specified family members, while the non shared environment refers to the environmental conditions that are distinct between the family members. As noted by scholars (e.g., Barnes, Beaver, and Boutwell, 2011; Barnes et al., 2014), the shared environment generally only influences phenotypic expression for siblings, whereas the effects of the non shared environment can influence phenotypic expression for every family member. The effects of the shared environment are especially important during the establishment of phenotypic expression amongst MZ twins (Plomin et al., 2013). As argued by sceptics of behavioral genetics (e.g., Barnes et al., 2014), phenotypic expression between MZ twins can result from genetic equivalence or the overrepresented environmental equivalence (i.e., the

shared environment). As discussed further in sections 3.3.1. and 3.3.2., the conceptual framework outlined above provides the foundation for developing the mathematical assumptions associated with the ACE decomposition model and MZ difference models.

A.3. Genes: Mechanisms of heredity

During the early 20th century various scholars (e.g., Bridges, 1916; Dobzhansky, 1937; Morgan, 1915) debated about the true mechanisms of heredity. Specifically, what was inherited by humans and how did natural selection influence the intergenerational transmission of phenotypes. Various perspectives argued that the chromosome – similar to Weismann (1891) – and the individual could be the mechanism of heredity (Darwin, 1968). While these discussions waged Watson and Crick (1953) discovered deoxyribonucleic acid (DNA). This discovery eventually led to the identification of discrete units of inheritance – following Mendel's (1866) guidance – and the publication of two of theoretical appraisals of contemporary evolution: *Adaptation and Natural Selection* (1966)*, and The Selfish Gene* (1976). With similar intentions, both George Williams (1966) and Richard Dawkins (1976) reevaluated evolutionary theory in an effort to cement genes as the mechanisms of heredity.

Williams (1966) and Dawkins (1976) put forth theoretical hypotheses that were founded upon three concrete evolutionary principals. First, evident by Mendelian inheritance, to guarantee heredity a molecular commonalty between generations must exist. Notably, although chromosomes were theoretically viable, Dawkins (1976) argued that it was unlikely for an individual chromosome to be the mechanism of inheritance. Furthermore, evident by contemporary scholarship (e.g., Boehnke, 1991; Roach et al., 2010), the exact replication of specific genes can be observed in subsequent generations. At other molecular levels (e.g., chromosomes, epigenetics) exact replications cannot be observed in subsequent generations.

With this position, Williams (1966) and Dawkins (1976) argued that for Mendelian inheritance to be valid, genes were the only possible mechanism of inheritance.

Second, in line with the principals of evolution, the ability to replicate must exist and be inherited. From such position, Dawkins (1976) argued that genes can explain the inherited ability to replicate within a species. This argument can be contextualized around the hypothesis that only genes can be inherited by subsequent generations, creating the viable conclusion that genes cause the desire to reproduce. To state differently, it can be theorized that only replicated molecules can produce the desire to replicate within a species. Consistent with this hypothesis is the knowledge that chromosomes – or other phenotypic/molecular factors – are not directly replicated in subsequent generations (unless asexually; Carey, 2003; Dawson et al., 2002; Morely et al., 2004). Furthermore, guided by the same logic, Dawkins (1976) argues that genes can almost definitely ensure their presence in subsequent generations by increasing the desire to replicate.

Finally, Williams (1966) and Dawkins (1976) argued that only genes as the mechanism of heritability can explain the pervasiveness of altruism in familial lineages. If the organism itself was the replicator than altruism would only be reserved for direct descendants. Nevertheless, evidence indicates that most species are altruistic towards non-descendants, suggesting a molecular mechanism of inheritance (Rushton, 1989). Consistent with this evidence, Dawkins (1976) indicated that the display of altruism towards one's lineage general suggests that there is an expected reproductive value for one's own genes when regarding relatives. For example, Dawkins (1976) argues that the likelihood of subsequent generations receiving an individual's genes from a sibling is higher than a third cousin, which is consistent with the average amount of altruism displayed towards each of these family members. While theoretical when proposed, the position of Williams (1966) and Dawkins (1976) has been entrenched within behavioral and

molecular genetics and has guided a substantive number of empirical endeavors, including the human genome project.

Consistent with the hypotheses proposed by Williams (1966) and Dawkins (1976), the field of molecular genetics has demonstrated the importance of genes during the establishment of heredity, phenotypic mutations, and phenotypic variation. To detail, molecular genetics is the examination of the association between single and multiple genetic alleles – or mutations – and phenotypic variation. Genes code for proteins which perform one or multiple bodily functions (Mukherjee, 2017; Strachan and Read, 2011). An allele is the common genetic variant of a single gene and a mutation is all other genetic variants of a single gene (Strachan and Read, 2011).[67] Generally, genetic variants (i.e., alleles or mutations) occur as either single nucleotide polymorphism (SNP) or multi-nucleotide polymorphism (MNP). SNPs and MNPs are easily understood with a simple demonstration (Strachan and Read, 2011). DNA is made up of four nucleobases: adenine (A), cytosine (C), guanine (G), and thymine (T), where only A can bond to T and C can bond to G (Carey, 2003). If we take Gene *X*, which as an allele is coded as ATGGTTCAAG[68], a SNP would be ATGATTCAAG (the fourth nucleobase bond was changed from a G to an A). A MNP would be ATGAGTCAAT, where multiple nucleobase bonds were changed (Strachan and Read, 2011). When SNPs and MNPs do exist, they code for the same protein as the identified allele, but generate a different rate of production (Beaver, 2009). If the SNPs or MNPs results in the creation of a different protein it is labeled as a missense mutation (Strachan and Read, 2011).

While seemingly negligible SNPs and MNPs have enormous effects on the appearance of phenotypic mutations and the observed phenotypic variation in a population (Strachan and Read,

[67] Please note that polymorphisms exist within humans, where two or more genetic alleles are equally common within the population. At this point neither can be labeled as the genetic mutation (Carey, 2003).
[68] Consistent with molecular genetics, the matches are not provided because they can be simply deduced by knowing the initial (Watson et al., 2008).

2011). Furthermore, genes can influence the appearance of phenotypic mutations through dominant[69] and additive effects. One such case of a dominant genetic effect corresponds to the appearance of red hair, where mutations of a single gene influence the observed phenotypic variation (Puig-Butille et al., 2017). Specifically, when an individual inherits two recessive copies of the MC1R gene mutations the probability of inheriting red hair increases substantially (Puig-Butille et al., 2017). Divergent from dominant genetic effects, additive genetic effects correspond to the idea that multiple genes – or divergent mutations – influence the observed phenotypic variation. Recent scholarship has provided a detailed example of the additive genetic factors influencing educational attainment in the population (Rietveld et al., 2013). Furthermore, these results have been supported by subsequent results (Davies et al., 2016; Okbay et al., 2016). The concepts of dominant and additive genetic effects has established a vast array of empirical scholarship on the association between molecular genetics and the observed phenotypic variation.

A.3.1. Candidate Gene Research

Preceding the human genome project, the primary method of establishing an association between molecular genetic factors and phenotypic variation was through candidate gene research (Beaver, 2009). A candidate gene is a coding sequence of DNA that is hypothesized to influence the phenotype of interest (Beaver, 2009). The expectation is generally developed from prior theoretical or empirical scholarship. Commonly, scholars employ candidate genes to examine the interaction between genetic and environmental factors on the establishment of phenotypic variation (Beaver et al., 2007; Risch et al., 2009). One of the foremost examinations of candidate genes was conducted Caspi and colleagues (2003). Through reliance on prior scholarship, Caspi

[69] Please note that within the current context dominant genetic effect refers the hypothesis that one gene (and its alleles) causes the observed population variation within a single phenotype. This term is divergent from Mendel's (1866) dominant and recessive alleles, which refers to the effects individual alleles on the observed phenotype.

and colleagues (2003) examined the association between two polymorphisms – both alleles can be commonly observed in the population – of the 5-HTT gene (which is a well-known transcription site for the serotonin transporter) and subsequent depression. The two polymorphisms tested were the short and long alleles, where the short polymorphism was associated with a lower transcriptional efficiency – reduced production of the serotonin transporter – than the long polymorphism. The findings of the regression analyses indicated that the short polymorphism had limited direct effects on depression, but the interaction between short polymorphism and the number of stressful life events drastically increased the probability of depression compared to individuals with one or two long polymorphisms. Even though this scholarship has been valuable, the assumptions associated with candidate genes often result in the inability to replicate the findings (e.g., Risch et al., 2009). Precisely, the inability to replicate findings is often a function of the assumption that the candidate gene has a direct dominate effect on the phenotype of interest (Davies et al., 2016; Okbay et al., 2016).

Appendix B: Coding Schemes for Measures of Interest

Dependent Variables (*Add Health reference ID*)**:**

Delinquency (Wave IV; Barnes and Beaver, 2010; Nedelec, Park, and Silver, 2016; Piquero and Brezina, 2001)

1) In the past 12 months, how often did you go deliberately damage property that didn't belong to you? (H4DS1)
2) In the past 12 months, how often did you steal something worth less than $50? (H4DS6)
3) In the past 12 months, how often did you buy, sell, or hold stolen property? (H4DS8)
4) In the past 12 months, how often did you use someone else's credit card, bankcard, or automatic teller card without their permission or knowledge? (H4DS9)
5) In the past 12 months, how often did you deliberately write a bad check? (H4DS10)
6) In the past 12 In the past 12 months, how often did you steal something worth more than $50 (H4DS2)
7) In the past 12 months, how often did you go into a house or building to steal something (H4DS3)
8) In the past 12 months, how often did you use or threaten to use a weapon to get something from someone (H4DS4)
9) In the past 12 months, how often did you sell marijuana or other drugs (H4DS5)
10) In the past 12 months, how often did you take part in a physical fight where a group of your friends was against another group (H4DS7)
11) In the past 12 months, how often did you hurt someone badly enough in a physical fight that he or she needed care from a doctor or nurse (H4DS12)

(0 = never; 1 = 1 or 2 times; 2 = 3 or 4 times; 3 = 5 or more times)

Drug Use (*Standardized;* Wave IV; Barnes and Beaver, 2010; Nedelec, Park, and Silver, 2016; Piquero and Brezina, 2001)

1) During the past 30 days, on how many days did you smoke cigarettes? (H4TO5)
2) During the past 30 days, on how many days did you drink alcohol? (H4TO35)
3) During the past 12 months, on how many days did you use marijuana? (H4TO70)
4) During the past 12 months, on how many days did you use [favorite drug]? (H4TO98)

(0 = never, 1 = one or more times)

Treatment Conditions (*Add Health reference ID*)**:**

Intelligence (Wave III):

1) Intelligence (pvtstd3c)
(Measured continuously)

Educational Attainment (Wave III):

1) What is the highest grade or year of regular school you have completed? (H3ED1)
(0 = 12^{th} grade or less; 1 = 1 enrolled in at least 1 year of college or more)

Environmental Covariates (*Add Health reference ID;* Boutwell and Beaver, 2008; Beaver et al., 2009; Markowitz, and Salvatore, 2012; Barnes, 2012)**:**

Race (Wave I)
1) Interviewer: Please Code the race of the respondent from your own observation alone (H1GI9)
(1 = White; 2 = Black/African American; 3 = American Indian; 4 = Asian or Pacific Islander, 5 = Other)

Gender (Wave I)
1) Interviewer, please confirm that R's sex is accurate (BIO_SEX).
(1 = Male, 2 = Female)

Age (Wave I)
1) What is your Birth year (H1GI1Y)
2) Interviewer: Please Identify the year of the interview (IYEAR)

Parental Income (Wave I)

1) About how much total income, before taxes did your family receive in 1994? Include your own income, the income of everyone else in your household, and income from welfare benefits, dividends, and all other sources. (PA55)
(Measured Continuously)

Parental Employment Status (Wave I)

1) Do you work outside the home? (PA13)
(0 = no [skip to PA14]; 1 = yes [skip to PA17])
2) In the past 12 months, have you worked outside the home? (PA14)
(0 = no; 1 = yes [skip to PA15])
3) Were you employed full time at your last job? (PA15)
(0 = no; 1 = yes)
4) Are you employed full time? (PA17)
(0 = no; 1 = yes)

Parental Education (Wave I)

1) How far did you go in school? (PA12)
(0 = less than college; 1 = went to college, but did not graduate)

Maternal Conflict (standardized; Wave I; $\alpha = .85$)

1) How close do you feel to your mother? (H1WP9) *(reverse recoded)*
2) How much do you think she cares about you? (H1WP10) *(reverse recoded)*
(5 = Not at all; 4 = very little; 3 = somewhat; 3 = quite a bit; 1 = very much)
Reverse recoded

3) Most of the time, your mother is warm and loving toward you? (H1PF1)
4) Your mother encourages you to be independent? (H1PF2)

5) When you do something wrong that is important, your mother talks about it with you and helps you understand why it is wrong? (H1PF3)
6) You are satisfied with the way your mother and you communicate with each other? (H1PF4)
7) Overall, you are satisfied with your relationship with your mother? (H1PF5)
(1 = Strongly Agree; 2 = Agree; 3 = neither agree nor disagree 4= Disagree; 5 = Strongly Disagree)

Paternal Conflict (standardized; Wave I; $\alpha = .89$)

1) How close do you feel to your father? (H1WP13) *(reverse recoded)*
2) How much do you think he cares about you? (H1WP14) *(reverse recoded)*
(5 = Not at all; 4 = very little; 3 = somewhat; 3 = quite a bit; 1 = very much)
Reverse recoded

3) Most of the time, your father is warm and loving toward you (H1PF23)
4) You are satisfied with the way your father and you communicate with each other. (H1PF24)
5) Overall, you are satisfied with your relationship with your father. (H1PF25)
(1 = Strongly Agree; 2 = Agree; 3 = neither agree nor disagree 4= Disagree; 5 = Strongly Disagree)

School Attachment (Standardized; Wave 1; $\alpha = .78$)

1) How often did you have trouble, Getting along with teachers (H1ED15) (reverse recoded)
2) How often did you have trouble, paying attention in school (H1ED16) (reverse recoded)
3) How often did you have trouble, getting your homework done (H1ED17) (reverse recoded)
4) How often did you have trouble, getting along with other students (H1ED18) (reverse recoded)
(5 = never; 4 = just a few times; 3 = about once a week; 2= almost every day; 1 = every day)

5) Agree or Disagree, You feel close to people at your school (H1ED19)
6) Agree or Disagree, you feel like you are part of your school (H1ED20)
7) Agree or Disagree, you are happy to be at your school (H1ED22)
8) Agree or Disagree, the teachers at your school treat student fairly (H1ED23)
9) Agree or Disagree, you feel safe in your school (H1ED24)
(1 = Strongly agree; 2 = agree; 3 = neither agree nor disagree; 4 = disagree; 5 = strongly disagree)

Social Support (Wave I; $\alpha = .85$)

1) How much do you feel that adults care about you? (H1PR1)
2) How much do you feel that your teachers care about you? (H1PR2)
3) How much do you feel that your parents care about you ?(H1PR3)
4) How much do you feel that your friends care about you? (H1PR4)
5) How much do you feel that people in your family understand you? (H1PR5)
6) (Reverse coded) How much do you feel that you want to leave home (H1PR6)
7) How much do you feel that you and your family have fun together? (H1PR7)
8) How much do you feel that your family pays attention to you? (H1PR8)
(1 = Not at all; 2 = very little; 3 = somewhat; 4 = Quite a bit; 5 = Very much)

Delinquent Drug Use (Wave I; $\alpha = .76$)

1) Of your 3 best friends, how many smoke at least 1 cigarette a day? (H1TO9)
2) Of your 3 best friends, how many drink alcohol at least once a month? (H1TO29)
3) Of your 3 best friends, how many use marijuana at least once a month? (H1TO33)
(0 = no friends; 1 = 1 friend; 2 = 2 friends; 3 = 3 friends)

Baseline Delinquency (Wave 1)

1) In the past 12 months, how often did you paint graffiti or signs on someone else's property or in a public place? (h1ds1)
2) In the past 12 months, how often did you deliberately damage property that didn't belong to you? (h1ds2)
3) In the past 12 months, how often did you lie to your parents or guardians about where you had been or whom you were with? (h1ds3)
4) In the past 12 months, how often did you take something from a store without paying for it? (h1ds4)
5) In the past 12 months, how often did you run away from home? (h1ds7)
6) In the past 12 months, how often did you steal something worth less than $50? (h1ds13)
7) In the past 12 months, how often did you act loud, rowdy, or unruly in a public place? (h1ds15)
(0 = never; 1 = 1 or 2 times; 2 = 3 or 4 times; 3 = 5 or more times)

8) In the past 12 months, how often did you hurt someone badly enough to need bandages or care from a doctor or nurse? (h1ds6)
9) In the past 12 months, how often did you drive a car without the owner's permission? (h1ds8)
10) In the past 12 months, how often did you steal something worth more than $50? (h1ds9)
11) In the past 12 months, how often did you go into a house or building to steal something? (h1ds10)
12) In the past 12 months, how often did you threaten to use a weapon to get something from someone? (h1ds11)
13) In the past 12 months, how often did you sell marijuana or other drugs? (h1ds12)
14) In the past 12 months, how often did you take part in a fight where a group of your friends was against another group? (h1ds14)
(0 = never; 1 = 1 or 2 times; 2 = 3 or 4 times; 3 = 5 or more times)

Baseline Drug Use (Wave 1)

1) Have you ever tried cigarette smoking, even just 1 or 2 puffs? (h1to1)
2) Have you had a drink of beer, wine, or liquor – not just a sip or a taste of someone else's drink – more than 2 or 3 times in your life? (h1to12)
(0 = No, 1 = Yes)
3) How old were you when you tried marijuana for the first time? (h1to30) (0 = never; 1 = Tried marijuana once [ages 1 thru 18 coded as "1"])
4) How old were you when you tried any kind of cocaine, including powder, freebase, or crack cocaine for the first time? (h1to34) (0 = never; 1 = Tried marijuana once [ages 1 thru 18 coded as "1"])
5) How old were you when you tried inhalants, such as glue or solvents for the first time?

(h1to37) (0 = never; 1 = Tried marijuana once [ages 1 thru 18 coded as “1”])

6) How old were you when you tried any other type of illegal drug such as LSD, PCP, ecstasy, mushrooms, speed, ice, heroin, or pills without a doctor’s prescription? (h1to40) (0 = never; 1 = Tried marijuana once [ages 1 thru 18 coded as “1”])

Appendix C: Assessing the Mean Differences Between Twin Subsamples and Full Sample

Table C1: Descriptive statistics for the MZ twin sample and the full sample.

	MZ Twins				Full Sample				*t-value*	*ZΔ*
	N	$\bar{X}$	SD	Range	N	$\bar{X}$	SD	Range		
Dependent Variables (Wave IV)										
Delinquency	480	.04	.15	.00,1.39	15,158	.07	.25	.00,5.60	-4.66*	.162
Drug Use	477	-.18	1.19	-1.60,3.45	15,045	.01	1.25	-1.60,3.45	-3.40*	.154
Treatment Conditions (Wave III)										
Intelligence	458	96.88	16.33	7,122	14,194	98.53	17.11	7,122	-2.13*	.099
Educational Attainment	472	.54	.50	0,1	14,711	.54	.50	0,1	.05	.002
Covariates (Wave I)										
Age	570	16.18	1.59	13,19	20,158	16.15	1.74	12,21	.47	.019
Non-White	570	.36	.48	0,1	20,134	.38	.49	0,1	-1.15	.049
Male	570	.50	.50	0,1	20,173	.49	.50	0,1	.17	.007
Parent Income	425	49.37	63.05	0,800	14,926	45.64	51.25	0,999	1.21	.065
Parent Employment Status	485	.66	.48	0,1	17,124	.62	.49	0,1	1.76	.080
Parent Education	477	.53	.50	0,1	17,050	.42	.49	0,1	4.39*	.205
Maternal Conflict	518	-.08	1.12	-1.35,4.23	18,868	.01	1.28	-1.35,7.49	-1.62	.067
Paternal Conflict	403	-.18	1.19	-1.37,4.72	14,000	.01	1.38	-1.37,6.12	-3.00*	.141
School Attachment	560	.05	1.10	-3.89,2.06	19,719	-.01	1.08	-4.88,2.06	1.05	.046
Social Support	559	8.09	1.18	3.75,10.00	19,533	7.98	1.19	2,10	2.18*	.093
Peer Drug Use	554	.85	.90	0,3	19,474	.85	.89	0,3	.09	.004
Baseline Delinquency	566	.53	.73	0,6	19,844	.55	.69	0,6	-.71	.031
Baseline Drug Use	559	.49	.44	0,2	19,662	.53	.47	0,2	-2.11*	.088

Notes: In the current context, the full sample includes same sex and different sex DZ twins.
*$p < .05$

Table C2: Descriptive statistics for the same sex MZ/DZ twin sample and the full sample.

	Same Sex MZ/DZ Twins				Full Sample				*t-value*	*ZΔ*
	N	$\bar{X}$	SD	Range	N	$\bar{X}$	SD	Range		
Dependent Variables (Wave IV)										
Delinquency	888	.04	.15	.00,2.80	14,750	.07	.25	.00,5.60	-2.05*	.161
Drug Use	876	-.18	1.19	-1.60,3.45	14,646	.01	1.25	-1.60,3.45	-.84	.151
Treatment Conditions (Wave III)										
Intelligence	839	96.88	16.33	7,122	13,813	98.56	17.16	7,122	-2.56*	.101
Educational Attainment	871	.54	.50	0,1	14,312	.54	.50	0,1	-.74	.001
Covariates (Wave I)										
Age	1,060	16.18	1.59	12,20	19,668	16.16	1.74	12,21	1.54*	.016
Non-White	1,060	.36	.48	0,1	19,644	.39	.49	0,1	-.87	.049
Male	1,060	.50	.50	0,1	19,683	.49	.50	0,1	1.36	.009
Parent Income	803	49.37	63.05	0,800	14,548	45.67	51.69	0,999	.60	.064
Parent Employment Status	912	.66	.48	0,1	16,697	.62	.49	0,1	2.52*	.082
Parent Education	904	.53	.50	0,1	16,623	.42	.49	0,1	4.16*	.207
Maternal Conflict	980	-.08	1.12	-1.35,6.36	18,406	.01	1.28	-1.35,7.49	-1.99*	.069
Paternal Conflict	738	-.18	1.94	-1.37,6.12	13,665	.01	1.39	-1.37,6.12	-2.35*	.141
School Attachment	1,040	.05	1.10	-3.89,2.06	19,239	-.01	1.08	-4.88,2.06	.50	.045
Social Support	1,038	8.09	1.18	3,10	19,054	7.98	1.20	2,10	2.45*	.094
Peer Drug Use	1,023	.85	.90	0,3	19,005	.85	.89	0,3	-.13	.004
Baseline Delinquency	1,047	.53	.73	0,6	19,363	.55	.70	0,6	-.98	.032
Baseline Drug Use	1,038	.49	.44	0,2	19,183	.53	.47	0,2	-1.72	.088

Notes: In the current context, the full sample includes different sex DZ twins.
*$p < .05$

Table C3: Descriptive statistics for the different sex MZ/DZ twin sample and full sample.

	Different Sex MZ/DZ Twins				Full Sample				*t-value*	*ZΔ*
	N	$\bar{X}$	SD	Range	N	$\bar{X}$	SD	Range		
Dependent Variables (Wave IV)										
Delinquency	1,229	.04	.15	.00,2.79	14,409	.07	.25	.00,5.60	-2.67*	.070
Drug Use	1,216	-.18	1.19	-1.60,3.45	14,306	.01	1.25	-1.60,3.45	-.60	.058
Treatment Conditions (Wave III)										
Intelligence	1,144	96.88	16.33	7,122	13,508	98.60	17.19	7,122	-3.09*	.103
Educational Attainment	1,184	.54	.50	0,1	13,999	.54	.50	0,1	-.45	.001
Covariates (Wave I)										
Age	1,461	16.18	1.59	12,20	19,267	16.16	1.75	12,21	-2.01*	.015
Non-White	1,460	.36	.48	0,1	19,244	.38	.49	0,1	-.01	.047
Male	1,461	.50	.50	0,1	19,282	.49	.50	0,1	1.31	.009
Parent Income	1,117	49.37	63.05	0,800	14,234	45.71	51.88	0,999	.13	.063
Parent Employment Status	1,262	.66	.48	0,1	16,347	.62	.49	0,1	1.51	.081
Parent Education	1,258	.53	.50	0,1	16,269	.42	.49	0,1	4.66*	.209
Maternal Conflict	1,361	-.08	1.12	-1.35,6.36	18,025	.01	1.28	-1.35,7.49	-2.94*	.071
Paternal Conflict	1,007	-.18	1.19	-1.37,6.12	13,396	.01	1.38	-1.37,6.12	-2.07*	.142
School Attachment	1,431	.05	1.10	-4.30,2.06	18,848	-.01	1.08	-4.88,2.06	.43	.045
Social Support	1,426	8.09	1.18	3,10	18,666	7.97	1.20	2,10	2.78*	.096
Peer Drug Use	1,409	.85	.90	0,3	18,619	.85	.89	0,3	-1.12	.002
Baseline Delinquency	1,443	.53	.73	0,6	18,967	.55	.70	0,6	-2.24*	.034
Baseline Drug Use	1,431	.49	.44	0,2	18,790	.53	.47	0,2	-3.10*	.091

Notes: In the current context, the full sample includes anyone who was not identified as a twin.
*$p < .05$

Appendix D: 40 Specifications of the Treatment Condition.

First 13 Specifications:

1. $t = .05*a + .45*e + .45*c + .04*E$
2. $t = .10*a + .43*e + .43*c + .04*E$
3. $t = .15*a + .40*e + .40*c + .04*E$
4. $t = .20*a + .38*e + .38*c + .04*E$
5. $t = .25*a + .35*e + .35*c + .04*E$
6. $t = .30*a + .33*e + .33*c + .04*E$
7. $t = .35*a + .30*e + .30*c + .04*E$
8. $t = .40*a + .28*e + .28*c + .04*E$
9. $t = .45*a + .25*e + .25*c + .04*E$
10. $t = .50*a + .23*e + .23*c + .04*E$
11. $t = .55*a + .20*e + .20*c + .04*E$
12. $t = .60*a + .18*e + .18*c + .04*E$
13. $t = .65*a + .15*e + .15*c + .04*E$

Second 13 Specifications:

14. $t = .05*a + .68*e + .22*c + .04*E$
15. $t = .10*a + .64*e + .21*c + .04*E$
16. $t = .15*a + .60*e + .20*c + .04*E$
17. $t = .20*a + .56*e + .19*c + .04*E$
18. $t = .25*a + .53*e + .18*c + .04*E$
19. $t = .30*a + .49*e + .16*c + .04*E$
20. $t = .35*a + .45*e + .15*c + .04*E$
21. $t = .40*a + .41*e + .14*c + .04*E$
22. $t = .45*a + .38*e + .13*c + .04*E$
23. $t = .50*a + .34*e + .11*c + .04*E$
24. $t = .55*a + .30*e + .10*c + .04*E$
25. $t = .60*a + .26*e + .09*c + .04*E$
26. $t = .65*a + .23*e + .08*c + .04*E$

Third 13 Specifications:

27. $t = .05*a + .22*e + .68*c + .04*E$
28. $t = .10*a + .21*e + .64*c + .04*E$
29. $t = .15*a + .20*e + .60*c + .04*E$
30. $t = .20*a + .19*e + .56*c + .04*E$
31. $t = .25*a + .18*e + .53*c + .04*E$
32. $t = .30*a + .16*e + .49*c + .04*E$
33. $t = .35*a + .15*e + .45*c + .04*E$
34. $t = .40*a + .14*e + .41*c + .04*E$
35. $t = .45*a + .13*e + .38*c + .04*E$
36. $t = .50*a + .11*e + .34*c + .04*E$
37. $t = .55*a + .10*e + .30*c + .04*E$
38. $t = .60*a + .09*e + .26*c + .04*E$
39. $t = .65*a + .08*e + .23*c + .04*E$

Point of Equivalence

40. $t = .32*a + .32*e + .32*c + .04*E$

Appendix E: Supplemental Results for Study 2

Table E1: Balance statistics for participants matched with nearest neighbor matching (caliper = .05) on educational attainment (Wave III).

DV: Educational Attainment (Wave III)	Did Not Attend One Year of College (c)	Attended One Year of College (t)	% Bias	*t-value*
	$\bar{X}$	$\bar{X}$		
Covariates (Wave I)				
Age	15.920	15.886	-1.981	-.636
Non-White	.295	.302	1.495	.489
Male	.501	.516	3.037	1.014
Parent Income	44.818	46.864	3.280	1.941
Parent Employment Status	.604	.604	-.184	-.061
Parent Education	.378	.366	-2.433	-.833
Maternal Conflict	-.047	-.078	-2.653	-.869
Paternal Conflict	-.031	-.035	-.352	-.114
School Attachment	.059	.043	-1.603	-.514
Social Support	8.075	8.082	.658	.211
Peer Drug Use	.797	.809	1.421	.444
Baseline Delinquency	.525	.535	1.659	.504
Baseline Drug Use	.504	.512	1.516	.494
N	2,250	2,250		

Notes: Caliper for the nearest neighbor matching was set at $p < .05$. (c) indicates the control cases and (t) indicates the treatment cases.

*$p < .05$

Table E2: Balance statistics for participants matched with nearest neighbor matching (caliper = .01) on educational attainment (Wave III).

DV: Educational Attainment (Wave III)	Did Not Attend One Year of College (c)	Attended One Year of College (t)	% Bias	*t-value*
	$\bar{X}$	$\bar{X}$		
Covariates (Wave I)				
Age	15.909	15.911	.163	.052
Non-White	.292	.295	.707	.231
Male	.511	.513	.271	.090
Parent Income	45.047	47.240	3.517	1.970*
Parent Employment Status	.611	.611	.093	.037
Parent Education	.380	.364	-3.378	-1.148
Maternal Conflict	-.079	-.042	3.095	1.007
Paternal Conflict	-.047	-.024	1.771	.572
School Attachment	.071	.038	-3.363	-1.084
Social Support	8.085	8.066	-1.831	-.586
Peer Drug Use	.789	.810	2.633	.819
Baseline Delinquency	.515	.542	4.483	1.371
Baseline Drug Use	.502	.518	3.686	1.206
N	2,221	2,221		

Notes: Caliper for the nearest neighbor matching was set at $p < .01$. (c) indicates the control cases and (t) indicates the treatment cases.

* $p < .05$

Table E3: Balance statistics for participants matched with nearest neighbor matching (caliper = .001) on educational attainment (Wave III).

DV: Educational Attainment (Wave III)	Did Not Attend One Year of College (c)	Attended One Year of College (t)	% Bias	*t-value*
	$\bar{X}$	$\bar{X}$		
Covariates (Wave I)				
Age	15.970	15.894	-4.568	-1.289
Non-White	.287	.288	.256	.075
Male	.504	.505	.229	.068
Parent Income	46.067	47.760	2.714	1.437
Parent Employment Status	.607	.608	.353	.104
Parent Education	.412	.404	-1.618	-.481
Maternal Conflict	-.079	-.085	-.482	-.140
Paternal Conflict	-.050	-.039	.889	.254
School Attachment	.094	.068	-2.651	-.758
Social Support	8.103	8.071	-2.975	-.842
Peer Drug Use	.782	.782	.024	.007
Baseline Delinquency	.500	.529	4.912	1.369
Baseline Drug Use	.492	.509	3.803	1.123
N	1,754	1,754		

Notes: Caliper for the nearest neighbor matching was set at $p < .001$. (c) indicates the control cases and (t) indicates the treatment cases.

*$p < .05$

Table E4: Balance statistics for participants matched with nearest neighbor matching (caliper = .0001) on educational attainment (Wave III).

DV: Educational Attainment (Wave III)	Did Not Attend One Year of College (c)	Attended One Year of College (t)	% Bias	*t-value*
	$\bar{X}$	$\bar{X}$		
Covariates (Wave I)				
Age	16.076	15.986	-5.390	-.786
Non-White	.306	.282	-5.352	-.830
Male	.539	.535	-.799	-.126
Parent Income	48.302	49.302	1.603	.331
Parent Employment Status	.606	.573	-6.982	-1.089
Parent Education	.439	.425	-2.822	-.445
Maternal Conflict	.016	-.139	-13.102	-2.002*
Paternal Conflict	-.065	-.081	-1.248	-.196
School Attachment	.107	.101	-.624	-.097
Social Support	8.058	8.175	11.123	1.700
Peer Drug Use	.783	.759	-2.928	-.442
Baseline Delinquency	.521	.510	-1.951	-.293
Baseline Drug Use	.490	.487	-.753	-.119
N	503	503		

Notes: Caliper for the nearest neighbor matching was set at $p < .0001$. (c) indicates the control cases and (t) indicates the treatment cases.

*$p < .05$

Table E5: Descriptive statistics for the matched sample (caliper = .05).

	Post-matching Sample	Pre-matching Sample	*t-value*
	$\bar{X}$	$\bar{X}$	
Dependent Variables (Wave IV)			
Delinquency	.071	.065	1.330
Drug Use	.051	.029	.850
Treatment Conditions (Wave III)			
Educational Attainment	.500	.620	-12.385*
Covariates (Wave I)			
Age	15.903	15.918	-.439
Non-White	.298	.266	3.631*
Male	.509	.467	4.255*
Parent Income	45.841	55.084	-10.507*
Parent Employment Status	.604	.611	-.708
Parent Education	.372	.480	-11.290*
Maternal Conflict	-.063	-.069	.256
Paternal Conflict	-.033	-.056	.865
School Attachment	.051	.093	-2.060*
Social Support	8.079	8.089	-.462
Peer Drug Use	.803	.759	2.600*
Baseline Delinquency	.530	.521	.659
Baseline Drug Use	.508	.490	2.046
N	4,500	6,202	

Notes: Caliper for the nearest neighbor matching was set at $p < .05$. Unmatched sample designates the cases remaining after listwise deletion for the model. The sample size for delinquency (Wave IV) was 3,896 and the sample size for Drug use (Wave IV) was 3,885 on the matched sample.
* $p < .05$

Table E6: Descriptive statistics for the matched sample (caliper = .01).

	Post-matching Sample	Pre-matching Sample	*t-value*
	$\bar{X}$	$\bar{X}$	
Dependent Variables (Wave IV)			
Delinquency	.070	.065	.929
Drug Use	.052	.029	.897
Treatment Conditions (Wave III)			
Educational Attainment	.500	.620	-12.337*
Covariates (Wave I)			
Age	15.910	15.918	-.236
Non-White	.294	.266	3.100*
Male	.512	.467	4.572*
Parent Income	46.144	55.084	-9.958*
Parent Employment Status	.611	.611	.046
Parent Education	.372	.480	-11.211*
Maternal Conflict	-.060	-.069	.361
Paternal Conflict	-.035	-.056	.772
School Attachment	.054	.093	-1.923
Social Support	8.075	8.089	-.616
Peer Drug Use	.800	.759	2.405*
Baseline Delinquency	.529	.521	.598
Baseline Drug Use	.510	.490	2.217*
N	4,442	6,202	

Notes: Caliper for the nearest neighbor matching was set at $p < .01$. Unmatched sample designates the cases remaining after listwise deletion for the model. The sample size for delinquency (Wave IV) was 3,840 and the sample size for Drug use (Wave IV) was 3,826 on the matched sample.
* $p < .05$

Table E7: Descriptive statistics for the matched sample (caliper = .001).

	Post-matching Sample $\bar{X}$	Pre-matching Sample $\bar{X}$	*t-value*
Dependent Variables (Wave IV)			
Delinquency	.066	.065	.091
Drug Use	.034	.029	.176
Treatment Conditions (Wave III)			
Educational Attainment	.500	.620	-11.460*
Covariates (Wave I)			
Age	15.932	15.918	.382
Non-White	.287	.266	2.212*
Male	.505	.467	3.533*
Parent Income	46.914	55.084	-8.900*
Parent Employment Status	.607	.611	-.320
Parent Education	.408	.480	-6.944*
Maternal Conflict	-.082	-.069	-.513
Paternal Conflict	-.045	-.056	.393
School Attachment	.081	.093	-.529
Social Support	8.087	8.089	-.071
Peer Drug Use	.782	.759	1.266
Baseline Delinquency	.515	.521	-.504
Baseline Drug Use	.501	.490	1.136
N	3,508	6,202	

Notes: Caliper for the nearest neighbor matching was set at $p < .001$. Unmatched sample designates the cases remaining after listwise deletion for the model. The sample size for delinquency (Wave IV) was 3,034 and the sample size for Drug use (Wave IV) was 3,026 on the matched sample.
* $p < .05$

Table E8: Descriptive statistics for the matched sample (caliper = .0001).

	Post-matching Sample	Pre-matching Sample	*t-value*
	$\bar{X}$	$\bar{X}$	
Dependent Variables (Wave IV)			
Delinquency	.073	.065	.880
Drug Use	.007	.029	-.480
Treatment Conditions (Wave III)			
Educational Attainment	.500	.620	-7.075*
Covariates (Wave I)			
Age	16.031	15.918	1.854
Non-White	.294	.266	1.806
Male	.537	.467	4.100*
Parent Income	48.802	55.084	-3.766*
Parent Employment Status	.589	.611	-1.275
Parent Education	.432	.480	-2.832*
Maternal Conflict	-.061	-.069	.178
Paternal Conflict	-.073	-.056	-.377
School Attachment	.104	.093	.344
Social Support	8.117	8.089	.746
Peer Drug Use	.771	.759	.393
Baseline Delinquency	.515	.521	-.285
Baseline Drug Use	.489	.490	-.088
N	1,006	6,202	

Notes: Caliper for the nearest neighbor matching was set at $p < .0001$. Unmatched sample designates the cases remaining after listwise deletion for the model. The sample size for delinquency (Wave IV) was 873 and the sample size for Drug use (Wave IV) was 871 on the matched sample.
* $p < .05$

Table E9: Predicting delinquency (Wave IV) and drug use (Wave IV) with educational attainment (Wave III) after nearest neighbor matching with a caliper at .05.

	DV: Delinquency (Wave IV)				DV: Drug Use (Wave IV)			
	b	SE	β	95%CI	b	SE	β	95%CI
Treatment Conditions (Wave III)								
Educational Attainment	-.009	.007	-.020	-.024,.005	-.094*	.040	-.038	-.173,-.016
R^2	.001				.001*			
N	3,896				3,885			

Notes: Caliper for the nearest neighbor matching was set at $p \leq .05$.
*$p < .05$

Table E10: Predicting delinquency (Wave IV) and drug use (Wave IV) with educational attainment (Wave III) after nearest neighbor matching with a caliper at .01.

	DV: Delinquency (Wave IV)				DV: Drug Use (Wave IV)			
	b	SE	β	95%CI	b	SE	β	95%CI
Treatment Conditions (Wave III)								
Educational Attainment	-.005	.008	-.011	-.020,.010	-.074	.041	-.029	-.153,.006
R^2	-.001				.001			
N	3,840				3,826			

Notes: Caliper for the nearest neighbor matching was set at $p \leq .01$.
*$p < .05$

Table E11: Predicting delinquency (Wave IV) and drug use (Wave IV) with educational attainment (Wave III) after nearest neighbor matching with a caliper at .001.

	DV: Delinquency (Wave IV)				DV: Drug Use (Wave IV)			
	b	SE	β	95%CI	b	SE	β	95%CI
Treatment Conditions (Wave III)								
Educational Attainment	-.005	.008	-.012	-.021,.011	-.054	.046	-.022	-.144,.035
R^2	-.001				.001			
N	3,034				3,026			

Notes: Caliper for the nearest neighbor matching was set at $p \leq .001$.
*$p < .05$

Table E12: Predicting delinquency (Wave IV) and drug use (Wave IV) with educational attainment (Wave III) after nearest neighbor matching with a caliper at .0001.

	DV: Delinquency (Wave IV)				DV: Drug Use (Wave IV)			
	b	SE	β	95%CI	b	SE	β	95%CI
Treatment Conditions (Wave III)								
Educational Attainment	-.026	.017	-.051	-.060,.008	-.048	.087	-.019	-.218,.122
R^2	.001				-.001			
N	873				871			

Notes: Caliper for the nearest neighbor matching was set at $p \leq .0001$.
*$p < .05$

Appendix F: Complete Results of the Simulation Analysis

Table F1. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to the bivariate and true point estimates (all measures observed; variance predicted by c and e equal).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Bivariate (y ~ t)	Post Matching (e)	Post Matching (c)	Post Matching (a)	Post Matching (e + c)	Post Matching (e + a)	Post Matching (c + a)	Post Matching (e + c + a)
	b	b	b	b	b	b	b	b	b
a(.05), c(.45), e(.45)	1.00	1.36	1.27	1.29	1.35	1.15	1.21	1.23	1.01
a(.10), c(.43), e(.43)	1.00	1.38	1.29	1.31	1.35	1.21	1.21	1.24	1.01
a(.15), c(.40), e(.40)	1.00	1.39	1.31	1.33	1.35	1.23	1.23	1.25	1.01
a(.20), c(.38), e(.38)	1.00	1.40	1.32	1.34	1.35	1.24	1.23	1.26	1.01
a(.25), c(.35), e(.35)	1.00	1.41	1.33	1.35	1.35	1.24	1.24	1.27	1.01
a(.30), c(.33), e(.33)	1.00	1.41	1.34	1.35	1.35	1.24	1.25	1.27	1.01
a(.32), c(.32), e(.32)	1.00	1.41	1.34	1.35	1.36	1.24	1.25	1.27	1.01
a(.35), c(.30), e(.30)	1.00	1.41	1.34	1.35	1.36	1.24	1.25	1.27	1.01
a(.40), c(.28), e(.28)	1.00	1.41	1.33	1.35	1.36	1.24	1.23	1.26	1.01
a(.45), c(.25), e(.25)	1.00	1.40	1.32	1.34	1.36	1.24	1.21	1.24	1.01
a(.50), c(.23), e(.23)	1.00	1.38	1.31	1.33	1.36	1.24	1.18	1.22	1.01
a(.55), c(.20), e(.20)	1.00	1.37	1.30	1.32	1.36	1.24	1.15	1.19	1.01
a(.60), c(.18), e(.18)	1.00	1.35	1.29	1.3	1.36	1.24	1.13	1.15	1.01
a(.65), c(.15), e(.15)	1.00	1.33	1.28	1.29	1.36	1.24	1.10	1.12	1.01

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such $t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*\varepsilon)$; The formula predicting the dependent variable (y) was specified as such $y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*\varepsilon)$. Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. All of the 95%CI ranged between -.01 and +.01 of the point estimate. Starting N = 50,000 cases. N size varies minimally between each point estimate.

Table F2. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to true point estimates (varying observed measures e; variance predicted by c and e equal).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Post Matching (e(1) + a)	Post Matching (e(1) + c(4) + a)	Post Matching (e(2) + a)	Post Matching (e(2) + c(4) + a)	Post Matching (e(3) + a)	Post Matching (e(3) + c(4) + a)
	b	b	b	b	b	b	b
a(.05), c(.45), e(.45)	1.00	1.33	1.22✢	1.31	1.18✢∓	1.29	1.13✢∓≒
a(.10), c(.43), e(.43)	1.00	1.33	1.22✢	1.31	1.18✢∓≒	1.29	1.13✢∓≒
a(.15), c(.40), e(.40)	1.00	1.33	1.22✢∓≒	1.31	1.18✢∓≒	1.29	1.13✢∓≒
a(.20), c(.38), e(.38)	1.00	1.34	1.22✢∓≒	1.32	1.18✢∓≒	1.29	1.13✢∓≒
a(.25), c(.35), e(.35)	1.00	1.34	1.22✢∓≒	1.32	1.18✢∓≒	1.29	1.13✢∓≒
a(.30), c(.33), e(.33)	1.00	1.34	1.22✢∓≒	1.32	1.18✢∓≒	1.29	1.13✢∓≒
a(.32), c(.32), e(.32)	1.00	1.34	1.22✢∓≒	1.32	1.18✢∓≒	1.29	1.14✢∓≒
a(.35), c(.30), e(.30)	1.00	1.34	1.22✢∓≒	1.32	1.18✢∓≒	1.29	1.13✢∓≒
a(.40), c(.28), e(.28)	1.00	1.34	1.22✢∓≒	1.32	1.18✢∓≒	1.30	1.13✢∓≒
a(.45), c(.25), e(.25)	1.00	1.34	1.22✢≒	1.32	1.18✢∓≒	1.30	1.13✢∓≒
a(.50), c(.23), e(.23)	1.00	1.34	1.22✢≒	1.32	1.18✢∓≒	1.30	1.12✢∓≒
a(.55), c(.20), e(.20)	1.00	1.35	1.22≒	1.32	1.18✢≒	1.29	1.13✢∓≒
a(.60), c(.18), e(.18)	1.00	1.35	1.22≒	1.32	1.18≒	1.30	1.12✢∓≒
a(.65), c(.15), e(.15)	1.00	1.35	1.22≒	1.32	1.18≒	1.29	1.12✢≒

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*ε); The formula predicting the dependent variable (y) was specified as such y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*ε). Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. The number in the bracket represents the number of observed variables included in the PSM model. All of the 95%CI ranged between -.01 and +.01 of the point estimate. Starting N = 50,000 cases. N size varies minimally between each point estimate.

✢ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (c + a; Table F1).

∓ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + a; Table F1).

≒ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + c; Table F1).

Table F3. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to true point estimates (varying observed measures c; variance predicted by c and e equal).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Post Matching (c(1) + a)	Post Matching (e(4) + c(1) + a)	Post Matching (c(2) + a)	Post Matching (e(4) + c(2) + a)	Post Matching (c(3) + a)	Post Matching (e(4) + c(3) + a)
	b	b	b	b	b	b	b
a(.05), c(.45), e(.45)	1.00	1.32	1.22✢	1.30	1.18✢∓≒	1.28	1.13✢∓≒
a(.10), c(.43), e(.43)	1.00	1.32	1.22✢	1.30	1.19✢∓≒	1.28	1.14✢∓≒
a(.15), c(.40), e(.40)	1.00	1.33	1.22✢∓≒	1.31	1.18✢∓≒	1.28	1.13✢∓≒
a(.20), c(.38), e(.38)	1.00	1.32	1.22✢∓≒	1.31	1.18✢∓≒	1.28	1.13✢∓≒
a(.25), c(.35), e(.35)	1.00	1.33	1.22✢∓≒	1.31	1.18✢∓≒	1.28	1.13✢∓≒
a(.30), c(.33), e(.33)	1.00	1.33	1.22✢∓≒	1.31	1.18✢∓≒	1.28	1.13✢∓≒
a(.32), c(.32), e(.32)	1.00	1.33	1.22✢∓≒	1.31	1.19✢∓≒	1.29	1.13✢∓≒
a(.35), c(.30), e(.30)	1.00	1.33	1.22✢∓≒	1.31	1.18✢∓≒	1.29	1.13✢∓≒
a(.40), c(.28), e(.28)	1.00	1.33	1.22✢∓≒	1.31	1.18✢∓≒	1.29	1.13✢∓≒
a(.45), c(.25), e(.25)	1.00	1.34	1.22✢≒	1.31	1.18✢∓≒	1.28	1.13✢∓≒
a(.50), c(.23), e(.23)	1.00	1.34	1.22✢≒	1.31	1.18✢≒	1.28	1.13✢∓≒
a(.55), c(.20), e(.20)	1.00	1.34	1.22≒	1.31	1.18✢≒	1.28	1.12✢∓≒
a(.60), c(.18), e(.18)	1.00	1.33	1.22≒	1.31	1.17≒	1.28	1.12✢∓≒
a(.65), c(.15), e(.15)	1.00	1.33	1.21≒	1.30	1.17≒	1.28	1.12✢≒

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*ε); The formula predicting the dependent variable (y) was specified as such y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*ε). Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. The number in the bracket represents the number of observed variables included in the PSM model. Starting N = 50,000 cases. N size varies minimally between each point estimate.

✢ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (c + a; Table F1).

∓ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + a; Table F1).

≒ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + c; Table F1).

Table F4. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to the bivariate and true point estimates (varying observed measures e and c; variance predicted by c and e equal).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Post Matching (e(1) + c(1) + a)	Post Matching (e(2) + c(2) + a)	Post Matching (e(3) + c(3) + a)	Post Matching (e(1) + c(2) + a)	Post Matching (e(1) + c(3) + a)	Post Matching (e(2) + c(3) + a)	Post Matching (e(2) + c(1) + a)	Post Matching (e(3) + c(1) + a)	Post Matching (e(3) + c(2) + a)
	b	b	b	b	b	b	b	b	b	b
a(.05), c(.45), e(.45)	1.00	1.29	1.24	1.17✢〒	1.26	1.22✢	1.21✢〒	1.27	1.25	1.22✢
a(.10), c(.43), e(.43)	1.00	1.29	1.24✢	1.17✢〒≒	1.26	1.23✢	1.20✢〒≒	1.27	1.25	1.22✢
a(.15), c(.40), e(.40)	1.00	1.29	1.24✢	1.17✢〒≒	1.26	1.23✢〒≒	1.20✢〒≒	1.27	1.25	1.22✢〒≒
a(.20), c(.38), e(.38)	1.00	1.29	1.24✢	1.17✢〒≒	1.26✢	1.23✢〒≒	1.20✢〒≒	1.27	1.25	1.22✢〒≒
a(.25), c(.35), e(.35)	1.00	1.29	1.24✢〒≒	1.17✢〒≒	1.27✢	1.23✢〒≒	1.20✢〒≒	1.27	1.25	1.22✢〒≒
a(.30), c(.33), e(.33)	1.00	1.30	1.24✢〒≒	1.17✢〒≒	1.26✢	1.23✢〒≒	1.21✢〒≒	1.27	1.25	1.22✢〒≒
a(.32), c(.32), e(.32)	1.00	1.30	1.24✢〒≒	1.17✢〒≒	1.27✢	1.23✢〒≒	1.21✢〒≒	1.28	1.25	1.22✢〒≒
a(.35), c(.30), e(.30)	1.00	1.29	1.24✢〒≒	1.17✢〒≒	1.27✢	1.23✢〒≒	1.21✢〒≒	1.27	1.25	1.22✢〒≒
a(.40), c(.28), e(.28)	1.00	1.30	1.24✢≒	1.17✢〒≒	1.27	1.23✢〒≒	1.20✢〒≒	1.27	1.25	1.21✢〒≒
a(.45), c(.25), e(.25)	1.00	1.30	1.24✢≒	1.17✢〒≒	1.27	1.23✢≒	1.20✢〒≒	1.28	1.25	1.21✢〒≒
a(.50), c(.23), e(.23)	1.00	1.30	1.24≒	1.17✢〒≒	1.27	1.23≒	1.20✢≒	1.27	1.25	1.21✢≒
a(.55), c(.20), e(.20)	1.00	1.30	1.24≒	1.17✢≒	1.27	1.23≒	1.20≒	1.17	1.25	1.22≒
a(.60), c(.18), e(.18)	1.00	1.29	1.24≒	1.17≒	1.26	1.23≒	1.20≒	1.27	1.25	1.21≒
a(.65), c(.15), e(.15)	1.00	1.29	1.24≒	1.16≒	1.26	1.23≒	1.20≒	1.27	1.25	1.20≒

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such $t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*\varepsilon)$; The formula predicting the dependent variable (y) was specified as such $y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*\varepsilon)$. Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. The number in the bracket represents the number of observed variables included in the PSM model. All of the 95%CI ranged between -.01 and +.01 of the point estimate. Starting N = 50,000 cases. N size varies minimally between each point estimate.

✢ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (c + a; Table F1).

〒 point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + a; Table F1).

≒ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + c; Table F1).

Table F5. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to the bivariate and true point estimates (all measures observed; variance predicted by e triple that of c).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Bivariate (y ~ t)	Post Matching (e)	Post Matching (c)	Post Matching (a)	Post Matching (e + c)	Post Matching (e + a)	Post Matching (c + a)	Post Matching (e + c + a)
	b	b	b	b	b	b	b	b	b
a(.05), c(.22), e(.68)	1.00	1.33	1.28	1.28	1.32	1.05	1.11	1.26	1.01
a(.10), c(.21), e(.64)	1.00	1.34	1.32	1.29	1.32	1.08	1.12	1.26	1.01
a(.15), c(.20), e(.60)	1.00	1.35	1.34	1.31	1.32	1.12	1.13	1.26	1.01
a(.20), c(.19), e(.56)	1.00	1.37	1.34	1.32	1.32	1.15	1.15	1.27	1.01
a(.25), c(.18), e(.53)	1.00	1.38	1.34	1.33	1.32	1.17	1.17	1.26	1.01
a(.30), c(.16), e(.49)	1.00	1.38	1.33	1.34	1.32	1.19	1.19	1.26	1.01
a(.32), c(.32), e(.32)	1.00	1.41	1.34	1.35	1.36	1.24	1.25	1.27	1.01
a(.35), c(.15), e(.45)	1.00	1.39	1.32	1.35	1.32	1.21	1.21	1.26	1.01
a(.40), c(.14), e(.41)	1.00	1.39	1.30	1.35	1.32	1.22	1.23	1.25	1.01
a(.45), c(.13), e(.38)	1.00	1.38	1.30	1.35	1.33	1.22	1.21	1.24	1.01
a(.50), c(.11), e(.34)	1.00	1.37	1.29	1.34	1.33	1.23	1.16	1.22	1.01
a(.55), c(.10), e(.30)	1.00	1.36	1.28	1.33	1.33	1.23	1.12	1.20	1.01
a(.60), c(.09), e(.26)	1.00	1.35	1.27	1.32	1.33	1.23	1.09	1.17	1.01
a(.65), c(.08), e(.23)	1.00	1.33	1.26	1.31	1.32	1.23	1.06	1.15	1.01

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such $t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*\varepsilon)$; The formula predicting the dependent variable (y) was specified as such $y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*\varepsilon)$. Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. All of the 95%CI ranged between -.01 and +.01 of the point estimate. Starting N = 50,000 cases. N size varies minimally between each point estimate.

Table F6. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to the bivariate and true point estimates (varying observed measures e; variance predicted by e triple that of c).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Post Matching (e(1) + a)	Post Matching (e(1) + c(4) + a)	Post Matching (e(2) + a)	Post Matching (e(2) + c(4) + a)	Post Matching (e(3) + a)	Post Matching (e(3) + c(4) + a)
	b	b	b	b	b	b	b
a(.05), c(.22), e(.68)	1.00	1.29	1.22✢	1.27	1.18✢	1.25✢	1.13✢
a(.10), c(.21), e(.64)	1.00	1.29	1.22✢	1.27	1.18✢	1.25✢	1.13✢
a(.15), c(.20), e(.60)	1.00	1.29	1.22✢	1.27	1.18✢	1.26✢	1.13✢∓
a(.20), c(.19), e(.56)	1.00	1.29	1.22✢	1.27✢	1.18✢	1.25✢	1.13✢∓≒
a(.25), c(.18), e(.53)	1.00	1.30	1.22✢	1.28	1.18✢	1.25✢	1.13✢∓≒
a(.30), c(.16), e(.49)	1.00	1.30	1.22✢	1.28	1.18✢∓≒	1.26✢	1.13✢∓≒
a(.32), c(.32), e(.32)	1.00	1.34	1.22✢∓≒	1.32	1.18✢∓≒	1.29	1.13✢∓≒
a(.35), c(.15), e(.45)	1.00	1.30	1.22✢	1.28	1.18✢∓≒	1.26✢	1.13✢∓≒
a(.40), c(.14), e(.41)	1.00	1.30	1.22✢∓≒	1.28	1.18✢∓≒	1.26	1.13✢∓≒
a(.45), c(.13), e(.38)	1.00	1.30	1.22✢≒	1.28	1.18✢∓≒	1.26	1.13✢∓≒
a(.50), c(.11), e(.34)	1.00	1.30	1.22✢≒	1.28	1.18✢≒	1.26	1.13✢∓≒
a(.55), c(.10), e(.30)	1.00	1.31	1.22≒	1.28	1.18✢≒	1.26	1.13✢≒
a(.60), c(.09), e(.26)	1.00	1.30	1.23≒	1.28	1.19≒	1.25	1.13✢≒
a(.65), c(.08), e(.23)	1.00	1.30	1.22≒	1.28	1.18≒	1.25	1.12✢≒

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*ε); The formula predicting the dependent variable (y) was specified as such y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*ε). Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. The number in the bracket represents the number of observed variables included in the PSM model. All of the 95%CI ranged between -.01 and +.01 of the point estimate. Starting N = 50,000 cases. N size varies minimally between each point estimate.

✢ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (c + a; Table F5).

∓ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + a; Table F5).

≒ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + c; Table F5).

Table F7. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to the bivariate and true point estimates (varying observed measures c; variance predicted by e triple that of c).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Post Matching (c(1) + a)	Post Matching (e(4) + c(1) + a)	Post Matching (c(2) + a)	Post Matching (e(4) + c(2) + a)	Post Matching (c(3) + a)	Post Matching (e(4) + c(3) + a)
	b	b	b	b	b	b	b
a(.05), c(.22), e(.68)	1.00	1.29	1.22✢	1.28	1.18✢	1.26✢	1.13✢
a(.10), c(.21), e(.64)	1.00	1.29	1.22✢	1.27	1.18✢	1.26✢	1.12✢∓
a(.15), c(.20), e(.60)	1.00	1.29	1.22✢	1.28	1.19✢	1.26✢	1.12✢∓≒
a(.20), c(.19), e(.56)	1.00	1.30	1.22✢	1.28	1.18✢	1.27✢	1.12✢∓≒
a(.25), c(.18), e(.53)	1.00	1.30	1.22✢	1.28	1.18✢	1.27	1.12✢∓≒
a(.30), c(.16), e(.49)	1.00	1.30	1.22✢	1.28	1.18✢∓≒	1.27	1.13✢∓≒
a(.32), c(.32), e(.32)	1.00	1.33	1.22✢≒	1.31	1.19✢∓≒	1.29	1.13✢∓≒
a(.35), c(.15), e(.45)	1.00	1.30	1.21✢∓≒	1.28	1.17✢∓≒	1.27	1.12✢∓≒
a(.40), c(.14), e(.41)	1.00	1.30	1.21✢≒	1.29	1.17✢∓≒	1.27	1.11✢∓≒
a(.45), c(.13), e(.38)	1.00	1.30	1.21✢∓≒	1.29	1.17✢∓≒	1.27	1.11✢∓≒
a(.50), c(.11), e(.34)	1.00	1.30	1.21✢≒	1.29	1.17✢≒	1.27	1.11✢∓≒
a(.55), c(.10), e(.30)	1.00	1.30	1.20✢≒	1.29	1.16✢≒	1.27	1.10✢∓≒
a(.60), c(.09), e(.26)	1.00	1.31	1.20≒	1.29	1.15✢≒	1.27	1.10✢≒
a(.65), c(.08), e(.23)	1.00	1.30	1.19≒	1.29	1.14✢≒	1.27	1.08✢≒

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*ε); The formula predicting the dependent variable (y) was specified as such y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*ε). Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. The number in the bracket represents the number of observed variables included in the PSM model. All of the 95%CI ranged between -.01 and +.01 of the point estimate. Starting N = 50,000 cases. N size varies minimally between each point estimate.

✢ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (c + a; Table F5).

∓ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + a; Table F5).

≒ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + c; Table F5).

Table F8. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to the bivariate and true point estimates (varying observed measures e and c; variance predicted by e triple that of c).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Post Matching (e(1) + c(1) + a)	Post Matching (e(2) + c(2) + a)	Post Matching (e(3) + c(3) + a)	Post Matching (e(1) + c(2) + a)	Post Matching (e(1) + c(3) + a)	Post Matching (e(2) + c(3) + a)	Post Matching (e(2) + c(1) + a)	Post Matching (e(3) + c(1) + a)	Post Matching (e(3) + c(2) + a)
	b	b	b	b	b	b	b	b	b	b
a(.05), c(.22), e(.68)	1.00	1.26✢	1.21✢	1.15✢	1.24✢	1.21✢	1.19✢	1.24✢	1.22✢	1.19✢
a(.10), c(.21), e(.64)	1.00	1.26✢	1.21✢	1.16✢	1.24✢	1.21✢	1.19✢	1.24✢	1.22✢	1.19✢
a(.15), c(.20), e(.60)	1.00	1.26✢	1.21✢	1.16✢	1.24✢	1.21✢	1.19✢	1.24✢	1.22✢	1.19✢
a(.20), c(.19), e(.56)	1.00	1.26✢	1.22✢	1.15✢∓⏥	1.24✢	1.22✢	1.19✢	1.24✢	1.22✢	1.19✢
a(.25), c(.18), e(.53)	1.00	1.26✢	1.21✢	1.15✢∓⏥	1.24✢	1.22✢	1.19✢	1.24✢	1.22✢	1.19✢
a(.30), c(.16), e(.49)	1.00	1.26✢	1.22✢	1.15✢∓⏥	1.24✢	1.22✢	1.19✢⏥	1.24✢	1.22✢	1.19✢∓⏥
a(.32), c(.32), e(.32)	1.00	1.30	1.24✢∓⏥	1.17✢∓⏥	1.27✢	1.23✢∓⏥	1.21✢⏥	1.28	1.25✢∓	1.22✢∓⏥
a(.35), c(.15), e(.45)	1.00	1.26✢	1.22✢	1.15✢∓⏥	1.24✢	1.22✢	1.19✢⏥	1.24✢	1.22✢⏥	1.19✢∓⏥
a(.40), c(.14), e(.41)	1.00	1.26	1.21✢∓⏥	1.15✢∓⏥	1.24✢	1.21✢∓⏥	1.19✢⏥	1.24✢	1.22✢∓⏥	1.19✢∓⏥
a(.45), c(.13), e(.38)	1.00	1.26	1.22✢⏥	1.15✢∓⏥	1.24✢	1.22✢⏥	1.19✢⏥	1.24✢	1.21✢⏥	1.19✢∓⏥
a(.50), c(.11), e(.34)	1.00	1.26	1.21✢⏥	1.15✢∓⏥	1.24	1.21✢⏥	1.19✢⏥	1.24	1.21✢⏥	1.19✢⏥
a(.55), c(.10), e(.30)	1.00	1.26	1.21⏥	1.15✢⏥	1.24	1.21⏥	1.19✢⏥	1.24	1.21⏥	1.19✢⏥
a(.60), c(.09), e(.26)	1.00	1.26	1.21⏥	1.14✢⏥	1.24	1.21⏥	1.18⏥	1.23⏥	1.21⏥	1.18⏥
a(.65), c(.08), e(.23)	1.00	1.26	1.20⏥	1.14✢⏥	1.23	1.22⏥	1.18⏥	1.23⏥	1.21⏥	1.18⏥

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*ε); The formula predicting the dependent variable (y) was specified as such y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*ε). Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. The number in the bracket represents the number of observed variables included in the PSM model. All of the 95%CI ranged between -.01 and +.01 of the point estimate. Starting N = 50,000 cases. N size varies minimally between each point estimate.

✢ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (c + a; Table F5).

∓ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + a; Table F5).

⏥ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + c; Table F5).

Table F9. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to the bivariate and true point estimates (all measures observed; variance predicted by c triple that of e).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Bivariate (y ~ t)	Post Matching (e)	Post Matching (c)	Post Matching (a)	Post Matching (e + c)	Post Matching (e + a)	Post Matching (c + a)	Post Matching (e + c + a)
	b	b	b	b	b	b	b	b	b
a(.05), c(.68), e(.22)	1.00	1.32	1.25	1.30	1.31	1.04	1.24	1.13	1.01
a(.10), c(.64), e(.21)	1.00	1.33	1.27	1.34	1.31	1.07	1.24	1.14	1.01
a(.15), c(.60), e(.20)	1.00	1.35	1.29	1.35	1.31	1.11	1.24	1.16	1.01
a(.20), c(.56), e(.19)	1.00	1.36	1.30	1.36	1.31	1.14	1.24	1.17	1.01
a(.25), c(.53), e(.18)	1.00	1.37	1.31	1.35	1.31	1.17	1.24	1.19	1.01
a(.30), c(.49), e(.16)	1.00	1.38	1.32	1.34	1.32	1.19	1.24	1.22	1.01
a(.32), c(.32), e(.32)	1.00	1.41	1.34	1.35	1.36	1.24	1.25	1.27	1.01
a(.35), c(.45), e(.15)	1.00	1.38	1.33	1.33	1.32	1.21	1.23	1.24	1.01
a(.40), c(.41), e(.14)	1.00	1.38	1.33	1.32	1.32	1.22	1.22	1.26	1.01
a(.45), c(.38), e(.13)	1.00	1.37	1.33	1.31	1.32	1.22	1.21	1.24	1.01
a(.50), c(.34), e(.11)	1.00	1.37	1.33	1.30	1.32	1.23	1.19	1.20	1.01
a(.55), c(.30), e(.10)	1.00	1.35	1.32	1.28	1.32	1.23	1.17	1.15	1.01
a(.60), c(.26), e(.09)	1.00	1.34	1.31	1.28	1.32	1.23	1.14	1.11	1.01
a(.65), c(.23), e(.08)	1.00	1.32	1.30	1.27	1.32	1.23	1.12	1.08	1.01

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such $t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*\varepsilon)$; The formula predicting the dependent variable (y) was specified as such $y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*\varepsilon)$. Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. All of the 95%CI ranged between -.01 and +.01 of the point estimate. Starting N = 50,000 cases. N size varies minimally between each point estimate.

Table F10. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to the bivariate and true point estimates (varying observed measures e; variance predicted by c triple that of e).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Post Matching (e(1) + a)	Post Matching (e(1) + c(4) + a)	Post Matching (e(2) + a)	Post Matching (e(2) + c(4) + a)	Post Matching (e(3) + a)	Post Matching (e(3) + c(4) + a)
	b	b	b	b	b	b	b
a(.05), c(.68), e(.22)	1.00	1.30	1.22∓	1.29	1.18∓	1.27	1.13∔∓
a(.10), c(.64), e(.21)	1.00	1.31	1.23∓	1.29	1.18∓	1.28	1.13∔∓
a(.15), c(.60), e(.20)	1.00	1.31	1.22∓	1.29	1.18∓	1.28	1.13∔∓
a(.20), c(.56), e(.19)	1.00	1.31	1.22∓	1.29	1.18∓	1.28	1.12∔∓≒
a(.25), c(.53), e(.18)	1.00	1.31	1.22∓	1.29	1.18∔∓	1.28	1.12∔∓≒
a(.30), c(.49), e(.16)	1.00	1.31	1.22∔∓	1.29	1.18∔∓≒	1.28	1.12∔∓≒
a(.32), c(.32), e(.32)	1.00	1.34	1.22∔∓	1.32	1.18∔∓≒	1.29	1.14∔∓≒
a(.35), c(.45), e(.15)	1.00	1.31	1.22∔∓	1.30	1.17∔∓≒	1.28	1.12∔∓≒
a(.40), c(.41), e(.14)	1.00	1.31	1.22∔∓≒	1.30	1.17∔∓≒	1.28	1.11∔∓≒
a(.45), c(.38), e(.13)	1.00	1.31	1.21∔∓≒	1.30	1.17∔∓≒	1.28	1.11∔∓≒
a(.50), c(.34), e(.11)	1.00	1.32	1.21≒	1.30	1.16∔∓≒	1.28	1.11∔∓≒
a(.55), c(.30), e(.10)	1.00	1.32	1.2≒	1.30	1.16∓≒	1.29	1.11∔∓≒
a(.60), c(.26), e(.09)	1.00	1.32	1.18≒	1.31	1.14∓≒	1.29	1.08∔∓≒
a(.65), c(.23), e(.08)	1.00	1.32	1.19≒	1.30	1.15≒	1.28	1.09∓≒

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*ε); The formula predicting the dependent variable (y) was specified as such y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*ε). Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. The number in the bracket represents the number of observed variables included in the PSM model. All of the 95%CI ranged between -.01 and +.01 of the point estimate. Starting N = 50,000 cases. N size varies minimally between each point estimate.

∔ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (c + a; Table F9).

∓ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + a; Table F9).

≒ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + c; Table F9).

Table F11. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to the bivariate and true point estimates (varying observed measures c; variance predicted by c triple that of e).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Post Matching (c(1) + a)	Post Matching (e(4) + c(1) + a)	Post Matching (c(2) + a)	Post Matching (e(4) + c(2) + a)	Post Matching (c(3) + a)	Post Matching (e(4) + c(3) + a)
	b	b	b	b	b	b	b
a(.05), c(.68), e(.22)	1.00	1.29	1.22∓	1.27	1.18∓	1.25	1.13∔∓
a(.10), c(.64), e(.21)	1.00	1.29	1.22∓	1.27	1.18∓	1.25	1.13∔∓
a(.15), c(.60), e(.20)	1.00	1.29	1.22∓	1.27	1.18∓	1.25	1.13∔∓
a(.20), c(.56), e(.19)	1.00	1.29	1.22∓	1.27	1.18∓	1.25	1.14∔∓≒
a(.25), c(.53), e(.18)	1.00	1.29	1.22∓	1.27	1.18∔∓	1.25	1.13∔∓≒
a(.30), c(.49), e(.16)	1.00	1.29	1.22∔∓	1.27	1.19∔∓≒	1.25	1.13∔∓≒
a(.32), c(.32), e(.32)	1.00	1.33	1.22∔∓	1.31	1.19∔∓≒	1.29	1.13∔∓≒
a(.35), c(.45), e(.15)	1.00	1.30	1.22∔∓	1.27	1.18∔∓≒	1.25	1.13∔∓≒
a(.40), c(.41), e(.14)	1.00	1.29	1.22∔∓≒	1.27	1.19∔∓≒	1.25∔	1.13∔∓≒
a(.45), c(.38), e(.13)	1.00	1.30	1.22∔≒	1.27	1.19∔∓≒	1.25	1.13∔∓≒
a(.50), c(.34), e(.11)	1.00	1.30	1.22≒	1.27	1.18∔∓≒	1.25	1.13∔∓≒
a(.55), c(.30), e(.10)	1.00	1.30	1.22≒	1.27	1.18≒	1.24	1.13∔∓≒
a(.60), c(.26), e(.09)	1.00	1.30	1.22≒	1.26	1.18≒	1.24	1.12∓≒
a(.65), c(.23), e(.08)	1.00	1.29	1.22≒	1.27	1.18≒	1.24	1.13∓≒

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*ε); The formula predicting the dependent variable (y) was specified as such y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*ε). Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. The number in the bracket represents the number of observed variables included in the PSM model. All of the 95%CI ranged between -.01 and +.01 of the point estimate. Starting N = 50,000 cases. N size varies minimally between each point estimate.

∔ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (c + a; Table F9).

∓ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + a; Table F9).

≒ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + c; Table F9).

Table F12. Simulated comparison of the point estimates achieved from post-matching bivariate regressions to the bivariate and true point estimates (varying observed measures e and c; variance predicted by c triple that of e).

Variance predicted (s^2) in t (a,c,e)	True (y ~ t)	Post Matching (e(1) + c(1) + a)	Post Matching (e(2) + c(2) + a)	Post Matching (e(3) + c(3) + a)	Post Matching (e(1) + c(2) + a)	Post Matching (e(1) + c(3) + a)	Post Matching (e(2) + c(3) + a)	Post Matching (e(2) + c(1) + a)	Post Matching (e(3) + c(1) + a)	Post Matching (e(3) + c(2) + a)
	b	b	b	b	b	b	b	b	b	b
a(.05), c(.68), e(.22)	1.00	1.26	1.22∓	1.16∓	1.23∓	1.20∓	1.18∓	1.25	1.24∓	1.20∓
a(.10), c(.64), e(.21)	1.00	1.26	1.22∓	1.16∓	1.24∓	1.20∓	1.18∓	1.25	1.24∓	1.20∓
a(.15), c(.60), e(.20)	1.00	1.26	1.22∓	1.16∓	1.24∓	1.20∓	1.18∓	1.25	1.24∓	1.20∓
a(.20), c(.56), e(.19)	1.00	1.27	1.22∓	1.16✛∓	1.24∓	1.20∓	1.18∓	1.25	1.24∓	1.20∓
a(.25), c(.53), e(.18)	1.00	1.27	1.22∓	1.16✛∓≒	1.24∓	1.20∓	1.18✛∓	1.25	1.24∓	1.20∓
a(.30), c(.49), e(.16)	1.00	1.27	1.22✛∓	1.16✛∓≒	1.24∓	1.20✛∓	1.18✛∓≒	1.25	1.24∓	1.20✛∓
a(.32), c(.32), e(.32)	1.00	1.30	1.24✛∓≒	1.17✛∓≒	1.27✛	1.23✛∓≒	1.21✛∓≒	1.28	1.25✛	1.22✛∓≒
a(.35), c(.45), e(.15)	1.00	1.27	1.22✛∓≒	1.16✛∓≒	1.24✛	1.20✛∓≒	1.18✛∓≒	1.25	1.24✛	1.20✛∓≒
a(.40), c(.41), e(.14)	1.00	1.27	1.22✛∓≒	1.16✛∓≒	1.24✛	1.20✛∓≒	1.18✛∓≒	1.26✛	1.24✛	1.20✛∓≒
a(.45), c(.38), e(.13)	1.00	1.27	1.22✛≒	1.16✛∓≒	1.24✛	1.20✛∓≒	1.18✛∓≒	1.26	1.24✛	1.21✛∓≒
a(.50), c(.34), e(.11)	1.00	1.27	1.22≒	1.16✛∓≒	1.24	1.20✛≒	1.18✛∓≒	1.26	1.24	1.20✛≒
a(.55), c(.30), e(.10)	1.00	1.27	1.22≒	1.16∓≒	1.24	1.20≒	1.18≒	1.26	1.25	1.20≒
a(.60), c(.26), e(.09)	1.00	1.27	1.22≒	1.15≒	1.23≒	1.19≒	1.17≒	1.25	1.24	1.20≒
a(.65), c(.23), e(.08)	1.00	1.27	1.22≒	1.15≒	1.23≒	1.19≒	1.17≒	1.25	1.24	1.20≒

Notes: All simulated data is based on 50,000 cases; The formula predicting the treatment variable (t) was specified as such t = 0 + (s^2*a) + (s^2*c) + (s^2*e) + (.05*ε); The formula predicting the dependent variable (y) was specified as such y = 0 + (1.00*t) + (1.25*a) + (1.25*c) + (1.25*e) + (.005*ε). Four predictor variables were created for e and four predictor variables were created four c. e was specified as such e = 25*x1 +25*x2 +25*x3 + 25*x4, and c was specified as such c = 25*x1 +25*x2 +25*x3 + 25*x4. The number in the bracket represents the number of observed variables included in the PSM model. All of the 95%CI ranged between -.01 and +.01 of the point estimate. Starting N = 50,000 cases. N size varies minimally between each point estimate.

✛ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (c + a; Table F9).

∓ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + a; Table F9).

≒ point estimate equal to, or closer to the true point estimate than point estimate derived from matching with (e + c; Table F9).

Appendix G: Interpretation of GPS Balancing Results

To emphasize the consistent percent reduction in bias across the majority of the comparisons, the current discussion will focus on the few comparisons in which the bias increased when moving from the pre-matching sample to the post-matching sample. In reference to the 10th percentile (intelligence = 7 – 79) the pre-matching sample achieved a smaller mean difference on paternal conflict than the post-matching sample (pre-matching: $\bar{X}\Delta$ = -.029, *t-value* = -.608; post-matching: $\bar{X}\Delta$ = -.049, *t-value* = -1.826), resulting in approximately a 71 percent increase in bias. For the 30th percentile (intelligence = 88 – 91) the pre-matching sample achieved a smaller mean difference on male (pre-matching: $\bar{X}\Delta$ = .002, *t-value* = .182; post-matching: $\bar{X}\Delta$ = .003, *t-value* = .460) and paternal conflict (pre-matching: $\bar{X}\Delta$ = -.027, *t-value* = -.576; post-matching: $\bar{X}\Delta$ = -.033, *t-value* = -1.677) than the post-matching sample,), resulting in approximately a 23 percent increase in bias for both variables.

Concerning the 40th percentile (intelligence = 92 – 96) the pre-matching sample achieved a smaller mean difference on age (pre-matching: $\bar{X}\Delta$ = .035, *t-value* = .601; post-matching: $\bar{X}\Delta$ = .059, *t-value* = 2.150), paternal conflict (pre-matching: $\bar{X}\Delta$ = .006, *t-value* = .127; post-matching: $\bar{X}\Delta$ = -.020, *t-value* = -.918), and social support (pre-matching: $\bar{X}\Delta$ = .011, *t-value* = .284; post-matching: $\bar{X}\Delta$ = -.012, *t-value* = -.678) than the post-matching sample. The evidence suggests that there was approximately a 70 percent increase in bias for age, a 209 percent increase in bias for paternal conflict, and a 10 percent increase in bias for social support. For the 50th percentile (intelligence = 97 – 103) the pre-matching sample achieved a smaller mean difference on maternal conflict (pre-matching: $\bar{X}\Delta$ = .001, *t-value* = .041; post-matching: $\bar{X}\Delta$ = .007, *t-value* = .451) and school attachment (pre-matching: $\bar{X}\Delta$ = -.009, *t-value* = -.321; post-matching: $\bar{X}\Delta$ = -.010, *t-value* = -.813) than the post-matching sample, resulting in approximately a 427 and a 17 percent increase in bias, respectively.

In reference to the 60th percentile (intelligence = 104 – 106) the pre-matching sample achieved a smaller mean difference on social support (pre-matching: $\bar{X}\Delta$ = .006, *t-value* = .188; post-matching: $\bar{X}\Delta$ = -.008, *t-value* = -.554) than the post-matching sample, resulting in approximately a 24 percent increase in bias. Concerning the 80th percentile (intelligence = 109 – 111) the pre-matching sample achieved a smaller mean difference on social support (pre-matching: $\bar{X}\Delta$ = -.006, *t-value* = -.164; post-matching: $\bar{X}\Delta$ = -.011, *t-value* = -.618), peer drug use (pre-matching: $\bar{X}\Delta$ = .007, *t-value* = .219; post-matching: $\bar{X}\Delta$ = .008, *t-value* = .570), and baseline delinquency (pre-matching: $\bar{X}\Delta$ = .004, *t-value* = .192; post-matching: $\bar{X}\Delta$ = .011, *t-value* = 1.051) than the post-matching sample. The evidence suggests that there was approximately a 70 percent increase in bias for social support, a 16 percent increase in bias for peer drug use, and a 153 percent increase in bias for baseline delinquency.

For the 90th percentile (intelligence = 112 – 116) the pre-matching sample achieved a smaller mean difference on maternal conflict (pre-matching: $\bar{X}\Delta$ = .002, *t-value* = .053; post-matching: $\bar{X}\Delta$ = -.014, *t-value* = .923), social support (pre-matching: $\bar{X}\Delta$ = .007, *t-value* = .248; post-matching: $\bar{X}\Delta$ = .014, *t-value* = 1.007), and baseline drug use (pre-matching: $\bar{X}\Delta$ = -.001, *t-value* = -.073; post-matching: $\bar{X}\Delta$ = .010, *t-value* = 1.739) than the post-matching sample, resulting in approximately a 690, a 94, and a 977 percent increase in the bias respectively. Concerning the 100th percentile (intelligence = 117 – 122) the pre-matching sample achieved a smaller mean difference on paternal conflict (pre-matching: $\bar{X}\Delta$ = .001, *t-value* = .031; post-matching: $\bar{X}\Delta$ = -.004, *t-value* = -.216), resulting in approximately a 200 percent increase in the bias.