2016

Improving Causal Claims in Observational Research: An Investigation of Propensity Score Methods in Applied Educational Research

Julie Diane Wren
Loyola University Chicago

LOYOLA UNIVERSITY CHICAGO

IMPROVING CAUSAL CLAIMS IN OBSERVATIONAL RESEARCH: AN INVESTIGATION OF PROPENSITY SCORE METHODS IN APPLIED EDUCATIONAL RESEARCH

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

RESEARCH METHODOLOGY

BY

JULIE D. WREN

CHICAGO, ILLINOIS

MAY 2017

ACKNOWLEDGMENTS

I would never have been able to finish my dissertation without the guidance of my committee members, the support of my family and the love of my son.

I would like to express my deepest gratitude to my advisor, Dr. Terri Pigott, her guidance and patience were essential for helping me move this work forward. Additionally, I would like to thank Dr. Susan Farrugia and Dr. Meng-Jia Wu for their feedback throughout this process. Special thanks to the institution that was willing to share their student data to complete this work.

None of this would be possible without the support of my family – particularly the village of women that helped to raise my son while I was at class or working on my dissertation, Thank you to two amazing grandmothers, Eizabeth and Kathy, along with an entourage of aunts, Natalie, Kelly, Katie and Kristin.

Finally, thank you to my son who gleefully went from daycare to family so his mom could work and continue her studies.

For my son, Kelan.

It takes a village.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TERMS

The following definitions are quoted from the Integrated Postsecondary Education Data System (IPEDS) glossary and are available at http://nces.ed.gov/ipeds/glossary/; slight modifications were made to fit the current format.

Cohort refers to a specific group of students established for tracking purposes.

Credit hour refers to a unit of measure representing the equivalent of an hour (50 minutes) of instruction per week over the entire term. It is applied toward the total number of credit hours needed for completing the requirements of a degree, diploma, certificate or other formal award.

Degree/certificate-seeking students refers to students enrolled in courses for credit who are recognized by the institution as seeking a degree or other formal award.

Entering students (undergraduates) refers to students at the undergraduate level, both fulltime and part-time, coming into the institution for the first time in the fall term (or the prior summer term who returned again in the fall). This includes all first-time undergraduate students, students transferring into the institution at the undergraduate level for the first time, and non-degree/certificate seeking undergraduates entering in the fall.

First-time students (undergraduates) refers to students who have no prior postsecondary experience (except as noted below) attending any institution for the first time at the undergraduate level. This includes students enrolled in academic or occupational programs. It also includes students enrolled in the fall term who attended college for the first time in the prior summer term, and students who entered with advanced standing (college credits earned before high school graduation).

Four-year institutions refers to postsecondary institutions that offer programs of at least four years duration or programs at or above the baccalaureate level. Thus, schools that offer post baccalaureate certificates only or those that offer graduate programs only are also included. In addition, free-standing medical, law or other first-professional schools are considered four-year institutions.

Fall cohort refers to the group of students entering in the fall term established for tracking purposes.

Fall term refers to the part of the academic year that begins between late August and November 1.

Fulltime students (undergraduates) refers to students enrolled for 12 or more semester credits, or 12 or more quarter credits, or 24 or more contact hours a week each term.

Postsecondary education refers to the provision of a formal instructional program whose curriculum is designed primarily for students who are beyond the compulsory age for high school. This includes programs whose purpose is academic, vocational, and continuing professional education, and excludes vocational and adult basic education programs.

Public institutions refers to educational institutions whose programs and activities are operated by a publicly elected or appointed school official and are primarily supported by public funds.

Undergraduate refers to a student enrolled in 4- or 5-year bachelor’s degree program, an associate degree program or a vocational or technical program below the baccalaureate.

CHAPTER ONE

INTRODUCTION

Educational researchers often face the challenge of determining the efficacy of a program, treatment, or intervention (hereto referred to as treatment) on a desired outcome (Murnane & Willett, 2011). These research questions often aim to explain whether or not treatment X caused outcome Y, but to investigate causal relationships, three requirements must be met. The requirements are: (1) the cause must precede the effect, (2) the cause must be related to the effect, and (3) no other plausible explanation exists except the causal explanation (Shadish, Campbell, & Cook, 2002, p.6). Although the first two requirements are relatively straightforward, the third requirement is much more difficult to ascertain.

The need to rule out all other probable explanations to make a causal claim is why random assignment is referred to as the gold standard (Murnane & Willett, 2011; Shadish, Campbell & Cook, 2002). Random assignment, if employed properly, has the benefit of balancing the observed and unobserved covariates between groups, making any differences between the groups arbitrary (Rubin, 1974, p. 694). This balancing ability of random assignment is critical as it ensures that the groups are equal in expectation thus bolstering confidence that the third requirement of causation, no other plausible explanation exists except the causal explanation, has been met.

Although random assignment provides the best support for ensuring that there are no other probable explanations, randomized experiments are less common in educational research due to financial, practical and ethical concerns (Murnane & Willett, 2011; Shadish, Campbell & Cook, 2002). These challenges and concerns have led to a reliance on observational research for educational inquiry.

Observational Research

Since observational research does not involve random assignment, it is subject to selection bias (Shadish, Campbell & Cook, 2002). Selection bias is systematic bias that results from individuals electing rather than being assigned to participate. Consider a new curriculum developed to improve reading levels. In the school where the reading program was administered, students whose parents signed them up to participate received the curriculum. At the end of the year, the students that participated in the program had demonstrated higher reading scores. Although the reading program might have had to led to these improvements, it is possible that other factors led to these differences. Taking a look at the two groups of students, students who participated in the reading program were more likely to be female and have more than 50 books in the home and less likely to demonstrate financial need. Rather than the differences in the treatment outcome resulting from the reading program, the improvements might be the result of the financial, social and educational advantages the children who participated were afforded by birth rather than the program. In this instance, parental affluence would be a confounding variable. To determine the impact of the program on performance, the variation in the outcome due to the confounding variable must be controlled for or removed from the analysis.

Observational research does not, by design, provide substantial evidence that there are no other probable explanations. Therefore, there is incongruence between the most popular design choice and the needs of educational researchers. Educational researchers need to be able to attest to the impact of treatment on individuals; therefore, the study of methodological and/or statistical approaches to allow for the investigation of causal inference is both critical and necessary.

Statement of the Problem

Due to the expense and ethical concerns associated with randomized research, causal questions are often addressed without the benefits of random assignment. Often, researchers attempt to minimize the impact of selection bias by controlling for the differences between groups on key covariates with regression (Morgan & Winship, 2007). Unlike random assignment, where the balancing between groups occurs before the analysis, regression balances and analyzes at the same time. While regression can provide information about the association between a treatment and an outcome, it cannot substantiate causal claims when used alone.

Causal Claims in Observational Research

Although regression, used as a statistical tool, does not allow for causal claims, it is powerful when combined with alternative design features such as regression discontinuity and instrument variable estimation. Regression discontinuity exploits exogenous characteristics of a treatment to support causal claims (Thistlethwaite & Campbell, 1960). Again, consider the new reading program. If a cutoff score was required for participation, then regression discontinuity could be employed. The cutoff score serves as the exogenous characteristic, and the analysis would focus on the students at and around the cutoff. The exogenous characteristic is both a necessary and limiting aspect of regression discontinuity. It is necessary because focusing on this smaller area, just around the cut off, allows for causal claims to be made.

Although causal claims can be made, they are bounded to the individuals closely surrounding the cutoff score, limiting generalizability and resulting in a local average treatment effect (Thistlethwaite & Campbell, 1960).

Propensity Score

In addition to alternative design features, statistical procedures that do not require design modifications can be employed. Based on the early work of Neyman in 1923 and Fisher in 1925, Rubin (1974) developed Rubin's Causal Model (RCM). Rubin framed all investigations of causal relationships as a missing data issue. Consider the new reading program; regardless of whether students are randomly assigned, students are signed up by their parents, or a cut off score is employed, each student can only be observed in one condition. Therefore, a student that is participating in the new reading program cannot also be observed for not participating in the new reading program. So for each student that participates in the reading program the outcome is known; but for that same student, the outcome for not participating in the reading program is unknown. This is why causal inference can be conceptualized as a missing data problem. Since the missing data can never be fully known, the goal becomes devising a set of conditions in which the missing data can be closely approximated.

Although random assignment is the gold standard, it is not always feasible or desirable. When random assignment is not possible, the principles, derived by Rubin (1974), can be applied to model the bias (i.e., selection process) (Rosenbaum & Rubin, 1983a). Modeling the selection process has the advantage of approximating random assignment because, like random assignment, the selection process is analyzed prior to the outcome. Consider the new reading program; parents had chosen whether or not to have their children participate, and initial results indicated a favorable outcome among students in the reading program. Although there was a positive treatment effect, it is unclear whether or not the outcome is a result of the reading program or the selection process because there were significant differences between the groups at the outset of the study. Rather than controlling for these observed differences between groups, which is a common strategy, the selection process can be modeled. Regression is often used to model the selection bias with the summation of this process resulting in a single score, known as a propensity score.

A propensity score is the "conditional probability of assignment to a particular group, given a vector of covariates" (Rosenbaum & Rubin, 1983 p. 42). Propensity score methods are different than regression because they use a single value to create non-equivalent groups. Therefore, unlike regression, the bias between the groups before and after propensity score methods can be assessed.

Although propensity score methods offer an alternative to experimental designs for causal analysis, its utility is based upon successfully proving that the two assumptions have been met: the stable unit treatment value assumption (SUTVA) and strongly ignorable treatment assignment (Rosenbaum & Rubin, 1983a; Rubin, 1980). The SUTVA assumption asserts that there is only one version of treatment and no interference between units (Cox, 1958, p. 19; Rubin, 1980, p.591). This means that the outcome of one unit is not impacted by the treatment of another unit, leaving only two potential outcomes (Little & Rubin, 2000, p.123). In addition to SUTVA, there has to be a strongly ignorable treatment assignment, also known as independence (Rosenbaum & Rubin, 1983a). The assumption of independence requires that the determination of cause (treatment or control) to which a unit is exposed is unrelated to all other variables (Holland, 1986, p.458). Stated alternatively, the treatment assignment is exogenous. Since there is no direct statistical test to ensure that these assumptions have been met, the quality of the methodology and related statistical analysis help to build support that these assumptions have been met.

Purpose of the Study

Although propensity score methods are conceptually simple and easy to understand, ensuring that the selection process is strongly ignorable is a challenge. This study used existing institutional data from a large, urban, public, very high research university to compare sixteen matching schemes, built from three separate datasets, to estimate the propensity score, achieve balance between groups and test the sensitivity of the average treatment effect (ATE). For each PS model, four different conditioning strategies were applied. The first four matching schemes used commonly collected data available within a student information system (referred to as SIS dataset). The next four matching schemes combined the SIS dataset with data from an entering student survey (referred to as ESS dataset). The next four matching schemes, again, combined the SIS dataset with data gathered from a noncognitive survey (referred to as NCS dataset). The final four matching schemes included data from the SIS, ESS and the NCS datasets. Each model builds upon the next, offering additional covariates for the model building process.

To assess the effectiveness of these propensity score techniques in an applied educational research setting, the methodological research questions are nested within the framework of an overarching contextual research question. This guiding research question aimed to understand the extent to which first-time, fulltime students who enrolled in optimal credit levels (defined as 15 or more credit hours during the first term of attendance) experienced greater levels of success. Student success is a complex phenomenon and includes multiple and sometimes competing constructs. This research uses first-year retention as a proxy for student success. Students are considered retained if they were enrolled at the university, the following fall term. Many researchers have studied retention resulting in various models with a diverse set of covariates (e.g., Pascarella & Terenzini, 2005; Ting, 1998; Tinto, 1975; 1993). Although this research does not seek to understand the complexity of student success, it does try to understand the influence the availability of additional covariates has on propensity score techniques and their influence on the stability of findings in applied educational research.

Research Questions

1. To what extent do the treatment and the control groups vary naively across covariates?
2. To what extent do different PS models achieve overlap between the treatment and control groups?
3. To what extent do different PS models and conditioning strategies impact the sample size?
4. To what extent do different PS models and conditioning strategies achieve balance between groups?
5. To what extent do different PS models and conditioning strategies reach the same overall conclusions?
6. To what extent is the average treatment effect robust against unobserved covariates under different PS models and conditioning strategies?

Delimitations. Delimitations of the study include:

1. The study was limited to a large, public, very high research postsecondary institution; therefore, results are not generalizable to other postsecondary institutions.
2. The study was limited to first-time, fulltime students and does not offer information about transfer students or part-time students.
3. Conclusions drawn from the analysis were based solely on student factors that are measurable; other aspects of the student experience derived from a qualitative approach were not included.
4. Each of the conditioning strategies used nearest neighbor, greedy, matching. As a result, no information can be garnered about performance relative to other strategies.

Limitations. Limitations of the study include:

1. Continuation of the analysis is dependent on the performance throughout.
2. Survey data are not an integrated part of the student record system. Therefore, data loss exists as a result of varied survey participation among students.
3. Survey data were gathered using self-report measures. These data only represent students' self-perceptions, and these perceptions are not corroborated by any behavioral indices or additional reporters.

Significance of the Study

This study adds to a growing body of knowledge of the significance of expansive covariate sets and the impact of propensity score techniques in applied educational research. Additionally, it contributes to an underdeveloped area of research, the use of propensity score methods in applied postsecondary institutions. Previous research has demonstrated that simply controlling for covariates does not replicate findings from randomized experiments (Angrist & Pischke, 2009). Therefore, there is a need to explore alternative methodologies for answering routine causal questions that arise in educational research.

Despite the rapid growth of propensity score methods in education, there have been relatively few studies focused on issues within higher education. Those studies that have occurred typically adopt a single-level model (e.g., Clark & Cundiff, 2011; Dehejia & Wahba, 1999) use logistic regression for estimation (e.g., Clark & Cundiff, 2011; Schafer, Wilkinson & Ferraro, 2013; Vaughan, Lalonde & Jenkins-Guarnierie, 2014), and condition the propensity score using matching or stratification (e.g., Clark & Cundiff, 2011; Schafer, Wilkinson & Ferraro, 2013; Vaughan, Lalonde & Jenkins-Guarnierie, 2014). Although there has been some attempt to broaden the application of propensity score methods to hierarchical relationships in this context, studies using multilevel modeling are far fewer (e.g., Vaughan, Lalonde & Jenkins-Guarnierie, 2014; Heil, Reisel & Attewell, 2014). Further, most of the research in this area has focused on a specific research question rather than on the method itself. Although information about the use and utility of propensity score methods exists, based on research using simulated data or multiple arms studies with randomized research as one of those arms, there lacks knowledge about what works within the context. Additionally, there is limited information about how the availability of expansive variable sets can influence the conclusions of a study.

Additionally, not much research has been done on the use of propensity score methods within a single institution, which is of interest to practitioners. When a single institution has been the focus of a research study, many of the necessary elements to judge quality are not included (Ali et al., 2015; Thoemmes & Kim, 2011). This research adds to information about the potential value of expansive datasets while detailing each of the steps for performing and assessing propensity score techniques.

Anticipated Outcomes

Although this study was explorative in nature, differences between the matching schemes were expected. Based on previous research (Steiner, Cook, Shadish and Clark, 2010; Steiner & Cook, 2013), the addition of relevant covariates was expected to impact the findings at various stages of analysis. The inclusion of additional covariates was expected to lead to stronger PS models that better accounted for the selection bias ultimately bolstering confidence in the study’s conclusion. Despite this, the inclusion of the additional covariates was expected to negatively impact sample size and match rate. Although sample loss was expected as more restrictions were placed on the conditioning strategy (i.e., caliper widths), it was unclear whether the conditioning strategies would perform differently across PS models.

CHAPTER TWO

REVIEW OF LITERATURE

This study investigated the availability of an expansive covariate set on propensity score (PS) models and the behavior and performance of propensity score conditioning strategies in applied educational research. Accordingly, the review focuses on causal local institution and the use of propensity score methods in observational research and their appropriateness and utility in applied educational research. To provide a foundation, the historical roots of causal inference and its extension to observational research through Rubin's Causal Model (RCM) are explained. Next, research design choices that aim to understand causal relationships are explored followed by a discussion about the logic and use of the propensity. Additionally, a synthesis of current recommendations for applying propensity score methods and the use of propensity score methods in higher education are discussed. Lastly, the empirical gaps are identified and the ability of this research to bridge this gap will be addressed.

Rubin's Causal Model (RCM)

With roots predating the 16th century, modern science and experimentation evolved from philosophy taking foothold in the 17th century (Shadish, Cook, & Campbell, 2002). As interest moved away from observations about the world, interest moved toward active manipulations and their effect on the phenomenon understudy. As knowledge and interest in experimntation grew so did the desire to control extraneous variables and minimize bias. By the early 1900s, this coalesced into the development of the modern experiment, including both random assignment and control groups (Shadish, Campbell & Cook, 2002). This desire to maximize control helped to make causal inference synonymous with randomized experiments, and it was not until 1974 that causal reasoning was first applied to observational research (Rubin, 1974).

Rubin's Casual Model (RCM; Rubin 1974, 1978), with its potential outcome notation, is an extension of the work of both Neyman in 1923 and Fisher in 1925 (Rubin, 1990). RCM is also referred to as the potential outcomes framework and the counterfactual model of causal inference. Due to Rubin's significant application of this framework to observational research, it will be referred to as RCM throughout (Holland, 1986, p.946). Neyman developed a non-parametric model where each unit had two potential outcomes, and the difference between these outcomes was the causal effect. The specification of two outcomes is particularly helpful since the requirement of two causes (treatment, control) is often taken for granted (Holland, 1986, p.459; Yuke, 1903, p.126). The work of both Neyman and Fisher was rooted in experimental design and was first applied to nonrandomized research by Rubin (1974).

RCM draws attention to the missing data issue formalized in the potential outcomes framework. More formally stated, let Y = the potential outcomes, Z= the indicator for treatment received, i = the unit, and j = the exposed treatment. Therefore, when $(Z = 0, Y_i^0)$ is the potential treatment outcome for the ith unit that received (Z = 0) treatment and $(Z = 1, Y_i^1)$ is the potential treatment outcome for the ith unit that received (Z = 1) treatment. Since a unit cannot be observed in both conditions, Y_i^1 and Y_i^0 are referred to as potential outcomes.

The goal of analysis is to compare these two potential outcomes (Y_i^1, Y_i^0) using an average treatment effect (τ). Depending on the nature of the investigation, the average treatment effect for the overall population (ATE), the average treatment effect for the treated (ATT) or the average treatment effect for the untreated (ATU) might be of interest. The average treatment effect is defined as the expected difference in the potential outcomes with the following,

$$\text{ATE } \tau = E(Y_i^1 - Y_i^0) = E(Y_i^1) - E(Y_i^0)$$

$$\text{ATT } \tau_T = E(Y_i^1 - Y_i^0 | Z_i = 1) = E(Y_i^1 | Z_i = 1) - E(Y_i^0 | Z_i = 1)$$

$$\text{ATU } \tau_U = E(Y_i^1 - Y_i^0 | Z_i = 0) = E(Y_i^1 | Z_i = 0) - E(Y_i^0 | Z_i = 0)$$

If both potential outcomes could be observed, then calculating the average treatment effect would simply be an average of the individual treatment differences. Since this is not the reality, the most that can be calculated is the treatment outcomes for the treated and the control outcomes for the untreated. The simple difference between these two outcomes provides a biased estimator of the average treatment effect. There is no statistical procedure or methodology that can fully resolve this missing data problem.

Assumptions

Since there is no way to completely resolve the missing data issue, there has to be a set of assumptions to allow for causal local institution. As Holland (1986) pointed out, a statistical solution is required in addition to the scientific framework. Specifically, the statistical solution needs to address how information from different units can be used to understand the impact of treatment by supplementing an average causal effect (p.457). The two assumptions necessary within the potential outcomes framework are: the stable unit treatment value assumption (SUTVA) and the strongly ignorable treatment assignment.

Stable Unit Treatment Value Assumption. The SUTVA asserts that there is only one version of treatment and no interference between units (Cox, 1958, p. 19; Rubin, 1980, p.591). This means that the outcome of one unit is not impacted by the treatment of another unit, leaving only two potential outcomes (Little & Rubin, 2000, p.123). This is an essential assumption to ensure that the treatment, as designed, is responsible for the causal effect. In practice, this can be violated.

For instance, consider a summer treatment program for children with behavioral disorders where children are blind to their medication treatment, receiving either a placebo or active pill daily. It is possible that child A receiving a placebo pill could cause increased negative behaviors for child B because child A is disturbing child B due to child A's treatment assignment (placebo). This violation of SUTVA increases the potential outcomes for child B because child B's outcomes would be a function of whether child A received a placebo pill or not as well as his own treatment assignment. The number of outcomes increases exponentially with the number of units (Little & Rubin, 2000, p.123). Therefore, a strong claim for meeting SUTVA is required.

Strongly Ignorable Treatment Assignment. In addition to SUTVA, there must be a strongly ignorable treatment assignment, also known as independence (Rosenbaum & Rubin, 1983a). Since units cannot be observed under both conditions, their assignment to treatment Z must be independent of outcomes (Little & Rubin, 2000, p.125). The assumption of independence requires that the determination of cause (treatment or control) to which a unit is exposed is unrelated to all other variables (Holland, 1986, p.458). Stated alternatively, the treatment assignment is exogenous. When treatment assignment is non-ignorable or endogenous, the selection mechanism must be incorporated into the analysis (Little and Rubin, 2000, p.127). The assignment of units to treatment must be known.

Criticisms

Not all researchers support the use of RCM for making causal claims. One of the major opponents of the potential outcome framework adopted by Rubin is Pearl (2010). Pearl stated the following, "one cannot substantiate causal claims from associations alone, even at the population level—behind every causal conclusion there must lie some causal assumption that is not testable in observational studies" (p. 99). Pearl (2009, 2010) advocates for a structural equation model basis of causality and has criticized RCM for its adoption of counterfactual reasoning. Despite these criticisms, Little and Rubin defend counterfactual reasoning and believe "the quality of the assumptions, not their existence, is the issue" (2000, p.123). Essentially, they advocate for the acceptance of causal claims when the conditions to which they are arrived at are strong, strengthening their validity.

Design Choice and Causal Inference

While both SUVTA and the ignorable treatment assignment assumptions must be met, how these assumptions are met is not prescriptive. Therefore, causal claims are possible with varied design choices because it is not the nature of causation that changes but, rather, the amount of control over the phenomenon understudy (Holland, 1986, p. 954). While causal local institution are possible under varied design choices, the clearest and simplest pathway is randomization (Fisher, 1925; Holland, 1986, p.946, Little & Rubin, 2000, p.127).

Randomized Experiments

Randomized experiments involve the assignment of units to treatment by a process known as random assignment (Shadish, Campbell, & Cook, 2002, p.12). It is this assignment strategy that makes the design so powerful; random assignment offers the strongest support for the assumption of ignorable treatment assignment because it ensures that the potential outcomes

(Y^0, Y^1) are independent of treatment assignment Z, that is $(Y^0, Y^1) \perp Z.$ Random assignment, if employed properly, has the benefit of balancing the observed and unobserved covariates between groups, making any differences arbitrary (Rubin, 1974, p. 694). Achieving balance means that the groups are equivalent in expectation. Therefore, the groups (treatment and control) are balanced across both observed and unobserved covariates.

As early as 1971, when the President's Commission on Federal Statistics called for increased utilization of randomization in research, there was a premium placed on randomized experiments despite their practical difficulties, and they remain the gold standard (Cochran & Rubin, 1973, p. 417; Guo & Fraser, 2015). Although randomization provides strong evidence to make causal claims, it too can be flawed. Even if perfectly designed and executed, randomized experiments can result in biased estimates of the treatment effect due to drop out and failure to comply with treatment guidelines. Further, randomization is not always possible due to ethical, financial or other practical concerns (Murnane & Willett, 2011; Shadish, Campbell & Cook, 2002). So despite some of the advantages of the design, researchers might choose not to use a randomized study design and opt for a nonrandomized study design also known as observational research.

Observational (nonrandomized) Research

The absence of randomization places a study into the categorization of observational research (Shadish, Campbell & Cook, 2002). Although randomization does not occur, the goal of the research often remains the same, to investigate causal relationships (Shadish, Campbell & Cook, 2002, p.14). Since observational research does not exert the same control as randomized research (e.g., random assignment), differences between groups exist prior to treatment. This difference, known as selection bias, makes it difficult to make causal claims between groups because units choose their treatment condition (Rubin, 1974, p.698). Stated otherwise, the potential outcomes (Y^0, Y^1) are not independent of treatment selection.

In practice, an observational study occurs when random assignment has not been used to assign units to active or control. Consider enrollment in private or public elementary schools. Families choose which type of educational setting to enroll their children. The decision to enroll a child into these differing educational systems can include a complex set of covariates including preference, proximity, finances and parental educational obtainment. This ability to choose the educational setting, public or private, is selection bias. Without random assignment, the best researchers can do is identify and track these variables that are different between the groups, referred to as confounding variables, and attempt to minimize or account for their impact (Cochran & Rubin, 1973, p.418). Comparing the two treatment groups without statistical adjustment leads to a biased estimate of the treatment effect. Therefore, to make a causal claim an unbiased effect of the treatment needs to be achieved and selection bias must be addressed.

Causal Local institution in Observational Research

Although treatment assignment is not independent in observational research, the selection process can be modeled and used to remove the bias resulting from self-selection into treatment or control groups (Murnane & Willett, 2011; Shadish, Campbell & Cook, 2002). The modeling of the selection process is best guided by direct study of the selection phenomenon and supported through a rich set of covariates, $\boldsymbol{X} = (X_1, \ldots, X_p)'$ (Steiner, Cook, Shadish & Clark, 2010). When the selection process is adequately modeled, the potential outcomes are independent of treatment conditional on **X**, $(Y^0, Y^1) \perp Z|\boldsymbol{X}$.

Accounting for the selection bias allows for the difference between groups to be an unbiased estimate of the treatment effect. Consequently, the average treatment effect is then the difference in conditional expectations of the treatment and control group's outcomes. That is,

ATE $\tau = E\{E(Y|Z = 1, \boldsymbol{X})\} - E\{E(Y|Z = 0, \boldsymbol{X})\} = E(Y_i^1) - E(Y_i^0)$

ATT $\tau_T = E\{E(Y^1|Z = 1, \boldsymbol{X})\} - E\{E(Y^0|Z = 1, \boldsymbol{X})\} = E(Y_i^1|Z_i = 1) - E(Y_i^0|Z_i = 1)$

ATU $\tau_U = E\{E(Y^1|Z = 0, \boldsymbol{X})\} - E\{E(Y^0|Z = 0, \boldsymbol{X})\} = E(Y_i^1|Z_i = 0) - E(Y_i^0|Z_i = 0)$

In theory, once the selection bias has been accounted for and the treatment selection has been determined ignorable, the difference between treatment and control groups now represents an unbiased estimate of the treatment effect. This is a much more complex process as there are no statistical tests to determine if the selection bias has been sufficiently addressed (Guo & Fraser, 2015). In fact, research has demonstrated that misspecified models of the selection process can increase the bias (Leon & Hedeker, 2007). Therefore, modeling of the selection process warrants careful attention.

While making causal claims with observational research is possible, not all researchers choose to go down this path; some elect to simply acknowledge the limitations of the research, explicitly stating that causal claims cannot be made. When researchers are interested in causal relationships, there are two main methods for its study: alternative design features and applied statistical analysis (Murnane & Willett, 2011, Shadish, Cook & Campbell, 2002).

Alternative Designs

The study of causal relationships can occur in observational research when alternative designs are used, specifically the regression-discontinuity approach and instrumental variables estimation. Regression discontinuity exploits the selection process to provide unbiased causal estimates (Shadish, Campbell & Cook, 2002) while instrumental variables estimation exploits a covariate, referred to as the instrument, to provide an asymptotically unbiased estimate (Murnane & Willett, 2011). Both methods allow for causal inference in observational research.

Consider a reading intervention that uses a cut off score to assign students to treatment or control. Since students are assigned to rather than selecting into groups, the assignment mechanism, the cut score, is fully known and a regression discontinuity approach can be used. A shift of the mean or slope of the line at the cut off score, the assignment mechanism, indicates that there is a treatment effect (Shadish, Campbell & Cook, 2002). Although this type of design does not provide information about the full sample of students, it does provide causal evidence for the impact of the treatment for students around the cut off score. Whether using regression discontinuity or instrument variables estimation, a limitation is that little is known about the full range of outcomes. With instrument variables estimation, knowledge is limited to that accounted for by the instrument, and with regression discontinuity, it is limited to those around the cut off score (Murnane & Willettt, 2011; Shadish, Campbell & Cook, 2002).

Applied Statistical Analysis

In absence of being able to use experimental or alternative designs, the next route to studying causal relationships is through applied statistical analysis. This method rests on the assumptions stipulated by Rubin (1974) in making causal claims in observational research: both SUTVA and strong ignorable treatment assignment must be achieved. Therefore, causal claims based on applied statistical analysis rely heavily on appropriate covariate selection. This process should be grounded in theory and strong knowledge of the selection process to ensure that the covariates adequately model the selection process (Murnane & Willettt, 2011; Steiner, Cook, Shadish & Clark, 2010). Murnane and Willettt (2011) advise that methods are not “magic” and warn that the subsequent methods applied are only as good as the covariates used to model the selection process (p.288). Failure to adequately model the selection process ensures the failure of any subsequent method.

Controlling for Covariates. One way to account for selection bias is to use statistical methods that control for covariates (e.g., regression, analysis of covariance). Regression is the most common statistical technique for controlling for covariates (Murnane & Willett, 2011). Multiple linear regression estimates treatment effects by regressing the outcome on the covariates. Relevant covariates and an indicator for treatment as well as any interactions between the treatment variable and each of the covariates are regressed on the outcome.

While regression is frequently employed in the literature, it is insufficient for meeting the criteria for making causal claims. Although controlling for covariates can create balanced groups across an observed set of covariates, the groups remain unequal in expectation due to hidden bias. This hidden bias results from achieving balance across only observed covariates meaning that systematic difference between groups on unmeasured covariates might remain.

Statistical methods that control for covariates are unlike experimental designs because the outcome and selection bias are addressed simultaneously. With randomized designs, equivalent groups are created by design at the outset of treatment. Therefore, the potential outcomes are independent of the selection modeling. Since this does not occur with post hoc adjustment, making causal local institution are not possible because the assumption of a strongly ignorable treatment assignment has not been met.

Creating Equivalent Groups. Another strategy for accounting for selection bias involves the use of statistical procedures to minimize its impact by creating equivalent groups prior to analysis. When this strategy is properly employed, the potential outcomes are independent of treatment conditional on a set of covariates ($\mathbf{X}$, $(Y^0, Y^1) \perp Z|\boldsymbol{X}$). There are different strategies for doing this including stratification and multivariate matching.

One way to reduce selection bias is to stratify on one or many covariates. Stratification takes a covariate or set of covariates and subdivides the sample on them (Murnane & Willettt, 2011). These strata are then used for the analysis to help minimize bias. This strategy works well with one or two covariates but becomes impossible due to data sparseness and lack of common support with increasing numbers of covariates (Murnane & Willettt, 2011).

Multivariate matching is most commonly used when examining the ATT (Guo & Fraser, 2015). In this case, multivariate matching attempts to resolve the missing data issue by matching each unit in the treatment group to at least one unit in the control group that is identical or near identical on observed covariates. If the ATE were of interest, a similar process would need to occur for matching each unit in the control group to at least one unit in the treatment group. Since finding an identical matched pair is difficult, matching involves a series of decisions related to distance, strategy and selected algorithm (Guo & Fraser, 2015).

Both multivariate matching and stratification offer a way to create groups that are equivalent in expectation allowing for causal local institution, but the complexity of data makes the approach impossible to use. Even with as little as ten covariates the possible combinations exceed one million (Guo & Fraser, 2015). This obstacle is why propensity score techniques are desirable and why they continue to grow in popularity (Thoemmes & Kim, 2011).

The Propensity Score

Propensity score (PS) techniques have an advantage over multivariate matching as the propensity score is a single, balancing score derived from all of the observed covariates **X**. The propensity score can be estimated using various statistical procedures that provide a probability, including regression, discriminant analysis and decision tree (Guo & Fraser, 2015). The propensity score is the probability of a unit receiving a treatment conditional on a set of covariates, $e(\boldsymbol{X}) = P(Z = \boldsymbol{X})$ (Rosenbaum & Rubin, 1983, p. 42). If the treatment assignment is strongly ignorable given the propensity score $e(\boldsymbol{X})$, then the potential outcomes are independent of treatment assignment given the propensity score, $(Y^0, Y^1) \perp Z|e(\boldsymbol{X})$.

Additionally, the propensity score is a balancing score with the joint distribution being equivalent in both the treatment and control groups, $P(\boldsymbol{X}|Z = 1) = P(\boldsymbol{X}|Z = 0)$. While balance is automatically achieved in randomized experiments, balance needs to be created in observation studies. For the propensity score to be balanced, a variety of statistical procedures can be applied including but not limited to matching and stratification (Guo & Fraser, 2015). Since there are many ways to arrive and use a propensity score, both the estimation process and the various methods are detailed.

Covariate Selection

Appropriate covariate selection is essential for ensuring that the treatment assignment is independent; ultimately, satisfying the assumption of an ignorable treatment selection. In theory, all variables related to the selection process and outcomes need to be included but, in practice, there is no statistical test to ensure that this has been accomplished (Luellen, Shadish, & Clark, 2005). Therefore, it is the responsibility of the researcher to ensure that an adequate model of the selection process has been developed.

The selection of covariates is best guided by empirical study of the selection process and theory as well as a comprehensive set of covariates (Luellen, Shadish & Clark, 2005; Murnane & Willettt, 2011). Rosenbaum (2002) advocates for the inclusion of important covariates even if they do not reach the level of statistical significance between groups. Therefore, if a covariate is related to the selection process and/or outcomes, it should be retained even if the *p* value falls below the specified threshold of statistical significance. Although there is no way to assure that hidden bias has been eliminated, sensitivity analyses can be done to bolster support.

Estimating the Propensity Score

Estimating the propensity score is most commonly completed using binomial regression models (Luellen, Shadish & Clark, 2005). Binomial regression models are used for discrete choice outcomes (i.e., treatment participation, yes or no) and model the probability that the binary response is a function of a set of predictors. Unlike, the traditional use of regression that models the outcome of interest, propensity score methods use regression to model the selection process. Although logistic regression is most often employed, it assumes linearity between the independent variables and the log odds. Due to this requirement, alternative approaches have been explored (Luellen, Shadish & Clark, 2005).

To accommodate for the complex relationship between the selection process and covariates, statistical learning algorithms, such as random forest, regression trees or boosting, have been adopted (Westreich, Lessler & Funk, 2010). These statistical learning algorithms have advantages over traditional regression approaches because they are an automatic, nonparametric procedure for addressing complex interactions and nonlinear relationships. Although they are better able to accommodate complex data, they have the tendency to lack fit when applied to new data (Luellen, Shadish & Clark, 2005).

Regardless of the chosen estimation process, significant overlap between the propensity scores for the treatment and control group must exist. This area of overlap is referred to as the region of common support. When the distribution of the propensity scores is similar between groups, then all levels of the propensity score can be included (Guo & Fraser, 2015). When the distribution is dissimilar, propensity scores that fall outside the region of common support are dropped from subsequent analyses, a process often referred to as trimming. Sufficient overlap between the distribution of the propensity scores for the treatment and control groups must exist to continue with the analysis. If there is insufficient overlap, then the selection model might be misspecified and a re-estimation of the propensity score might yield different results. Overlap between the distributions of the propensity score must occur before moving to conditioning of the propensity score.

Conditioning the Propensity Score

Following the estimation of the propensity score, different conditioning methods can be applied. Conditioning methods aim to achieve balance between the treatment and control groups. There are different conditioning strategies that can be employed but these strategies influence the analysis of the outcome. For instance, matching (i.e., 1:1 and 1: many) and weighting by odds are commonly used when estimating the average treatment effect on the treated (ATT) (Austin, 2011). Full matching, stratification, inverse probability, propensity score weighting, ANCOVA and ANCOVA, including the propensity score as a covariate, are used when estimating the average treatment effect (ATE) (Harder et al., 2010; Stuart, 2010; Steiner et al. 2010). Ultimately, the conditioning method chosen is important and it influences subsequent analyses.

Matching. Matching is one method for conditioning the propensity score. Matching, in essence, is the pairing of similar units; units with a similar propensity score would be paired together. The unit in the control group would serve as the potential outcome had the unit in the treatment group not received the treatment.

Most commonly, 1:1 matching is used. With one to one matching, a single treatment unit is paired with a single control unit. One-to-many matching is also employed; with this approach, a unit in the treatment group is matched to a specified number of control units. The equation below demonstrated a basic matching strategy:

$$|p_i - p_j| = \min\{|p_i - p_k|\}$$

Depending on the nature of the data, one matching strategy might be preferred to another. For instance, one-to-many is beneficial when there are a large number of control units, and the potential data loss is substantial. Consider the case where there were 100 units in the treatment group and 300 in the control. Despite the matching strategy, the maximum number of matches would be 100. With one-to-one matching, there would be substantial data loss since 200 control units would be dropped from the analysis. One-to-many matching has the ability to curtail this data loss by matching more control units to the treatment unit.

While the matching strategy is an important consideration, the distance between matches is a critical consideration. Distance (δ) is a measurement of similarity between units on a given covariate, and this information is utilized within a matching strategy. The equation below shows a matching strategy that accounts for distance.

$$\delta > 0\ |p_i - p_j| = \min\{|p_i - p_k|\}$$

Without setting this distance, also known as a caliper width, there is a potential for dissimilar units to be matched. Although caliper widths help to place some assurances around matching, it can cause a reduction in matching.

In addition to these strategies, matching can also be done with or without replacement. Matching without replacement occurs, as discussed above, with one unit being matched to one treatment. When matching occurs without replacement, a control unit cannot be used again even if it matches well to more than one treatment unit. Therefore, matching with replacement can help increase balance between groups by allowing the same control unit to be matched to multiple treatment units. The downside to matching with replacement is that it again causes a loss in data. This loss in data is important because the conclusions might be less generalizable.

Finally, the algorithm for matching needs to be determined. When matching with replacement, nearest matched to its nearest neighbor or set of nearest neighbors in the control group. When matching occurs without replacement, greedy or optimal matching can be used. Greedy matching is similar to nearest neighbor except once cases are matched; they are dropped from the dataset. Due to this ‘first come’ strategy, some matches are not ideal because the overall distance is not minimized. To circumvent these issues, optimal matching can be used. Optimal matching ensures better overall matching by minimizing the global distance (Guo & Fraser, 2015). This means that some treated units are matched with their second, third or other best control units.

Propensity score matching is similar to matching using multivariate methods, insomuch that propensity score matching can be done with variable distances using calipers, different matching methods (e.g., 1:1 or 1: Many) and using various algorithms for nearest neighbor, optimal or greedy. The difference between the matching methods is that rather than using the entire set of covariates **X**, propensity score matching can use just the propensity score or the propensity and a subset of key covariates.

Stratification. Alternatively, propensity score stratification can be employed which uses the estimated propensity score $\hat{e}(\mathbf{X})$ to divide the observations into distinct strata. Within each stratum, the units are homogenous; thus, the aim is to divide observations into groups with the same covariate distribution (Austin, 2011). Cochran (1968) demonstrated that 90% of overt bias is removed from a confounding variable when using 5 equal-size strata. This finding extends to the application of propensity score methods; Rosenbaum and Rubin (1984) additionally demonstrated that 90% of bias could be removed. Austin (2011) conceptualizes this strategy as 5 distinct quasi-randomized experiments. Treatment effects can be considered within a stratum or across strata. Typically, stratum-specific estimates of treatment effects are poled across stratum to estimate an overall treatment effect (Rosenbaum & Rubin, 1984).

Weighting. Another method, first introduced by Rosenbaum (1987), propensity score inverse-propensity weighting is used to achieve balance. Unlike matching and stratification, it does not aim to create equivalent groups. Rather, weighting achieves balance by taking a portion of a unit's information based on that unit's likelihood of receiving treatment. Formally stated, the weights are defined as:

$$w_i = \frac{Z_i}{e_i} + \frac{(1-Z_i)}{1-e_i}.$$

The main benefits to weighting are that all of the data can be retained, and it does not require a continuous or normally distributed outcome variable (Guo & Fraser, 2015).

Covariance adjustment. Unlike the previous strategies for conditioning the propensity score, an alternative method is to use the propensity score as a covariate and adjust for its impact. Similar to weighting, covariance adjustment does not attempt to create equivalent groups. Instead, covariance adjustment is a strategy that regresses the outcome variable on the estimated propensity score and treatment indicator (Austin, 2011). Conducting an analysis of covariance (ANCOVA) is the simplest way to use this method. Although this method is simple to use, Rosenbaum and Rubin (1984) advocated for the use of matching and stratification rather than weighting or covariance adjustment.

Assessing the Treatment Effect

Once the propensity score has been conditioned, multivariate analyses can be carried out to examine the treatment effect, but the procedure for this is dependent on the conditioning strategy that has been employed and the level of the model needed. For instance, with greedy matching, multivariate analyses can proceed as they do in experimental designs, but this is not true with optimal matching. For optimal matching, a regression adjustment must be applied when examining the treatment effect (Guo & Fraser, 2015). Additionally, depending on the nature of the data, multilevel model might be warranted.

Evaluating Accuracy of the Propensity Score

The overall aim of using propensity scores is to eliminate the selection bias inherent in observational research to arrive at an unbiased estimate of the treatment effect (Guo & Fraser, 2015). Although there is no test that can definitively affirm that a selection process has been adequately modeled, sensitivity analyses must be carried out. A sensitivity analysis provides information about the robustness of the treatment outcome - asking specifically what the nature of the unobserved covariate would have to be to change the outcome of the study (Rosenbaum, 2005, p. 1809). Based on the results of the sensitivity analysis, the treatment effect might be insensitive or sensitive to small or large biases (Rosenbaum, 2005).

Considerations for the Application of Propensity Score

Although it is appealing to move from correlation to causation, it takes more than the technical skills required to perform propensity score techniques for this to be achieved. Using propensity score techniques to discern causation is predicated on having a selection process that is strongly ignorable. Steiner and Cook (2013) identify three requirements for a strongly ignorable selection process: 1) valid measurement of constructs correlated to both treatment and potential outcomes; 2) latent constructs involved in the selection process and potential outcomes must be measured in addition to covariates to remove all bias; and 3) a region of common support must exist between the treatment and control group. Since its utility is predicated on moving the non-ignorable treatment selection to strongly ignorable, covariate selection is the most critical issue.

Covariate Concerns

As Thoemmes and Kim (2011) stated, “a propensity score analysis can only be as good as the covariates that are at the disposal of the researcher” (p.93). To establish an ignorable selection process, a rich set of covariates must be available to the researcher. Steiner & Cook (2013) recommend an investigation of the selection process through a planning study while Steiner, Cook, Shadish and Clark (2010) suggest covering a wide array of variables covering different factors. Since, in practice, the dataset might be fixed gathering additional variables might be impossible. Early research has identified two critical variables for reducing bias: pretest measures and variables related to treatment assignment. Steiner and Cook (2013) warn that when using secondary data where all the necessary variables are not available, causal claims should not be made.

In addition to having a robust set of covariates, each of the covariates needs to be reliably measured. As reliability decreases, bias has the potential to increase (Steiner & Cook, 2013). Often, observed covariates are unable to explain the selection process. Theory needs to guide the process to help assist understanding of the selection mechanism and identify latent constructs that might be involved.

Estimation Methods

Logistic regression is the most common estimation method for propensity score analysis. Following the work of Dehejia and Wahba (1999), the goal of estimation should be to balance the covariates thus supporting independence of treatment. If balance is not achieved, higher-order terms and interactions should be added and the modeled retested until balance is achieved. Although alternative approaches to logistic regression (e.g., tree-based methods, boosted regression models and neural networks) are feasible, research is limited (McCaffrey, Ridgeway & Morral, 2004; Westereich, Lessler & Frank, 2010; Watkins et al., 2013). While some research has demonstrated superiority for tree based regression methods (Watkins et al., 2013), other research has demonstrated more mixed outcomes (Westereich, Lessler & Frank, 2010). More research needs to be done to determine if these alternative methods outperform logistic regression.

Conditioning Methods

Research has examined the impact of different propensity score schemes on matching rates, balance and treatment effects among other aspects of analysis. Overall, the research is mixed with no clear indication of a single best approach to conditioning the propensity score. Research has generally demonstrated that matching is a better strategy than stratification (Austin, 2007; Austin, 2014), which is likely why matching is the most common approach for conditioning.

Although matching is the preferred conditioning method (e.g., Ali, 2015), there is less evidence about which type of matching is best – although, nearest neighbor matching is most common. In a test of 12 different matching schemes, both nearest neighbor and optimal matching achieved the same level of balance across covariates (Austin, 2014). Additionally, adding calipers to nearest neighbor matching improved mean squared error, but it does sacrifice sample size in comparison to optimal matching. Further, when examining the impact of the sub-algorithms used in nearest neighbor (i.e., low to high, high to low, closest distance, random), the results were generally inconsistent, not favoring any of the methods. Despite this, selecting matches ordered from high to low led to the most bias consistently (Austin 2014).

Accuracy of Propensity Score Methods

Although no direct test exists for the reduction of bias, Monte Carlo studies have demonstrated that there is not a clear ‘winner’ when it comes to propensity score conditioning methods (e.g., Zhao, 2004; Guo & Fraser, 2015). For instance, when Guo & Fraser (2015) tested seven different conditioning strategies in two settings using Monte Carlo simulation, their results revealed that best conditioning method varied by the setting. Due to this, they advise for the use of sensitivity analysis to help gauge how robust the conclusions are from confounds.

Another strategy for determining the accuracy of propensity score methods uses within study comparisons. Within study comparisons is an approach using a single study question but alters the design so that some participants are assigned randomly and others get to choose treatment condition. The goal is to compare the results of the observational study to the results of the randomized study. This line of research, within study comparisons, has demonstrated that bias elimination is possible when there is extensive knowledge of the selection process or when the comparison groups are like the treatment group on pretest measures of the outcome (Steiner, Cook, Shadish & Clark, 2010, p. 251; Shadish, Clark & Steiner, 2008). These studies have also demonstrated that covariate selection is more important than the propensity score method employed (Shadish, Clark and Steiner, 2008; Steiner, Cook, Shadish & Clark, 2008).

Effectiveness of the Propensity Score Model

Although propensity score methods hold much promise and have grown in popularity, research regarding their superiority has been mixed (Peikes, Moreno & Orzol, 2008; Shah, Laupacis, Hux, and Austin, 2005; Stürmer et al. 2006). Meta-analyses in the medical field have not found many cases in which the propensity score method is superior to other methods (e.g., regression, ANCOVA) for accounting for differences between groups (Shah, Laupacis, Hux, and Austin, 2005; Stürmer et al. 2006). Further using four-arm within-study comparisons, Shadish, Clark and Steiner (2008) and Pohl et al. (2009) found similarity in bias reduction using both propensity score methods and analysis of covariance (ANCOVA). Although this research has demonstrated a general parity of performance, Peikes, Moreno and Orzol (2008) found that using propensity score contradicted the conclusions of the experimental design. Although this would seemingly deter from the use of propensity score methods, the lack of superiority might be due to the newness of this technique. These inconsistent results might be a result of the misapplication of propensity score methods (Austin, 2008; Cook, Shadish, Wong, 2008; Luellen, 2007).

PSM in Higher Education

Although the superiority of propensity score methods has not been definitively demonstrated, there are other reasons that researchers might choose propensity score methods over traditional regression (Peikes, Moreno & Orzol, 2008). Propensity score methods are particularly appealing for contexts in which randomized research is not feasible or desired, which is a common constraint in higher education. Since much of the research in higher education continues to be observational, it is not surprising that the use of propensity score methods continues to grow despite these mixed results.

The use of propensity score methods in higher education can be organized into two major approaches: single institution and multi-institutional. Research using a single institution focuses on a question or problem encountered at a single institution. The analysis and subsequent findings are local to students at that institution and are not generalizable to students at other institutions. Most often, this type of research adopts a single-level model but multilevel models have been applied (e.g. Vaughan, Lalonde & Jenkins-Guarnierie, 2014). When using multi-institutional data sets, multilevel models are more common. This type of model is better able to account for the dependences between students from similar types of institutions. For instance, students that attend large, urban, public institution might share more similarities with one another than with students that attend small, rural, Catholic institutions.

Multi-institutional Research

Most research on propensity score methods uses large national datasets. These large datasets are appealing when studying propensity score methods because data are collected on various individuals from many institutions allowing both the study of long-term effects of behaviors on success in higher education as well as greater generalizability. These large data sets allow researchers to explore questions like the utility of summer bridge programs (Douglas & Attewell, 2014), academic matching between students' achievements and institutions' selectivity (Heil, Reisel and Attewell, 2004), the impact of community college on degree attainment (Melguizo, Kienzl &Alfonso, 2011). Pairing these datasets with propensity score methods, further allows researchers the potential to move their findings from correlation to causation.

Although the ability to make causal claims exists, most research in this area does not do enough to satisfy the necessary claims. Since there is no test to ensure that selection bias has been successfully removed, there must be strong support that this has been accomplished both through methodology and appropriate statistical analysis. For instance, from a statistical standpoint, it is likely that there are dependencies based on the institutions in which students are nested. Often the multilevel structure of this data is not taken into consideration and single-level models are applied (e.g., Douglas & Attewell, 2014; Doyle, 2011; Melguizo, Kienzl &Alfonso, 2011). Whether the single-level model fits better remain unexamined making the subsequent claims tenuous.

Additionally, the critical decisions points are not explicated, making it hard to support claims that the selection bias has been removed. Although most research uses matching for conditioning the propensity score, the details of their specific approach are left unexplained. For instance, neither Heil et al. (2014) nor Melguizo et al. (2011) fully explained their matching method. It is difficult to discern if researchers are using one-to-one, with or without replacement or applying calipers when it is referred to generically as ‘matching’. Although Douglas & Attewell (2014) identified the type of matching, it was unclear how the optimal matching strategy (i.e. matching 3 control cases to each 1 treatment case within a .25 caliper width) impacted the overall sample size and the conclusions that were subsequently drawn.

In addition to this issue, there also has been a lack of attention on the impact of variable selection when estimating the propensity score. The removal of selection bias hinges on this model and although the researcher might state that there is no difference between groups after the conditioning strategy has been applied, this balance is solely achieved through these observed covariates. Since propensity score methods do not have the benefit of balancing both observed and unobserved covariates like experimental approaches, the conclusions are only as strong as the covariates included. None of the studies addressed whether they had a comprehensive set of variables necessary for investigation of their research question. For instance, Douglas and Attewell (2014) focused on a small set of academic and demographic variables and did not incorporate any noncognitive variables into their model. In addition, sensitivity analyses were not conducted to bolster the support of the causal claims.

Single Institution Research

Unlike multi-institutional research, single institution research attempts to resolve local issues. Although this method reduces generalizability, it does often benefit from additional knowledge or access to knowledge about the research process. For instance, consider the same researcher using a national dataset and a local dataset with similar covariates. When the researcher uses the local dataset, more information is known about potential covariates. This proximity to the data can help illuminate issues about the selection process and help to resolve or provide context to any data irregularities.

Although much of this research is applied in nature, there has been some studies that have specifically examined the utility of propensity score methods in higher education. For instance, Clark and Cundiff (2011) examined the impact of a first year course on academic performance and persistence using propensity score methods. The propensity score was estimated using a single-level model and conditioned using stratification with five strata and matching. The two conditioning methods led to different overall conclusions regarding the impact of the course with conditioning using stratification finding no difference and matching demonstrating the opposite.

Although there is reliance on single-level models with single institution research, multilevel modeling has been used. For instance, Vaughan, Lalonde & Jenkins-Guarnierie (2014), used hierarchial linear modeling (HLM) to examine a first year seminar course aimed at improving the academic achievement and persistence of first year students. Since students were assigned to the first year courses based on academic major, an HLM approach was warranted. Vaughan et al. (2014) argue for the utility of HLM propensity score methods because of the insufficient matching that resulted with the use of a single-level model.

Although there are benefits to single institution research, when using propensity score methods, this line of research is similarly plagued by a lack of essential details provided throughout the analysis. For instance, Clark and Cundiff (2011) do not provide information on the subsequent sample size with each matching procedure nor specifics on which treatment effect was assessed.

Overall Aim of Research

Despite the rapid growth of propensity score methods in education, there have been relatively few studies focused on issues within higher education. Those studies that have occurred typically adopt a single-level model (e.g., Clark & Cundiff, 2011; Dehejia & Wahba, 1999) use logistic regression for estimation (e.g., Clark & Cundiff, 2011; Schafer, Wilkinson & Ferraro, 2013; Vaughan, Lalonde & Jenkins-Guarnierie, 2014), and condition the propensity score using matching or stratification (e.g., Clark & Cundiff, 2011; Schafer, Wilkinson & Ferraro, 2013; Vaughan, Lalonde & Jenkins-Guarnierie, 2014). Although there has been some attempt to broaden the application of propensity score methods to hierarchical relationships in this context, studies using multilevel modeling are far fewer (e.g., Vaughan, Lalonde & Jenkins-Guarnierie, 2014; Heil, Reisel & Attewell, 2014). Further, most of the research in this area has focused on a specific research question rather than on the method itself. Although information about the use and utility of propensity score methods exists, based on research using simulated data or multiple arm studies with randomized research as one of those arms, there remains a lack of knowledge about what works within the context. Additionally, there is only limited information about how the availability of covariates influences the results.

Additionally, many important details have been left out of propensity score research in higher education literature. Although this is a problem within the field, it is a notable issue outside the field as well. Overall, there is a lack of consensus on what aspects of the analysis should be reported (Ali et al., 2015; Thoemmes & Kim, 2011). Specifically, Ali et al. (2015) found in their review of medical literature, only 34.4% of articles explicitly reported variable selection process and the only 59.8% checked and reported covariate balance. Additionally, when examining balance, p-values were much more likely to be reported than the standardized mean difference (70.6% vs. 25.4%). Combined, this makes replication difficult as key aspects from the analysis are missing and inferior methods are being used. Further challenges exist when the method is moved from a strictly theoretical framework to an applied setting. This research aims to add to the literature within applied educational research.

CHAPTER THREE

METHODOLOGY

This chapter outlines the methodology including an overview of the study, research questions, design, sample characteristics, analytical procedures and outcome measures.

Study Overview

This study used existing institutional data from a large, urban, public, very high research university to compare sixteen matching schemes, built from three separate datasets, to estimate the propensity score, achieve balance between groups and test the sensitivity of the average treatment effect (ATE). For each propensity score (PS) model, four different conditioning strategies were applied. The first four matching schemes used commonly collected data available within a student information system (referred to as SIS dataset). The next four matching schemes combined the SIS dataset with data from an entering student survey (referred to as ESS dataset). The next four matching schemes, again, combined the SIS dataset with data gathered from a noncognitive survey (referred to as NCS dataset). The final four matching schemes included data from the SIS, ESS and the NCS datasets. Each model builds upon the next, offering additional covariates for the model building process.

For the conditioning methods, two matching algorithms were used. Three of the matching strategies used a greedy algorithm developed by Bergstralh and Kosanke (1995) and one matching strategy used a digit matching approach developed by Parsons (2000). For the matching strategies using the greedy algorithm, 3 caliper widths were applied (no caliper applied, 0.25 caliper width, .1 caliper width). The four PS models were conditioned by the four matching strategies, resulting in 16 matching schemes that were assessed on sample size, balance, average treatment effect and sensitivity.

To assess the effectiveness of these propensity score techniques in an applied educational research setting, the methodological research questions are nested within the framework of an overarching contextual research question. This guiding research question aimed to understand the extent to which first-time, fulltime students who enrolled in optimal credit levels (defined as 15 or more credit hours during the first term of attendance) experienced greater levels of success. Student success is a complex phenomenon and includes multiple and sometimes competing constructs. This research uses first-year retention as a proxy for student success. Students are considered retained if they were enrolled at the university, the following fall term. Many researchers have studied retention resulting in various models with a diverse set of covariates (e.g., Pascarella & Terenzini, 2005; Ting, 1998; Tinto, 1975; 1993). Although this research does not seek to understand the complexity of student success, it does try to understand the influence the availability of additional covariates has on propensity score techniques and their influence on the stability of findings in applied educational research.

Research Questions

1. To what extent do the treatment and the control groups vary naively across covariates?
2. To what extent do different PS models achieve overlap between the treatment and control groups?
3. To what extent do different PS models and conditioning strategies impact the sample size?

4. To what extent do different PS models and conditioning strategies achieve balance between groups?
5. To what extent do different PS models and conditioning strategies reach the same overall conclusions?
6. To what extent is the average treatment effect robust against unobserved covariates under different PS models and conditioning strategies?

Design

A “four by four” design was employed. Specifically, four PS models (i.e., SIS, SIS + ESS, SIS+NCS, SIS + ESS + NCS) and four matching strategies (greedy – no caliper, greedy – 0.25 caliper width, greedy – 0.1 caliper width, greedy 5→1) were applied to the data. Overall, 16 propensity score matching schemes were examined.

1) SIS, greedy, no caliper
2) SIS, greedy, .25 caliper
3) SIS, greedy, .1 caliper
4) SIS, greedy 5→1, no caliper
5) SIS + ESS, greedy, no caliper
6) SIS + ESS, greedy, .25 caliper
7) SIS + ESS, greedy, .1 caliper
8) SIS + ESS, greedy 5→1, no caliper
9) SIS + NCS, greedy, no caliper
10) SIS + NCS, greedy, .25 caliper
11) SIS + NCS, greedy, .1 caliper

12) SIS + NCS, greedy 5→1, no caliper

13) SIS + ESS + NCS, greedy, no caliper

14) SIS + ESS + NCS, greedy, .25 caliper

15) SIS + ESS + NCS, greedy, .1 caliper

16) SIS + ESS + NCS, greedy 5→1, no caliper

Data Collection

Data were collected as part of the university's routine processes and shared with the researcher as a de-identified data file. Three primary sources of data were used for this research: student information system, an entering student survey and a noncognitive survey.

Student Information System (SIS). Routine data are collected on prospective, enrolled and graduate students within a student information system. These data can be expansive or limited depending on the practices of the particular institution. Standardly, universities maintain data on information that they need to report back to federal or state agencies or other organizations. These data are often collected through students' applications, admissions, enrollment, registration, course grades and financial aid. The data made available for this research study are listed in Appendix A. The data include basic demographic information, high school academic information, placement test results, academic college and financial need.

Survey Datasets. In addition to the host of institutional variables routinely collected as part of an institution's SIS, there are often university-approved additional data collection efforts. These data efforts typically aim to supplement the information available in the SIS to enhance the institution's understanding of issues relating to student success, satisfaction and engagement.

Often, students are asked to complete surveys such as: entering and exiting student surveys, personality and/or behavioral assessments, student engagement surveys, student satisfaction surveys, noncognitive surveys, and placement surveys, among others. Often these data do not reside within the institution's SIS but can be combined with these data to more fully understand aggregate student behaviors, patterns and performance as they relate to issues of policy, program review or other areas of substantial educational interest. For this particular institution, an entering g student survey and a noncognitive survey were administered to first-time students.

Entering Student Survey Dataset. In addition to the SIS data, data were provided from an entering student survey to create the ESS dataset (see Appendix B). The entering student survey was administered to students who had not yet matriculated into the university but intended to enroll. The survey provided information related to students' reasons for attending college, reasons for selecting the particular institution, students' self-perceptions and educational plans, as well as information on how students spent their time.

Noncognitive Survey Dataset. Data were also provided from a noncognitive survey to create a NCS dataset (see Appendix C). The noncognitive survey was administered to matriculated first year students. The noncognitive survey collects information across 12 domains. The scales on the survey measure family obligation, self-regulated learning, perceived efficacy of instructor, perceived self-efficacy, perceived sense of belonging, time management, academic motivation, academic control striving behavior, academic dishonesty, grit/perseverance, caring, subjective well-being and feeling lost in the system.

Research Population

This study investigated the entering fall 2014 first-time, fulltime student cohort. at a large, public, very high research postsecondary institution. Overall, there were 3,007 student cases that met these criteria. This particular student population was chosen because it had the most robust survey participation and because the performance of this cohort (first-time, fulltime students) is of national interest. The contextual research question was derived from recent work from Complete College America. States that have adopted 15 credit hours as fulltime have demonstrated gains in retention and completion (CCA, 2014). At this time, this institution and the state for this particular study define fulltime at 12 and had not begun any statewide initiatives to move this metric from 12 to 15 credit hours. In this analysis, optimal credit enrollment was defined as registering in 15 or credit hours in the first term.

Variables

Three distinct sets of variables were used to build the selection models. For the SIS dataset, covariates included race, gender, age, placement results (writing and math), ACT Composite Score, ACT Math Score, ACT English Score, ACT Reading Score, unweighted high school grade point average, high school type, advanced placement credits, academic college, honors college, Pell recipient and first generation. Covariate descriptions can be found in Appendix A.

For the ESS and NCS datasets, each of the items from the surveys were eligible to be entered into the model (see Appendix B & Appendix C). The noncognitive student survey had 12 scale scores and scale scores were prioritized over individual items. The entering student survey was not designed with scale scores so items were only eligible to be entered as individual covariates.

Analytic Procedures

Sixteen propensity score matching schemes were assessed to determine the impact of the scheme on sample size, balance, average treatment effect and sensitivity. Each step of the analytic procedures was aligned to the study's hypotheses and are detailed below.

Step one: Determine the difference between groups

As a precursor to the first research question, a chi-square test was conducted to determine if there was a difference on retention between students that enrolled in optimal credit hours and students that did not enroll in optimal credit hours. If there was no difference between groups at the outset, the analysis would not have continued. Knowing that the groups did differ on the outcome of interest, the next step was to determine whether or not students in these two groups demonstrated differences across the three covariate sets. Prior to running the logistic regression to discern differences between groups ($p < .05$)., descriptive statistics for each covariate were examined (i.e., N and the distribution of the covariate overall and between groups). Covariates eligible for entry were assessed for their relationship to the outcome of interest, multicollinearity and small cell sizes. It was anticipated that there would be significant differences between the treatment and control groups on key covariates.

Step two: Estimate the propensity score

Four single-level logistic regression models (SIS, SIS+ESS, SIS+NCS, SIS+ESS+NCS) were derived to estimate the propensity score. Since the propensity score aims to satisfy the strongly ignorable treatment assignment assumption, predictors associated with the assignment should be controlled; stated differently, selection bias needs to be removed. Therefore, bivariate correlations were run to examine the relationship between the treatment and the predictors.

Since parsimony is not a goal when estimating the propensity score, all variables with small relationships, even when not statistically significant, were retained. This is consistent with current recommendations (Rosenbaum & Rubin, 1983; Shadish & Steiner, 2010; Steiner, Cook, Shadish & Clark, 2010).

As part of the model building process, covariates were once again checked for multicollinearity and descriptive statistics were assessed. Since each PS model was built separately, it was important to re-inspect the covariates. In addition, if any of the models had demonstrated inadequate fit on Lemeshow Goodness-of-Fit test, interactions and hierarchical relationships would have been examined. This step was not necessary.

Step three: Assess the region of common support

The region of common support was visually inspected for each of the four models following the advice of Lechner (2000). It is preferred to have a wider region of common support because this supports general comparability between the groups and suggests that the treatment assignment is strongly ignorable (Thoemmes & Kim, 2011). Since the goal of propensity score methods is to support causal claims, it is suggested that units that fall outside the region of common support be dropped from the analysis (Shadish & Steiner, 2010; Stuart, 2010). Since the present study adopted matching strategies using caliper widths and $5 \rightarrow 1$ digit matching, restrictions on the proximity of matches already existed. As a result, a conservative trimming approach was applied; only extreme outliers were trimmed, the top 99^{th} percentile and the bottom percentile.

Step four: Propensity Score Conditioning

For each model, four matching strategies were examined. Each matching strategy used 1:1 nearest neighbor without replacement and one of two matching algorithms, greedy or greedy 5→1. For greedy matching, three different caliper widths were applied (no caliper, 0.25 caliper and 0.1 caliper). Propensity score conditioning was done in SAS 9.4 using the %gmatch macro developed by Bergstralh and Kosanke (1995) for greedy matching and greedy 5→1 digit matching developed by Parsons (2000). Although the two are similar, greedy 5→1 digit matching offers more precision as matching is based on closest proximity being matched first. To some extent, greedy 5→1 digit matching functions similarly to optimal matching by factoring in proximity into the matching process but, unlike optimal matching, the match is never reconsidered.

The use of nearest neighbor 1:1 matching without replacement leads to data loss as any unmatched units will be dropped from the analysis. To understand the impact of different conditioning strategies, the number of matched pairs retained will be reported as well as the percentage of matched pairs out of the potential pairs.

Step five: Assessment of balance

Balance was assessed to evaluate the ability of the estimation and conditioning strategies to remove the relationship between the treatment assignment (Z) and each covariate. Both statistical significance and the standardized mean difference (SMD) are often cited in the literature as strategies for assessing balance (Guo & Fraser, 2015). Therefore, both were assessed and reported. For statistical significance, the level was set at 0.5 and for SMD a threshold of 0.15 was applied. Based on the literature, balance is achieved if 10% or less of the covariates are unbalanced (Rubin, 2001; Shadish & Steiner, 2010). Balance is required for estimating the average treatment effect.

Step six: Estimate the ATE

To determine the stability of the outcome under different estimation and matching conditions, the average treatment effect of the treated (ATE) was estimated. Since greedy matching was used to condition the propensity score, analysis best proceeds with an approach that accounted for the paired nature of the data (Austin, 2009). Therefore, to analyze the impact of optimal credit enrollment on first year retention, the difference in the probability of 1-year retention between treatment groups was estimated directly by the difference in proportions between treated and untreated students in the propensity score matched sample. McNemar's test, $p < .05$, was used to assess the statistical significance of the risk difference.

Step seven: Sensitivity analysis of unobserved covariates

The final step of the analysis was assessing sensitivity of the ATE to unobserved covariates. The inclusion of the essential covariates is a key step in estimating the propensity score but this does not ensure that all bias has been removed. Since there is no direct test of the magnitude of selection bias, an additional step after determining the ATE is to assess the extent to which the finding is robust against hidden bias.

Γ is a measure of the degree of departure from a study that is free of hidden bias. To measure Γ, Wilcoxon’s signed rank test will be used. The analysis will demonstrate several possible values of Γ and identify where the local institution might change. A study is sensitive if values of Γ close to 1 could lead to conclusions that are very different from obtained assuming the study is free of hidden bias. A study is insensitive if extreme values of Γ are required to alter the inference (Guo & Fraser, 2015).

Comparison across models

Although there was no formal test to assess the differences across PS models and/or matching schemes, comparative information is provided at the conclusion of step two and step three. For step two, the PS models are compared on sample size, variance explained and significant covariates. For step three, a summary of the visual inspection of the region of common support is provided. At step four, the analysis becomes fully integrated with the analysis focused on the 16 matching schemes. Therefore, each table presented provides the relevant data for comparison.

Chapter Summary

This chapter outlined the methodology for this study and described the purpose, research questions, design, sample, analytical procedures and outcome measures. The goal of this chapter was to outline the specific strategies that were undertaken to help applied researchers understand the impact of propensity score techniques on sample size, achieving balance and establishing robust conclusions.

CHAPTER FOUR

RESULTS

This chapter outlines the results of the analysis. The analysis is presented by steps with each of the propensity score (PS) models presented separately within the steps. The steps align directly to the research questions as posed. Additionally, the code used to do the analysis is similarly organized by steps and presented in Appendix D.

Step zero: Baseline data

Analysis began with the SIS dataset which was derived from information in the student information system. The SIS dataset was reduced from 3030 first-time students to 3007 first-time students who enrolled in 12 or more credit hours during their first academic term. Thus, 99.2 percent of the first-time student population met the minimum criteria for inclusion in this analysis. In total, 72.2 percent of the study's population enrolled in optimal credit hours (defined as 15 or more student credit hours) in the first term of college enrollment. Not accounting for potential differences between the two groups, optimal credit hour enrollment and less credit hour enrollment, a chi-square test of independence demonstrated a significant relationship, $X^2(1) = 31.44$, $p<.0001$, between student credit hours and first year college retention. Students enrolled in optimal credit hours were more likely to retain at the university. Based on this finding, subsequent analyses were carried out to determine if the overall finding remained significant after accounting for differences between groups using propensity score method

Step one: Determine the difference between groups on the selection variable

Prior to examining the difference between students on the selection variable, optimal credit hour enrollment, exploratory analyses of each dataset were conducted. In addition to the SIS dataset, both the entering student survey (ESS) dataset and the noncognitive survey (NCS) dataset were examined. The first step determined which covariates were eligible for inclusion in the model. Ideal covariates are those that were collected prior to students enrolling at the institution. For both the SIS and ESS datasets, all variables met this condition (see Appendix A and Appendix B for variables and descriptions). This was not true for the NCS dataset. Since the NCS dataset was comprised of items from a noncognitive survey administered post enrollment, some of the items specifically referenced experiences that occurred after enrollment. Scales that addressed these experiences were not retained in subsequent analyses (see Appendix C for items and scale descriptions). The following scales were dropped: self-regulated learning, perceived efficacy of instructor, perceived sense of belonging, academic motivation, academic dishonesty, and feeling lost in the system.

The next step examined both missingness and distribution of covariates in the datasets. Missingness was examined in relation to other variables supplied from the dataset as well as using Cochran's (1954) general rule that the expected cell frequencies are no less than one and no more than 20% are less than five. From the SIS dataset, the first generation indicator was dropped due to high levels of missing data. Similarly, the ACT reading score was dropped due to a missing data pattern that was inconsistent with the other ACT subtest scores. Additionally, the raw advanced placement credits field was dropped because the data could not be substantiated.

For covariates from the ESS dataset, two variables (live arrangements and degree plans) had two distinct response items collapsed into one to ensure that Cochran's rule (1954) was upheld. Student's age from the SIS dataset was dropped because it could not be meaningfully collapsed and did not have enough variability as a continuous item. No other items from the ESS dataset or NCS dataset required adjustment.

After assessing missing data across the datasets, the scale scores in the NCS dataset were summed. Each of the scales demonstrated sufficient internal consistency (>.80) with the exception of caring. Therefore, the scale scores rather than the individual items were used for the analysis with the exception of the caring scale. Since the caring scale (α =.62) did not demonstrate adequate internal consistency, the scale was not used for modeling and the individual variables were retained. Since the survey used to develop the ESS dataset was not designed to represent constructs, the individual items were used in modeling.

The next step for ensuring the quality of the covariates in the model included running Pearson correlations to identify significant overlap between variables scored on an interval and dichotomous scale (r > .80; see Appendix D for correlation matrix). Based on this analysis, only ACT Composite was removed. The composite score is an average of its subtests and thus was highly correlated with the individual subtests. Since the model building process for propensity score methods aims to maximizes information, the decision was made to drop ACT Composite and retain the remaining individual subtests, ACT Math and ACT English. For the categorical variables, contingency tables were examined. The items related to advanced placement (exams and courses) demonstrated significant overlap. The item assessing the number of advanced placement courses the student took was retained over the number of advancement placement exams the student took because the latter had more missing data.

Finally, the relationship between the selection variable, optimal credit hours, and the independent variables was assessed. Overall, relatively few variables demonstrated a small relationship (r = .10) with the selection variable. Therefore, to include a fuller list of covariates but retain power and reduce increased variance from nonsignificant variables, the criterion for inclusion was set at 0.07 for the ESS dataset since the individual items were not designed to be collapsed by scales. Each variable dropped from the analysis are shown in Table 1.

Table 1. Dropped variables across datasets

Variable	Reason Dropped
NCS Self-Regulated Learning	Referenced experiences after enrollment
NCS Perceived Efficacy of Instructor	Referenced experiences after enrollment
NCS Perceived Sense of Belonging	Referenced experiences after enrollment
NCS Academic Motivation	Referenced experiences after enrollment
NCS Academic Dishonesty	Referenced experiences after enrollment
NCS Feeling Lost in the System	Referenced experiences after enrollment
First Generation	High amount of missing data
ACT Reading	Irregular missing data pattern
Advanced Placement Credits (raw)	Data could not be substantiated
Age	Not enough variability
ACT Composite	High correlation with individual ACT subtests – retained subtests instead
ESS AP Exams	High correlation with AP classes – greater missing data than AP classes so AP classes retained
ESS Q91 Q92 Q93 Q94 Q95 Q96 Q97 Q98 Q99 Q910 Q911 Q912 Q101 Q102 Q103 Q104 Q105 Q107 Q109 Q1010 Q1011 Q1012 Q1013 Q1014 Q1015 Q1016 Q1017 Q1018 Q112 Q113 Q114 Q115 Q116 Q117 Q118 Q119 Q1110 Q1111 Q1112 Q1113	Relationship with optimal enrollment fell below the 0.07 threshold

Variable				Reason Dropped
Q12	Q131	Q132	Q133	
Q134	Q135	Q136	Q137	
Q138	Q139	Q1310	Q1311	
Q1312	Q1313	Q1314	Q1315	
Q1316	Q1317	Q1318	Q141	
Q142	Q143	Q144	Q145	
Q146	Q147	Q148	Q149	
Q1410	Q1411	Q1412	Q1413	
Q1414	Q1415	Q1416	Q1417	
Q151	Q152	Q154	Q155	
Q158	Q1510	Q1512	Q1513	
Q1514	Q1515	Q1517	Q1518	
Q1519	Q152			

The remaining covariates were entered into a single-level logistic regression with optimal enrollment as the outcome. The overall effect demonstrated a statistically significant difference on optimal enrollment, $X^2(76)= 207.0767$, $p<.0001$. Exploration of the estimates, illustrated in Table 2, demonstrate that the groups are not equivalent on covariates across the disparate dataset.

For covariates in the SIS dataset, honors college, academic college and summer college demonstrated significant differences between the groups. Students enrolled in optimal credit hours were more likely to participate in both honors college and summer college. Additionally, students who did not enroll in optimal credit hours were more likely to be enrolled in a major within the college of applied health sciences or the college of architecture, design and the arts majors. For the ESS dataset, degree, language, Q156, Q157 and Q1511 demonstrated significant differences between the groups. Students who enrolled in optimal credit hours were less likely to indicate that they were not planning on pursing an academic degree (degree) at the university and more likely to have English as a first language (language). Additionally, students who enrolled in optimal credit hours indicated a lower chance of working fulltime (ESS Q156) and a lower chance of playing varsity sports (ESS Q157) and indicated a higher chance of completing a bachelor’s degree (ESS Q111). For the NCS dataset, significant differences were found on academic control. Students enrolled in optimal credit hours had lower levels of academic control.

Table 2. Parameter Estimates for Logistic Regression

Variable	Value	DF	Estimate	Std. Error	Wald Chi-Square	Pr > ChiSq
Gender	Male	1	-0.0157	0.1515	0.0108	0.9174
Ethnic	Af Am	1	-0.4422	0.2804	2.4878	0.1147
Ethnic	Asian	1	-0.1248	0.1912	0.4259	0.514
Ethnic	Hisp	1	0.00507	0.1858	0.0007	0.9782
Ethnic	Other	1	0.9515	0.5105	3.4734	0.0624
Honors College*	No	1	-1.0313	0.258	15.9811	<.0001
Pell Recipient	No	1	-0.1449	0.1427	1.0316	0.3098
Academic College*	Applied Health Sciences	1	-1.5838	0.3101	26.0925	<.0001
Academic College*	Architecture, Design,& the Arts	1	-0.5472	0.2492	4.8213	0.0281
Academic College	Business Administration	1	0.1892	0.2193	0.7444	0.3883
Academic College	Education	1	1.087	0.67	2.6324	0.1047
Academic College	Engineering	1	-0.4459	0.2319	3.6979	0.0545
High School: CPS	No	1	0.1237	0.1686	0.5386	0.463
Summer College*	No	1	-0.4306	0.1749	6.0573	0.0138
Placement Math	MATH 180 and STAT 130	1	0.088	0.1905	0.2133	0.6442
Placement Math	Math 075	1	-0.4065	0.27	2.2676	0.1321
Placement Math	Math 090	1	0.1631	0.1924	0.7188	0.3965
Placement Writing	ENGL 070	1	0.6264	0.4793	1.7083	0.1912
Placement Writing	ENGL 071	1	0.1816	0.2024	0.8047	0.3697
Placement Writing	ENGL 161	1	-0.0483	0.236	0.0419	0.8378
Placement Writing	ESL 060	1	0.9632	1.8348	0.2756	0.5996
ACT English		1	0.024	0.0275	0.7635	0.3822
ACT Math		1	-0.0173	0.0229	0.5719	0.4495
High School GPA		1	0.1819	0.1945	0.8738	0.3499
ESS Live (R)	Off campus	1	0.1351	0.2792	0.234	0.6286
ESS Live (R)	Parents	1	0.0269	0.1594	0.0284	0.8662

Variable	Value	DF	Estimate	Std. Error	Wald Chi-Square	Pr > ChiSq
ESS Live (R)	Other	1	-2.1053	1.1486	3.3597	0.0668
ESS Degree (R)*	None	1	-1.2743	0.6146	4.2995	0.0381
ESS Degree (R)	Bachelors	1	-0.0461	0.1592	0.084	0.772
ESS Degree (R)	PhD or EdD	1	0.0797	0.1834	0.1886	0.6641
ESS Degree (R)	Adv. Med	1	0.2036	0.2251	0.8181	0.3657
ESS Degree (R)	Adv. Law	1	0.1553	0.5732	0.0734	0.7864
ESS Degree (R)	Other	1	-1.1712	0.6469	3.2772	0.0703
ESS Had Math Help	No	1	0.0406	0.2019	0.0404	0.8406
ESS Need Math Help	No	1	-0.0655	0.1695	0.1494	0.6991
ESS Had Sci Help	No	1	-0.2171	0.2639	0.6765	0.4108
ESS Need Sci Help	No	1	-0.04	0.1868	0.0459	0.8304
ESS Had Write Help	No	1	0.1622	0.2365	0.4703	0.4929
ESS Nd Write Help	No	1	0.1567	0.1817	0.7441	0.3884
ESS Language*	No	1	-0.3319	0.146	5.1704	0.023
ESS Religion	Buddhist	1	0.1546	0.4898	0.0996	0.7523
ESS Religion	Hindu	1	0.1453	0.3728	0.1518	0.6968
ESS Religion	Jewish	1	-0.3901	0.5804	0.4517	0.5015
ESS Religion	Muslim	1	-0.1715	0.2601	0.4345	0.5098
ESS Religion	Protestant	1	0.3079	0.245	1.5798	0.2088
ESS Religion	Catholic	1	-0.0896	0.1773	0.2552	0.6134
ESS Religion	Other Religion	1	-0.1133	0.2371	0.2284	0.6327
ESS AP Course	1-2	1	0.0939	0.1828	0.2638	0.6075
ESS AP Course	3-4	1	0.2818	0.1951	2.0857	0.1487
ESS AP Course	5+	1	0.2187	0.2553	0.734	0.3916
ESS Q106		1	-0.1507	0.0971	2.4083	0.1207
ESS Q108		1	0.0303	0.0953	0.1012	0.7504
ESS Q111		1	0.118	0.0724	2.6555	0.1032
ESS Q153		1	-0.0339	0.1045	0.1051	0.7458
ESS Q156*		1	0.1554	0.0738	4.4323	0.0353
ESS Q157*		1	0.1627	0.0732	4.9381	0.0263
ESS Q159		1	-0.0661	0.1261	0.275	0.6
ESS Q1511*		1	-0.3765	0.1351	7.7726	0.0053
ESS Q1516		1	-0.0443	0.096	0.2133	0.6442
ESS Q1520		1	-0.0377	0.1354	0.0776	0.7806
ESS Q1521		1	0.0204	0.1079	0.0359	0.8497
NCS Self-Efficacy		1	0.00312	0.0166	0.0354	0.8507
NCS Time Manage		1	-0.00198	0.0133	0.0221	0.8818

Variable	Value	DF	Estimate	Std. Error	Wald Chi-Square	Pr > ChiSq
NCS Sub WellBeing		1	0.0171	0.0174	0.9687	0.325
NCS Family Oblig.		1	0.000547	0.00933	0.0034	0.9533
NCS Grit		1	-0.00435	0.0201	0.0467	0.829
NCS Acad Control*		1	-0.1222	0.044	7.7227	0.0055
NCS Caring1		1	-0.1012	0.0611	2.7421	0.0977
NCS Caring2		1	-0.1168	0.0892	1.7124	0.1907
NCS Caring3		1	0.0088	0.0965	0.0083	0.9273
NCS Caring4		1	0.0482	0.0672	0.5149	0.473
NCS Caring5		1	0.106	0.0799	1.7582	0.1848
NCS Caring6		1	-0.036	0.0576	0.3911	0.5317
NCS Caring7		1	0.04	0.0923	0.1875	0.665
NCS Caring8		1	0.1144	0.0869	1.7345	0.1878
NCS Caring9		1	-0.049	0.096	0.2603	0.6099

Since significant differences were found between the two groups, the use of propensity score methods to address the nonequivalence between groups was warranted. This finding permitted continuation of the analysis.

Step two: Estimate the propensity score

Single-level logistic regression was used to estimate the propensity score for four separate PS models derived from a combination of the three disparate datasets.

SIS Model

The first model, SIS, was restricted to only covariates in the SIS dataset. Using only complete cases, 94.6% of the original sample was retained (n = 2,845 with 72.7% optimal enrollment). There was no evidence of multicollinearity; thus, all covariates were retained. The overall model was significant, X^2(39)=183.3497, *p*<.0001, and the Hosmer and Lemeshow Goodness-of-Fit test demonstrated adequate fit, X^2(8)= 8.7249, *p*=0.3660. The SIS model accounted for 6.87% of the variance in optimal credit enrollment and 65.8% of the cases were accurately classified with no ties.

There were significant differences between groups on academic college, summer college, honors college, math placement level, ACT Math score and high school GPA (see parameter estimates in Table 3). Students who enrolled in optimal credit levels were more likely to participate in summer college and honors college and have higher scores on the ACT Math subtest and higher high school GPAs. Additionally, students enrolled in optimal credit hours were less likely to be applied health sciences or architecture, design and the arts majors and less likely to be placed in the lowest remedial math course (Math 075).

Table 3. Parameter Estimates for SIS Model

Variable	N	% (M)	%Optimal Enroll (SD)	Bivariate
Ethnicity				X^2=4.3727, p=0.3579
African American/Black	261	9.17	68.20	
Asian	763	26.82	75.35	
Hispanic	912	32.06	72.04	
Other	116	4.08	78.45	
White	793	27.87	71.37	
Gender				X^2=0.6798, p=0.4097
Male	1333	46.85	70.67	
Female	1512	53.15	74.40	
Summer College*				X^2=10.9882, p=0.0009
Yes	565	19.86	76.11	
No	2280	80.14	71.80	
Honors College*				X^2=26.8372, p<.0001
Yes	432	89.35	89.44	
No	2413	69.66	69.80	
Academic College*				X^2=66.0252 p<.0001
Applied Health Science	98	3.5	47.96	
Architecture, Design and the Arts	175	5.96	56.00	
Business Administration	310	10.75	77.10	
Education	44	1.57	86.36	

Variable	N	% (M)	%Optimal Enroll (SD)	Bivariate
Engineering	308	10.57	70.45	
Liberal Arts & Sciences	1910	67.14	74.76	
High School: CPS				X^2=3.7951, p=0.0514
Yes	912	32.06	69.85	
No	1933	67.94	73.98	
Pell Recipient				X^2=2.2323, p=0.1352
Yes	1547	54.38	72.72	
No	1298	45.62	72.57	
Placement Writing				X^2=0.8244, p=0.9351
ESL 060	13	0.46	69.23	
ENGL 070	74	2.6	74.32	
ENGL 071	386	13.57	69.69	
ENGL 160	1520	53.43	69.67	
ENGL 161	852	29.95	79.23	
Placement Math*				X^2=10.2250, p=0.0167
MATH 075	273	15.75	59.34	
MATH 090	1021	38.77	72.87	
MATH 121,160,165 and STAT 101	448	9.6	72.54	
MATH 180 and STAT 130	1102	35.89	75.79	
ACT English				X^2=1.1952, p=0.2743
Optimal Enrollment	2067	25.14	4.64	
Less Enrollment	778	24.02	4.15	
ACT Math*				X^2=3.9355, p=0.0473
Optimal Enrollment	2067	25.14	4.30	
Less Enrollment	778	24.05	4.16	
High School GPA*				X^2=4.0718, p=0.0473
Optimal Enrollment	2067	3.34	0.37	
Less Enrollment	778	3.24	0.38	

SIS+ESS Model

The second model expanded upon the first by adding covariates from the ESS dataset. With the addition of these covariates, only 67.9% of the original sample was retained (n = 2,041 with 73.8% optimal enrollment). There was no evidence of multicollinearity; thus all covariates were retained. The overall model, SIS+ESS, was significant, $X^2(61)= 212.5261$, $p<.0001$, and the Hosmer and Lemeshow Goodness-of-Fit test demonstrated adequate fit, $X^2(8)= 11.0645$, $p= 0.1981$. The SIS model accounted for 9.9% of the variance in optimal credit enrollment and 69.7% of the cases were accurately classified with no ties.

There were significant differences between groups on the following covariates: honors college, academic college, ESS language and ESS Q1511 (see parameter estimates in Table 4). Similar to the earlier SIS model, students who enrolled in optimal credit hours were more likely to participate in honors college and less likely to be enrolled in a major within the college of applied health sciences or architecture, arts and design (academic college). Additionally, they were more likely to have English as a first language (ESS language) and indicate a great change of obtaining a bachelor's degree (ESS degree).

Table 4. Parameter Estimates for SIS+ESS Model

Variable	N	% (M)	%Optimal Enroll (SD)	Bivariate
Ethnicity				X^2=6.2403, p=0.1819
African American/Black	172	8.43	72.09	
Asian	566	27.73	75.97	
Hispanic	662	32.44	72.81	
Other	58	2.84	87.93	
White	583	28.56	72.04	
Gender				X^2=0.0804, p=0.7767
Male	917	44.93	71.43	
Female	1124	55.07	75.80	
Summer College				X^2=2.8595, p=0.0908
Yes	412	20.19	75.97	
No	1629	79.81	73.30	
Honors College*				X^2=16.8889, p=<.0001
Yes	363	17.79	89.53	
No	1678	82.21	70.44	
Academic College*				X^2=40.8218, p=<.0001

Variable	N	% (M)	%Optimal Enroll (SD)	Bivariate
Applied Health Science	62	3.04	50.00	
Architecture, Design and the Arts	124	6.08	57.26	
Business Administration	203	9.95	79.31	
Education	32	1.57	90.63	
Engineering	209	10.24	70.33	
Liberal Arts & Sciences	1411	69.13	75.69	
High School: CPS				X^2=0.2818, p=0.5955
Yes	668	32.73	70.96	
No	1373	67.27	75.24	
Pell Recipient				X^2=3.2475, p=0.0715
Yes	1122	54.97	73.98	
No	919	45.03	73.67	
Placement Writing				X^2=3.983, p=0.4083
ESL 060	7	0.34	71.43	
ENGL 070	45	2.2	75.56	
ENGL 071	264	12.93	71.21	
ENGL 160	1098	53.8	70.58	
ENGL 161	627	30.72	80.54	
Placement Math				X^2=6.2759, p=0.0989
MATH 075	176	8.62	62.50	
MATH 090	727	35.62	74.55	
MATH 121,160,165 and STAT 101	318	15.58	71.38	
MATH 180 and STAT 130	820	40.18	76.59	
ESS Living (R)				X^2=4.941, p=0.1762
Residence Halls	829	40.62	75.87	
Off campus	152	7.45	77.63	
Parents	1053	51.59	72.08	
Other	7	0.34	14.29	
ESS Degree (R)				X^2=10.0965, p=0.1206
None	15	0.73	33.33	

Variable	N	% (M)	%Optimal Enroll (SD)	Bivariate
Bachelors	490	24.01	68.98	
Masters	777	38.07	73.23	
PhD or EdD	375	18.37	74.93	
Adv. Medical	332	16.27	83.43	
Adv. Law	31	1.52	74.19	
Other	21	1.03	66.67	
ESS Had Math Help				X^2=0.0001, p=0.9935
Yes	392	19.21	72.70	
No	1649	80.79	74.11	
ESS Need Math Help				X^2=0.532, p=0.4658
Yes	627	30.72	73.37	
No	1414	69.28	74.05	
ESS Had Science Help				X^2=1.4023, p=0.2363
Yes	259	12.69	74.13	
No	1782	87.31	73.79	
ESS Need Science Help				X^2=0.3148, p=0.5748
Yes	495	24.25	72.53	
No	1546	75.75	74.26	
ESS Had Writing Help				X^2=2.8151, p=0.0934
Yes	289	14.16	71.63	
No	1752	85.84	74.20	
ESS Need Writing Help				X^2=2.6048, p=0.1065
Yes	458	22.44	70.31	
No	1583	77.56	74.86	
ESS English Language*				X^2=7.2161, p=0.0072
Yes	1450	71.04	75.45	
No	591	28.96	69.88	
ESS Religion				X^2=0.7586, p=0.9978
Buddhist	40	1.96	75.00	
Hindu	102	5	81.37	
Jewish	24	1.18	70.83	
Muslim	207	10.14	76.81	
Protestant	248	12.15	75.81	
Catholic	759	37.19	72.20	
Other religion	196	9.6	71.43	
No Affiliation	465	22.78	73.55	
ESS AP Courses				X^2=4.134, p=0.2474

Variable	N	% (M)	%Optimal Enroll (SD)	Bivariate
None	311	15.24	67.20	
1-2	665	32.58	69.47	
3-4	691	33.86	75.83	
5+	374	18.32	83.42	
ACT English				X^2=0.2888, p=0.591
Optimal Enroll	1507	25.25	4.64	
Less Enroll	534	24.19	4.24	
ACT Math				X^2=1.0562, p=0.3041
Optimal Enroll	1507	25.244857	4.265947	
Less Enroll	534	24.320225	4.086366	
HS GPA				X^2=0.4922, p=0.483
Optimal Enroll	1507	3.3639681	0.369405	
Less Enroll	534	3.2760487	0.377002	
ESS Q106				X^2=0.5458, p=0.46
Optimal Enroll	1507	1.7232913	0.755508	
Less Enroll	534	1.8501873	0.803388	
ESS Q108				X^2=0.0164, p=0.8981
Optimal Enroll	1507	2.2070338	0.806434	
Less Enroll	534	2.3220974	0.82024	
ESS Q111				X^2=2.9173, p=0.0876
Optimal Enroll	1507	3.2554745	0.9855	
Less Enroll	534	3.0093633	0.925917	
ESS Q153				X^2=0.1248, p=0.7239
Optimal Enroll	1507	1.7511612	0.695119	
Less Enroll	534	1.8838951	0.715418	
ESS Q156				X^2=3.1357, p=0.0766
Optimal Enroll	1507	2.6794957	0.876958	
Less Enroll	534	2.5393258	0.871343	
ESS Q157				X^2=2.8709, p=0.0902
optimal enroll	1507	3.1605839	0.871238	
less enroll	534	3.0299625	0.934557	
ESS Q159				X^2=0.1601, p=0.6891
Optimal Enroll	1507	1.3191772	0.525243	
Less Enroll	534	1.417603	0.571156	
ESS Q1511*				X^2=6.3074, p=0.012
Optimal Enroll	1507	1.1532847	0.411965	
Less Enroll	534	1.2621723	0.576748	

Variable	N	% (M)	%Optimal Enroll (SD)	Bivariate
ESS Q1516				X^2=0.2633, p=0.6079
Optimal Enroll	1507	1.8500332	0.801769	
Less Enroll	534	2.0187266	0.826181	
ESS Q1520				X^2=0.1218, p=0.7271
Optimal Enroll	1507	1.279363	0.52771	
Less Enroll	534	1.3445693	0.559079	
ESS Q1521				X^2=0.8649, p=0.3524
Optimal Enroll	1507	1.6031851	0.712581	
Less Enroll	534	1.7434457	0.752547	

SIS+NCS Model

The third model expanded upon the base SIS dataset with the addition of the NCS dataset. The addition of these covariates resulted in the retention of 73.6% of the original sample (n = 2,213 with 73.8% optimal enrollment). There was no evidence of multicollinearity; thus, all items were retained. The overall model was significant, $X^2(39)=183.3497$, $p<.0001$, and the Hosmer and Lemeshow Goodness-of-Fit test demonstrated adequate fit, $X^2(8)=9.5040$, $p=0.3016$. The model explained 7.95% of the variance in optimal credit enrollment and correctly classified 67.5% of cases with no ties.

There were significant differences between groups on the following covariates: honors college, academic college, summer college, math placement level, high school GPA, high school CPS and NCS academic control (see Table 5). Students who enrolled in optimal credit hours were more likely to participate in both honors college and summer college and less likely to major in applied health sciences or architecture, design and the arts. They had higher high school GPAs and were less likely to be placed in the lowest remedial math (Math 075) or attend a city public school (High School: CPS). Additionally, students enrolled in optimal credit hours demonstrated lower academic control.

Table 5. Parameter Estimates for SIS+NCS Model

Variable		N	% (M)	%Optimal Enroll (SD)	Bivariate
Ethnicity					X^2=3.249, p=0.517
	African American/Black	177	8	69.49	
	Asian	612	27.65	75.98	
	Hispanic	724	32.72	72.93	
	Other	81	3.66	80.25	
	White	619	27.97	73.18	
Gender					X^2=0.0381, p=0.8452
	Male	1031	46.59	72.26	
	Female	1182	53.41	75.21	
Summer College*					X^2=12.0884,p=0.0005
	Yes	450	20.33	77.78	
	No	1763	76.97	72.83	
Honors College*					X^2=19.031, p=<.0001
	Yes	338	15.27	89.94	
	No	1875	84.73	70.93	
Academic College*					X^2=64.9969,p=<.0001
	Applied Health Science	87	3.93	47.13	
	Architecture, Design and the Arts	137	6.19	57.66	
	Business Administration	249	11.25	78.31	
	Education	32	1.45	87.50	
	Engineering	217	9.81	69.12	
	Liberal Arts & Sciences	1491	67.37	76.53	
High School: CPS*					X^2=4.1318, p=0.0421
	Yes	712	32.17	70.65	
	No	1501	67.83	75.35	
Pell Recipient					X^2=0.5081, p=0.476
	Yes	1221	55.17	73.30	
	No	992	44.83	74.50	
Placement Writing					X^2=0.6747, p=0.9544

Variable	N	% (M)	%Optimal Enroll (SD)	Bivariate
ESL 060	7	0.32	71.43	
ENGL 070	56	2.53	75.00	
ENGL 071	303	13.69	71.95	
ENGL 160	1239	55.99	71.03	
ENGL 161	608	27.47	80.43	
Placement Math*				X^2=9.232, p=0.0264
MATH 075	207	9.35	60.87	
MATH 090	799	36.1	73.97	
MATH 121,160,165 and STAT 101	343	15.5	73.47	
MATH 180 and STAT 130	864	39.04	76.97	
ACT English				X^2=3.5692, p=0.0589
Optimal Enroll	1634	24.97	7.56	
Less Enroll	579	23.80	4.24	
ACT Math				X^2=0.8721, p=0.3504
Optimal Enroll	1479	25.27	4.28	
Less Enroll	522	24.35	4.07	
HSGPA*				X^2=5.4012, p=0.0201
Optimal Enroll	1634	3.34	1.05	
Less Enroll	579	3.25	0.95	
NCS Self-Efficacy				X^2=0.7043, p=0.4013
Optimal Enroll	1634	25.32	0.91	
Less Enroll	579	25.03	0.81	
NCS Time Manage				X^2=0.1042, p=0.7468
Optimal Enroll	1634	15.93	1.05	
Less Enroll	579	15.73	1.00	
NCS Subj Well Being				X^2=2.1376, p=0.1437
Optimal Enroll	1634	16.59	1.22	
Less Enroll	579	16.25	0.98	
NCS Fam Obligation				X^2=0.8794, p=0.3484
Optimal Enroll	1634	37.64	4.34	
Less Enroll	579	37.95	4.72	
NCS Grit				X^2=0.0088, p=0.9253
Optimal Enroll	1634	22.70	5.52	
Less Enroll	579	22.72	4.13	
NCS Acad Control*				X^2=4.7078, p=0.03
Optimal Enroll	1634	13.05	4.12	

Variable	N	% (M)	%Optimal Enroll (SD)	Bivariate
Less Enroll	579	13.13	0.38	
NCS Caring1				X^2=0.6467, p=0.4213
Optimal Enroll	1634	0.94	4.25	
Less Enroll	579	1	4.79	
NCS Caring2				X^2=0.1735, p=0.677
Optimal Enroll	1634	4.21	5.49	
Less Enroll	579	4.21	4.16	
NCS Caring3				X^2=0.1251, p=0.7235
Optimal Enroll	1634	4.44	7.78	
Less Enroll	579	4.42	4.50	
NCS Caring4				X^2=0.0013, p=0.9715
Optimal Enroll	1634	3.65	1.90	
Less Enroll	579	3.63	1.23	
NCS Caring5				X^2=0.9482, p=0.3302
Optimal Enroll	1634	0.60	0.91	
Less Enroll	579	0.61	0.86	
NCS Caring6				X^2=0.2595, p=0.6105
Optimal Enroll	1634	3.35	1.08	
Less Enroll	579	3.27	0.93	
NCS Caring7				X^2=0.0988, p=0.7533
Optimal Enroll	1634	4.11	1.24	
Less Enroll	579	4.06	1.01	
NCS Caring8				X^2=2.8741, p=0.09
Optimal Enroll	1634	4.07	1.12	
Less Enroll	579	3.96	1.05	
NCS Caring9				X^2=0.5023, p=0.4785
Optimal Enroll	1634	4.07	4.59	
Less Enroll	579	4.01	0.37	

SIS+ESS+NCS

The final model included covariates from each dataset (SIS, ESS and NCS). The addition of these resulted in the retention of 54.1% of the original sample (n = 1,627 with 74.6% optimal enrollment). There was no evidence of multicollinearity; thus, all covariates were retained. The overall model was significant, $X^2(75)$= 207.0441, p<.0001, and the Hosmer and Lemeshow

Goodness-of-Fit test demonstrated adequate fit, $X^2(8)= 4.2669$, $p= 0.8323$. The model explained 11.95% of the variance in optimal credit enrollment and correctly classified 72.2% of cases with no ties.

There were significant differences between groups on the following covariates: honors college, academic college, summer college, ESS language, ESS Q156, ESS Q157, ESS Q1511 and NCS academic control (see Table 6). Students enrolled in optimal credits were more likely to participate in both honors college and summer college and less likely to major in applied health sciences or architecture, design and the arts. They were more likely to have English as a first language and indicated a lower chance of working fulltime while in college (ESS Q156) and a lower chance of playing varsity athletics (ESS Q157). Students who enrolled in optimal credit hours indicated a greater chance of obtaining a bachelor's degree (ESS Q1511) and lower academic control.

Table 6. Parameter Estimates for SIS+ESS+NCS Model

Variable	N	% (M)	%Optimal Enrollment (SD)	Bivariate
Ethnicity				X^2=7.3646, p=0.1178
African American/Black	120	7.38	70.83	
Asian	465	28.58	76.13	
Hispanic	528	32.45	73.86	
Other	45	2.77	88.89	
White	469	28.83	73.56	
Gender				X^2=0.0114, p=0.9148
Male	739	45.52	72.67	
Female	888	54.58	76.24	
Summer College*				X^2=6.0763, p=0.0137
Yes	335	20.59	78.21	
No	1292	79.41	73.68	
Honors College*				X^2=15.9843, p=<.0001

Variable	N	% (M)	%Optimal Enrollment (SD)	Bivariate
Yes	286	17.58	90.56	
No	1341	82.42	71.22	
Academic College*				X^2=37.8401, p=<.0001
Applied Health Science	56	3.44	48.21	
Architecture, Design and the Arts	96	5.9	3.56	
Business Administration	174	10.69	78.74	
Education	84	1.48	87.50	
Engineering	151	9.28	68.87	
Liberal Arts & Sciences	1126	69.21	77.00	
High School: CPS				X^2=0.5326, p=0.4655
Yes	536	32.94	71.27	
No	1091	67.06	76.26	
Pell Recipient				X^2=1.0318, p=0.3097
Yes	904	55.56	74.56	
No	723	44.44	74.69	
Placement Writing (R)				X^2=2.3237, p=0.508
ENGL 070	37	2.27	75.68	
ENGL 071	211	12.97	73.46	
ENGL 160	915	56.24	71.37	
ENGL 161	464	28.52	81.47	
Placement Math				X^2=6.413, p=0.0932
MATH 075	138	8.48	63.04	
MATH 090	581	35.71	75.22	
MATH 121,160,165 and STAT 101	248	15.24	72.18	
MATH 180 and STAT 130	660	40.57	77.42	
ESS Live (R)				X^2=3.6863, p=0.2974
Residence Halls	647	39.77	76.66	
Off Campus	108	6.64	76.85	
Parents	866	53.23	73.21	
Other	6	0.37	16.67	
ESS Degree (R)				X^2=9.0471, p=0.1709
None	14	0.86	35.71	
Bachelors	397	24.4	71.79	
Masters	622	38.23	72.99	
PhD or EdD	307	18.87	76.22	

Variable	N	% (M)	%Optimal Enrollment (SD)	Bivariate
Adv. Medical	250	15.37	84.00	
Adv. Law	25	1.54	80.00	
Other	12	0.74	50.00	
ESS Had Math Help				X^2=0.0378, p=0.8459
Yes	320	19.67	73.44	
No	1307	80.33	74.90	
ESS Need Math Help				X^2=0.1469, p=0.7015
Yes	499	30.67	73.75	
No	1128	69.33	75.00	
ESS Had Science Help				X^2=0.6657, p=0.4146
Yes	211	12.97	74.88	
No	1416	87.03	74.58	
ESS Need Science Help				X^2=0.0456, p=0.8309
Yes	396	24.34	72.47	
No	1231	75.66	75.30	
ESS Had Writing Help				X^2=0.4685, p=0.4937
Yes	236	14.51	73.31	
No	1391	85.49	74.84	
ESS Need Writing Help				X^2=0.7395, p=0.3898
Yes	372	22.86	70.97	
No	1255	77.14	75.70	
ESS English Lang*				X^2=5.1637, p=0.0231
Yes	482	29.63	70.33	
No	1145	70.37	76.42	
ESS Religion				X^2=4.4332, p=0.7287
Buddhist	32	1.97	78.13	
Hindu	80	4.92	81.25	
Jewish	19	1.17	68.42	
Muslim	170	10.45	77.06	
Protestant	186	11.43	80.11	
Catholic	619	38.05	72.86	
Other Religion	163	10.02	71.78	
No affiliation	358	22	73.46	
ESS AP Courses				X^2=2.4399, p=0.4863
None	263	16.16	68.44	
1-2	513	31.53	71.54	
3-4	559	34.36	76.03	

Variable	N	% (M)	%Optimal Enrollment (SD)	Bivariate
5+	292	17.95	82.88	
ACT English				X^2=0.7684, p=0.3807
Optimal Enroll	1214	25.101318	4.60246	
Less Enroll	413	23.941889	4.178832	
ACT Math				X^2=0.5732, p=0.449
Optimal Enroll	1214	25.149918	4.32088	
Less Enroll	413	24.387409	4.182144	
HS GPA				X^2=0.8667, p=0.3519
Optimal Enroll	1214	3.365626	0.36998	
Less Enroll	413	3.2764649	0.381819	
ESS Q106				X^2=2.404, p=0.121
Optimal Enroll	1214	1.7273476	0.749142	
Less Enroll	413	1.8619855	0.828985	
ESS Q108				X^2=0.1, p=0.7519
Optimal Enroll	1214	2.2100494	0.790211	
Less Enroll	413	2.3075061	0.824307	
ESS Q111				X^2=2.6806, p=0.1016
Optimal Enroll	1214	3.2586491	0.975708	
Less Enroll	413	3.0338983	0.918427	
ESS Q153				X^2=0.1055, p=0.7454
Optimal Enroll	1214	1.7693575	0.692332	
Less Enroll	413	1.8958838	0.69065	
ESS Q156*				X^2=4.424, p=0.0354
Optimal Enroll	1214	2.6968699	0.877702	
Less Enroll	413	2.5326877	0.860132	
ESS Q157*				X^2=4.931, p=0.0264
Optimal Enroll	1214	3.1713344	0.875781	
Less Enroll	413	3	0.924321	
ESS Q159				X^2=0.2779, p=0.5981
Optimal Enroll	1214	1.3228995	0.524285	
Less Enroll	413	1.4188862	0.571433	
ESS Q1511*				X^2=7.741, p=0.0054
Optimal Enroll	1214	1.1515651	0.408193	
Less Enroll	413	1.283293	0.603096	
ESS Q1516				X^2=0.2184, p=0.6403
Optimal Enroll	1214	1.8706755	0.809413	
Less Enroll	413	2.0169492	0.810851	

Variable	N	% (M)	%Optimal Enrollment (SD)	Bivariate
ESS Q1520				X^2=0.0783, p=0.7797
Optimal Enroll	1214	1.2817133	0.527601	
Less Enroll	413	1.3535109	0.549465	
ESS Q1521				X^2=0.0369, p=0.8477
Optimal Enroll	1214	1.6169687	0.722327	
Less Enroll	413	1.7118644	0.73523	
NCS Self-Efficacy				X^2=0.0356, p=0.8503
Optimal Enroll	1214	25.418451	4.671168	
Less Enroll	413	25.065375	4.781445	
NCS Time Manage				X^2=0.0226, p=0.8804
Optimal Enroll	1214	15.96458	5.502265	
Less Enroll	413	15.661017	5.406429	
NCS Subj Well Being				X^2=0.9697, p=0.3248
Optimal Enroll	1214	16.717463	4.058729	
Less Enroll	413	16.416465	4.100192	
NCS Fam Obligation				X^2=0.0032, p=0.9548
Optimal Enroll	1214	37.756178	7.454165	
Less Enroll	413	37.966102	7.52335	
NCS Grit Scale				X^2=0.0467, p=0.8289
optimal enrollment	1214	22.764415	4.128263	
less enrollment	413	22.728814	4.598867	
NCS Acad Control*				X^2=7.7568, p=0.0054
Optimal Enroll	1214	13.079901	1.8905	
Less Enroll	413	13.154964	1.883811	
NCS Caring1				X^2=2.7754, p=0.0957
Optimal Enroll	1214	0.907743	1.108934	
Less Enroll	413	1.0169492	1.229572	
NCS Caring2				X^2=1.6904, p=0.1936
Optimal Enroll	1214	4.193575	0.915881	
Less Enroll	413	4.2276029	0.903882	
NCS Caring3				X^2=0.0057, p=0.94
Optimal Enroll	1214	4.4489292	0.798976	
Less Enroll	413	4.4309927	0.829205	
NCS Caring4				X^2=0.5057, p=0.477
Optimal Enroll	1214	3.6861614	1.034216	
Less Enroll	413	3.6343826	1.051804	
NCS Caring5				X^2=1.7601, p=0.1846
Optimal Enroll	1214	0.5939044	0.93251	

Variable	N	% (M)	%Optimal Enrollment (SD)	Bivariate
Less Enroll	413	0.5956416	0.905045	
NCS Caring6				X^2=0.3927, p=0.5309
Optimal Enroll	1214	3.3574959	1.213705	
Less Enroll	413	3.3171913	1.247763	
NCS Caring7				X^2=0.1841, p=0.6678
Optimal Enroll	1214	4.1317957	0.963862	
Less Enroll	413	4.0702179	1.020379	
NCS Caring8				X^2=1.729, p=0.1885
Optimal Enroll	1214	4.0897858	1.040488	
Less Enroll	413	4.0024213	1.087213	
NCS Caring9				X^2=0.2635, p=0.6078
Optimal Enroll	1214	4.0823723	1.002784	
Less Enroll	413	4.0387409	1.035042	

Summary

Across the four PS models, there was a declining n size with relatively stable optimal enrollment (see Table 7). The full model (SIS+ESS+NCS) dropped from 2,845 students in the SIS model to 1,627 students in the full model but optimal enrollment remained relatively stable with slightly increasing proportions of optimal enrollment with additional datasets. The significant covariates varied across models despite keeping the SIS data constant. Despite this change across models, the directionality of these relationships did not change. Therefore, significant covariates that demonstrated a positive relationship with optimal credit enrollment continued to do so when found to be significant in another model. Overall, the full model (SIS+ESS+NCS) explained the most variance and classified the most cases correctly.

Table 7. Summary of PS Models

PS Model	%Variance Accounted	%Correctly Classified	Significant Covariates (p<.05)
SIS (n = 2,845 72.7% optimal enrollment)	6.87%	65.8%	academic college, summer college, honors college, math placement level, ACT Math score and high school GPA
SIS+ESS (n = 2,041 73.8% optimal enrollment)	9.9%	69.7%	honors college, academic college, ESS language and ESS Q1511
SIS+NCS (n = 2,213 73.8% optimal enrollment)	7.95%	67.5%	honors college, academic college, summer college, math placement level, high school GPA, high school CPS and NCS academic control
SIS+ESS+NCS (n = 1,627 74.6% optimal enrollment)	11.95%	72.2%	honors college, academic college, summer college, ESS language, ESS Q156, ESS Q157, ESS Q1511 and NCS academic control

Step three: Assess the region of common support

The region of common support for each PS model was visually inspected using frequency distributions and boxplots. In addition, data were trimmed using a conservative approach, removing only extreme outliers. Cases with propensity scores greater than the 99th percentile of the treated cases and lower than the 1st percentile of the control cases were trimmed from the datasets.

SIS Model

Figure 1 displays the density of propensity scores for both groups. Both groups have the highest density of propensity scores between 0.68 and 0.78, but the propensity scores for the treatment group (labeled F1_15=1) were denser than the control group at the higher propensity scores (>.80). Figure 2 illustrates that the mean propensity score is slightly higher for the treatment group (labeled ‘1’). Overall, the figures demonstrated that sufficient overlap existed between the groups. Following this, the data the data were trimmed to remove extreme outliers. As a result, the SIS Model lost 32 cases.

Figure 1. Region of Common Support: SIS Model

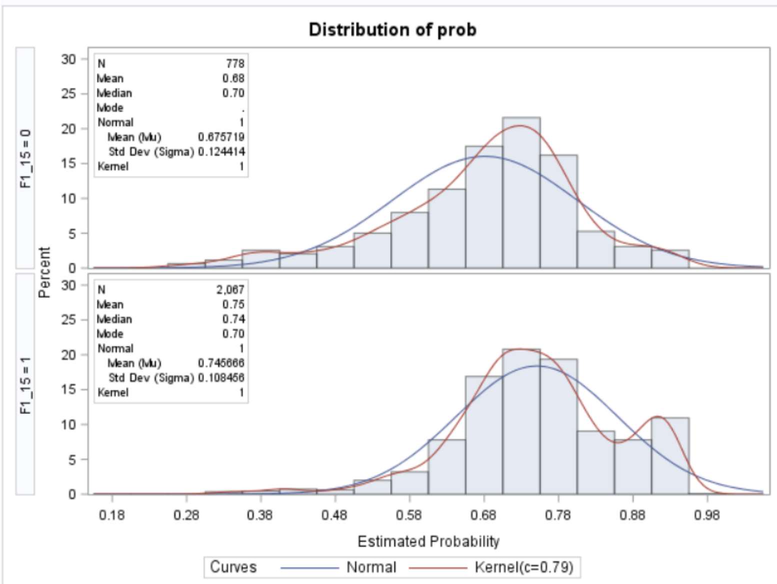
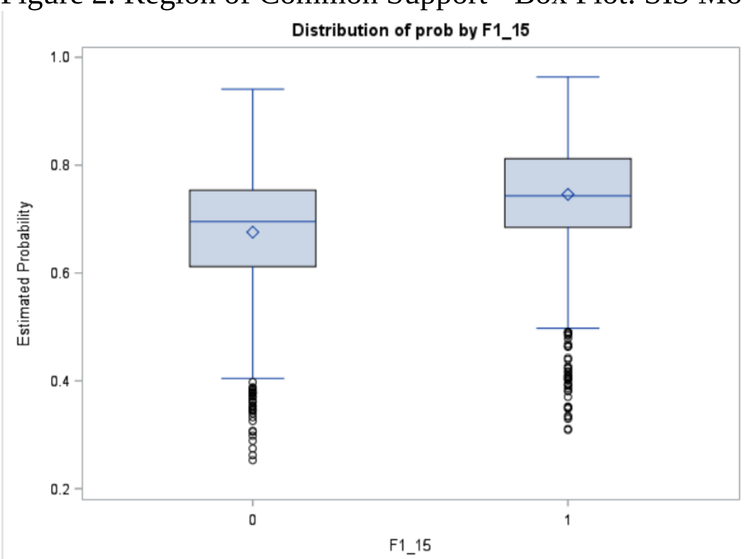


Figure 2. Region of Common Support - Box Plot: SIS Model



SIS+ESS Model

Similar to the SIS model, the Figure 3 and Figure 4 demonstrate that there is sufficient overlap between the treatment and control group for the SIS+ESS model. Specifically, Figure 3 illustrates that both groups have the highest density of propensity scores around 0.78, but the treatment group (labeled F1_15=1) had a greater density of higher propensity scores (>.90). Figure 4 illustrates that the treatment group (labeled '1') has a higher mean propensity score. Since sufficient overlap existed between the groups, the data were trimmed resulting in the loss of 22 cases.

Figure 3. Region of Common Support: SIS+ESS Model

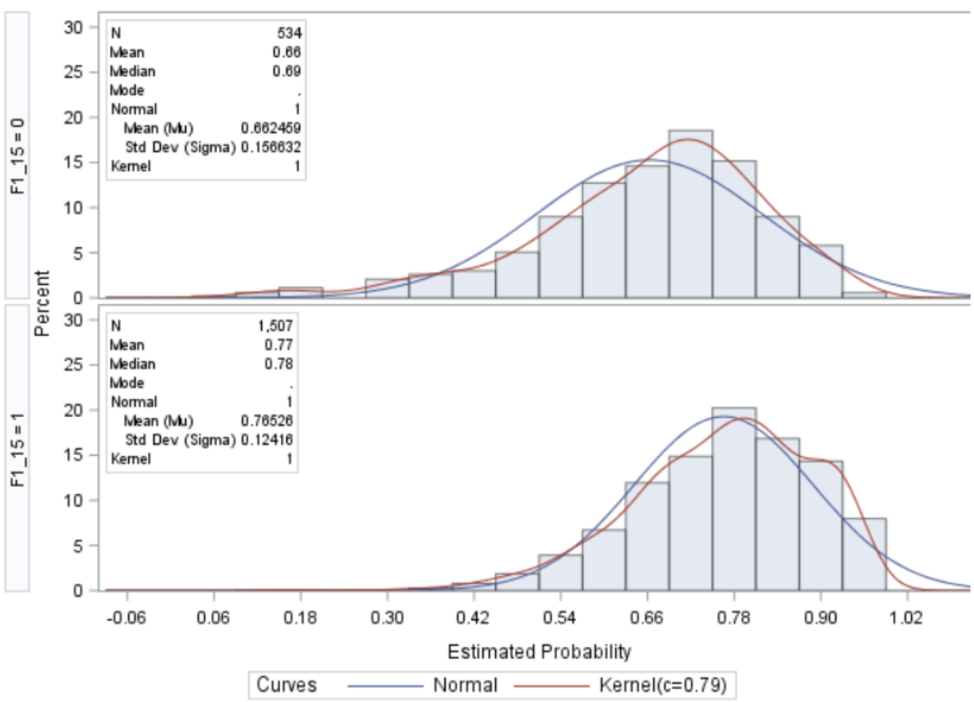
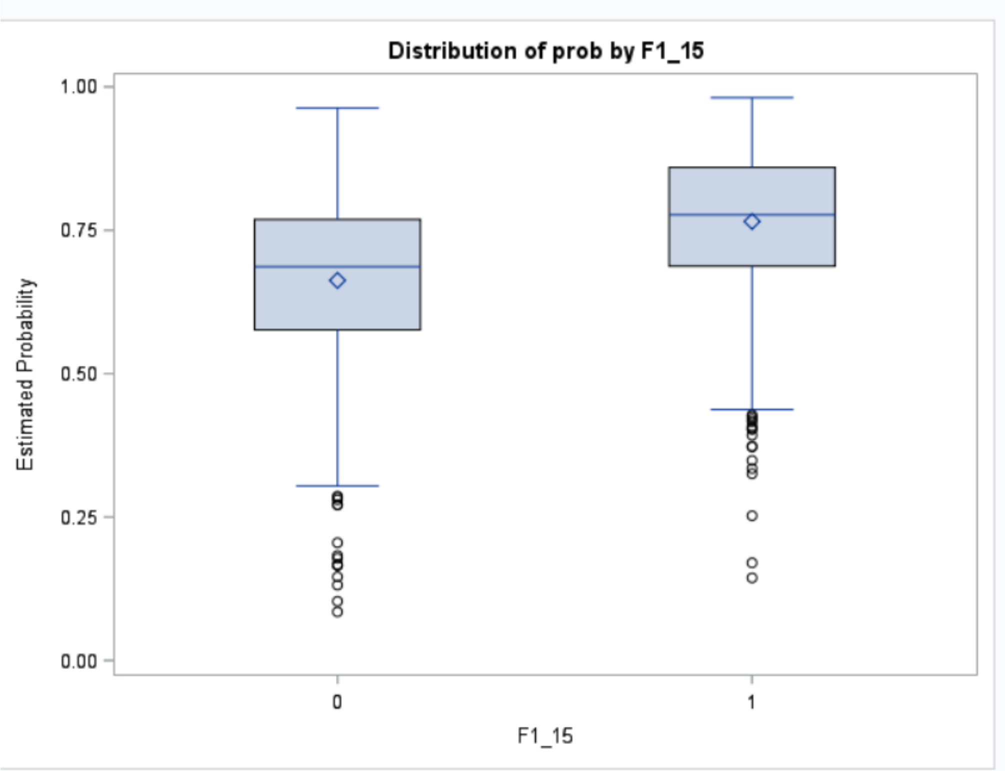


Figure 4. Region of Common Support - Box Plot: SIS+ESS Model



SIS+NCS Model

Similar to the prior models, Figure 5 and Figure 6 demonstrate that there is sufficient overlap between the treatment and control group for the SIS+NCS model. Specifically, Figure 5 illustrates that both groups have the highest density of propensity scores around 0.75, but the treatment group (labeled F1_15=1) had a greater density of higher propensity scores (>.87). Figure 6 also illustrates that mean propensity score is higher for the treatment group (labeled '1'). Since sufficient overlap existed between the groups, the data were trimmed resulting in the loss of 26 cases.

Figure 5. Region of Common Support: SIS+NCS Model

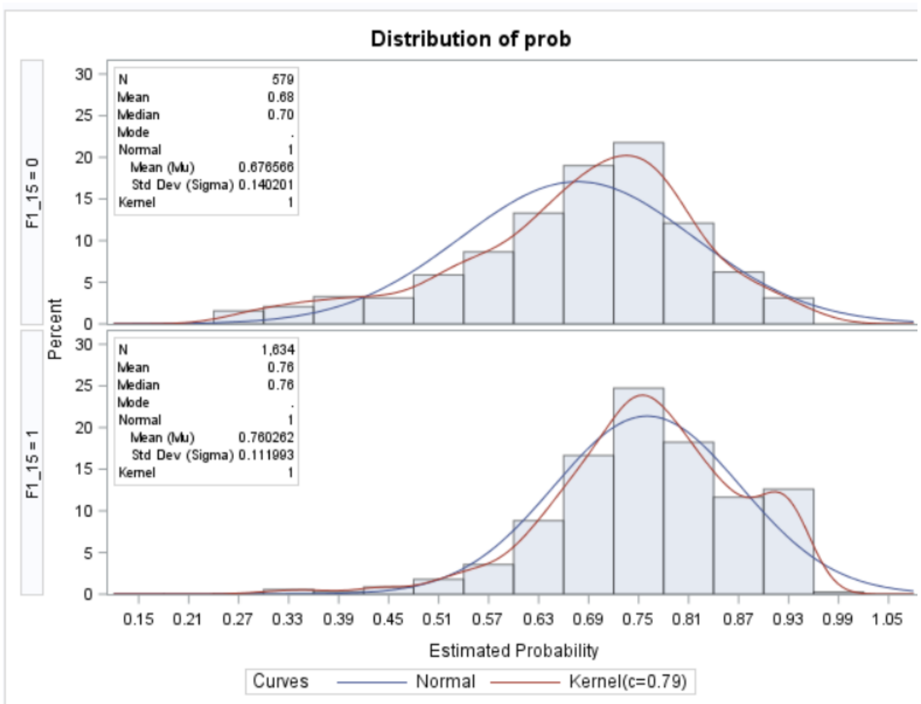
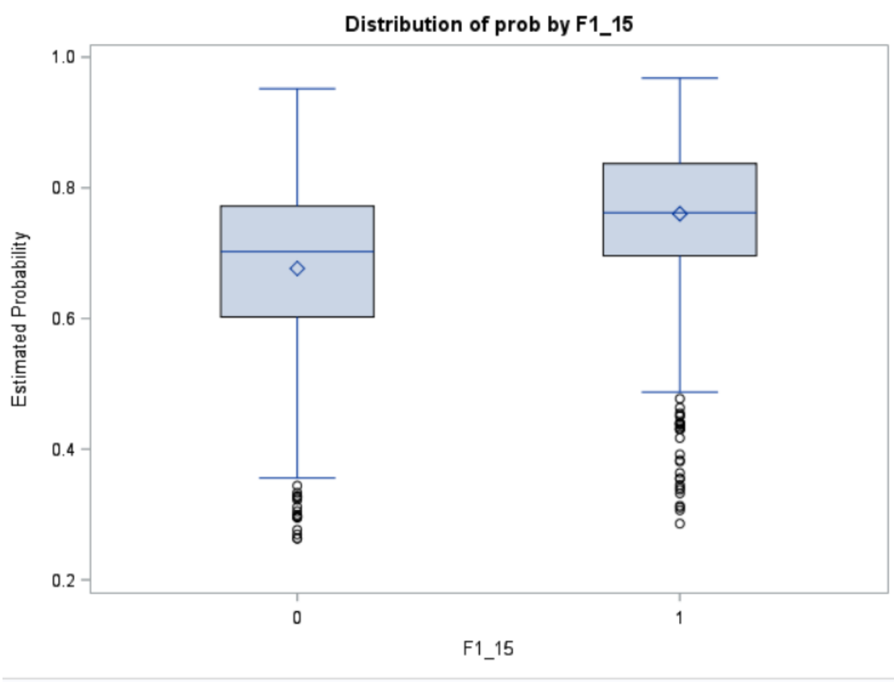


Figure 6. Region of Common Support - Box Plot: SIS+NCS Model



SIS+ESS+NCS Model

Again similar to the prior models, Figure 7 and Figure 8 demonstrate that there is sufficient overlap between the treatment and control group for the SIS+ESS+NCS model but the amount of overlap is less than with the prior models. Specifically, Figure 7 illustrates that the groups do not share a peak density for propensity score with the treatment group (labeled F1_15=1) having a peak density at a higher propensity score. Figure 8 substantiates illustrating the higher mean for the treatment group (labeled '1') but also illustrates that the range of scores is wider with the control group. Despite this increasing distance between the groups, sufficient overlap existed. Following this, the data were trimmed resulting in the loss of 19 cases.

Figure 7. Region of Common Support: SIS+ESS+NCS Model

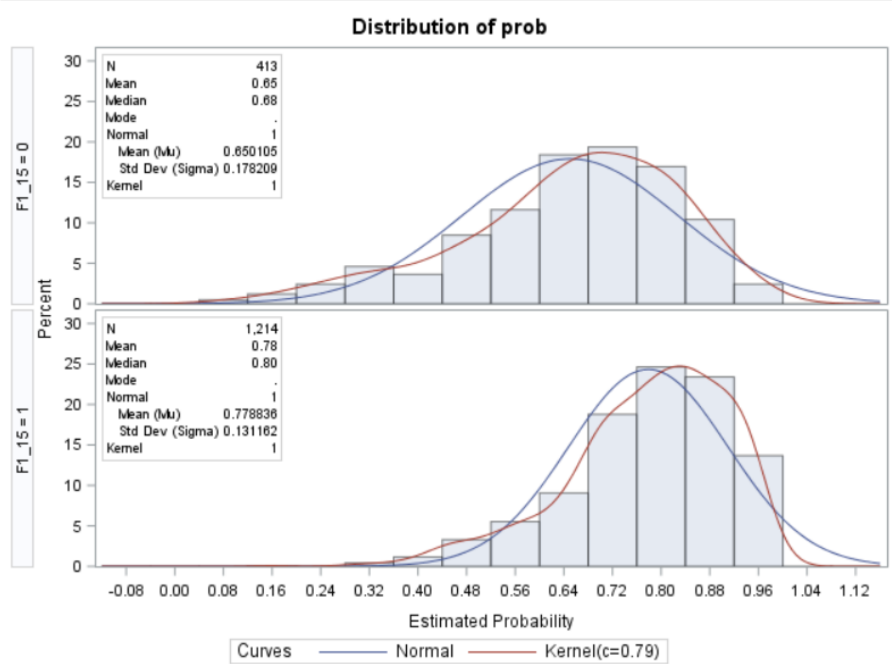
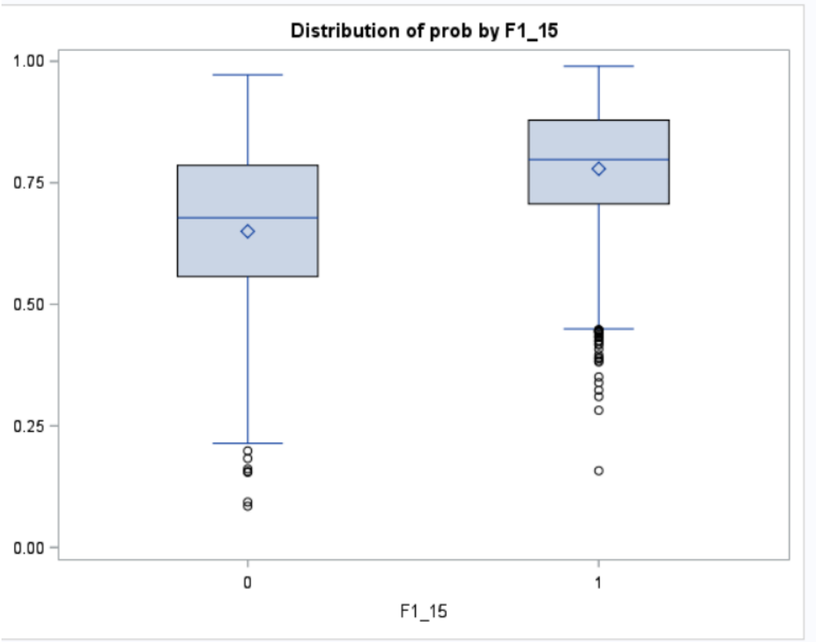


Figure 8. Region of Common Support - Box Plot: SIS+ESS+NCS Model



Summary

Each of the models demonstrated sufficient overlap to move to the next stage of analysis, conditioning the propensity score. Although overlap was achieved, the area of common support was greatest with the SIS model and smallest with the full model. This finding is a result of the increasing quality of the full PS model. As the ability to differentiate between these two groups increased, the area of overlap naturally decreased. Although this is an expected finding, it does have implications for propensity score conditioning. A reduced area of common support will lead to fewer matches when restrictions (i.e., caliper widths) are applied to the matching strategy.

Step four: Propensity Score Conditioning

The propensity score was conditioned using 1:1 nearest neighbor matching (without replacement) incorporating two different matching algorithms, greedy and greedy 5→1. As would be expected, the highest number of matches is achieved when the least stringent, no caliper width, criterion is applied to PS model and the fewest number of matches is achieved with the most stringent matching restriction, greedy 5→1. Table 8 illustrates the decreasing sample size within a PS model. For the SIS model, the treatment group dropped from 769 with no caliper width to 728 when greedy 5→1 was used to condition the propensity score. This same pattern occurred within each of the PS models.

Although these generalities are true, the match loss is greatest for the full model. Within the SIS model, the least restrictive conditioning strategy resulted in a match high of 769 and the most restrictive conditioning strategy resulted in a match low of 728. The difference between the high and low matches represents a 5.3% match loss. Applying this same approach to the other PS models, there is a 6.3% match loss for SIS+ESS, a 10.2% match loss for SIS+NCS and a 15.5% match loss for the full model, SIS+ESS+NCS. Therefore, there is not only a general impact of increasing restrictions and decreasing match sizes but the impact is varied across the PS models with the greatest impact on the most complex models.

Table 8. Description of Matching Schemes and Resample Size

	N of Sample (Before Conditioning)		N of the New Sample	
Matching Schemes	Treated	Control	Treated	Control
1. SIS, greedy, no caliper	2044	769	769	769
2. SIS, greedy, .25 caliper			751	751
3. SIS, greedy, .1 caliper			740	740
4. SIS, greedy 5→1			728	728
7. SIS + ESS, greedy, no caliper	1491	528	528	528
8. SIS + ESS, greedy, .25 caliper			505	505
9. SIS + ESS, greedy, .1 caliper			499	499
10. SIS + ESS, greedy 5→1			495	495
11. SIS + NCS, greedy, no caliper	1616	571	571	571
12. SIS + NCS, greedy, .25 caliper			548	548
13. SIS + NCS, greedy, .1 caliper			540	540
14. SIS + NCS, greedy 5→1			513	513

Matching Schemes	N of Sample (Before Conditioning) Treated	Control	N of the New Sample Treated	Control
15. SIS + ESS+ NCS, greedy, no caliper	1201	407	407	407
16. SIS + ESS + NCS, greedy, .25 caliper			378	378
17. SIS + ESS + NCS, greedy, .1 caliper			376	376
18. SIS + ESS + NCS, greedy 5→1			344	344

Step five: Assessment of balance

Balance was assessed using both statistical significance and standardized mean difference (SMD). For balance to be achieved, 90 percent of the covariates need to be balanced – meaning that the covariates do not demonstrate significant differences between groups (Rubin, 2001; Shadish & Steiner, 2010). Table 9 demonstrates the balance using both statistical significance and SMD.

When the PS models wer conditioned using greedy, no caliper, balance was not achieved. Significant differences persisted between the groups under this condition with the SMD approach demonstrating greater sensitivity. The failure to achieve balance means that the groups are not equivalent in expectation and selection bias remains. The remaining conditioning strategies adequately achieved balance across PS models. Thus, even a modest caliper width of .25 was capable of achieving balance.

Table 9. Covariate Balance across Matching Schemes

Matching Scheme	Covariates Significant ($p<.05$) after Matching	Covariates SMD > .15 after Matching
1. SIS, greedy, no caliper	ACT Math, HS CPS, Summer College, Honor College, Academic College	Honors College, Academic College, ACT Math, Placement Writing, HS GPA, ACT English, Placement Math, Ethnic
2. SIS, greedy, .25 caliper	None	None

Matching Scheme	Covariates Significant (p<.05) after Matching	Covariates SMD > .15 after Matching
3. SIS, greedy, .1 caliper	None	None
4. SIS, greedy 5→1	None	None
5. SIS + ESS, greedy, no caliper	Academic College, ESS Eng. Lang, Honors College, ESS Q111, ESS Q1511	Honors College, Academic College, ESS Q111, ESS Degree, ESS AP Course, Placement Math, ACT English, Placement Writing, ACT Math, HSGPA, Q156 ESS Eng. Lang, ESS Live, HSCPS, ESS Q108, ESS Q159, ESS Q1521, ESS Q153, ESS Q1511, ESS Q106, ESS Q1516
6. SIS + ESS, greedy, .25 caliper	None	None
7. SIS + ESS, greedy, .1 caliper	None	None
8. SIS + ESS, greedy 5→1	None	None
9. SIS + NCS, greedy, no caliper	Honors College, Academic College, Summer College, NCS Academic Control, ACT English, HS CPS, HS GPA	Honors College, Academic College, HS GPA, ACT English, Placement Math, Placement Writing, ACT Math
10. SIS + NCS, greedy, .25 caliper	Honors College	None
11. SIS + NCS, greedy, .1 caliper	None	None
12. SIS + NCS, greedy 5→1	None	None
13. SIS + ESS + NCS, greedy, no caliper	Honors College, Academic College, NCS Academic Control, ESS Q1511, ESS Eng. Lang	Honors College, Academic College, ACT English, Placement Writing, ESS Degree, HS GPA, Placement Math, ESS Q11, ESS Q156, Ethnic, ACT Math, ESS AP courses, ESS Live, ESS Q157, ESS Eng. Lang, NCS

Matching Scheme	Covariates Significant (p<.05) after Matching	Covariates SMD > .15 after Matching
		Caring8,ESS Q1520, ESS Q1516, ESS Q1511
14. SIS + ESS + NCS, greedy, .25 caliper	None	None
15. SIS + ESS + NCS, greedy, .1 caliper	None	None
16. SIS + ESS + NCS, greedy 5→1	ESS Q1511	None

Step six: Estimate the ATE

To determine the stability of the outcome under different estimation and matching conditions, the average treatment effect of the treated (ATE) was estimated using McNemar's test, $p < .05$ for paired samples. The difference in the probability of first year retention between treatment groups was estimated directly by the difference in proportions between treated and untreated students in the propensity score matched sample. Across the 16 matching schemes, 13 matching schemes demonstrated the significant impact of optimal credit hour enrollment on retention, with students who enrolled in optimal credit hours retaining at a higher rate (see Table 10). As the full set of covariates were added (SIS+ESS+NCS), the impact of enrolling in 15 or more credit hours was no longer significant. The only exception to this is when greedy matching, no caliper was the conditioning strategy.

Table 10. Average Treatment Effect across Matching Schemes.

Matching Schemes	Effect
1. SIS, greedy, no caliper*	26.6667, p<.0001
2. SIS, greedy, .25 caliper*	19.5932, p<.0001
3. SIS, greedy, .1 caliper*	15.8127, p<.0001

Matching Schemes	Effect
4. SIS, greedy 5→1*	10.6838, p= 0.0011
5. SIS + ESS, greedy, no caliper*	12.4233, p=0.0004
6. SIS + ESS, greedy, .25 caliper*	6.3210, p=0.0119
7. SIS + ESS, greedy, .1 caliper*	8.3988, p=0.0038
8. SIS + ESS, greedy 5→1*	4.5849, p=0.0323
9. SIS + NCS, greedy, no caliper*	7.9024, p=0.0049
10. SIS + NCS, greedy, .25 caliper*	5.0359, p=0.0248
11. SIS + NCS, greedy, .1 caliper*	4.3653, p=0.0367
12. SIS + NCSS, greedy 5→1*	6.7368, p= 0.0094
13. SIS + ESS + NCS, greedy, no caliper*	4.5660, p=0.0326
14. SIS + ESS + NCS, greedy, .25 caliper	3.1391, p=0.0764
15. SIS + ESS + NCS, greedy, .1 caliper	2.8058, p=0.0939
16. SIS + ESS + NCS, greedy 5→1	1.5319, p=0.2158

Step seven: Sensitivity analysis of unobserved covariates

To ascertain the robustness of the ATE, the sensitivity parameter (Γ) was assessed using Wilcoxon's signed rank test. Since there is no direct measure to ensure that the selection process has been adequately modeled removing all bias, sensitivity analyses serve to demonstrate how an unobserved covariate could change the inference. Values of Γ closer to 1 indicate that the findings are sensitive.

Overall, the findings were sensitive with gamma ranging from less than 1 to 1.5 across the matching schemes (see Table 11). When examining the sensitivity across models and matching schemes, the SIS model was the least sensitive. The inclusion of additional covariates beyond those found in the SIS dataset increased sensitivity. Further, when the PS model was conditioned using 5→1 digit matching, the findings were more sensitive than the PS models conditioned with caliper widths. Values of Γ for each of the matching schemes are displayed in Table 12 – Table 15 for each PS Model.

Table 11. Sensitivity Analysis

Matching Schemes	Gamma	Upper	Lower
1. SIS, greedy, no caliper	1.5	0	0.02824
2. SIS, greedy, .25 caliper	1.4	1.28E-12	0.0487
3. SIS, greedy, .1 caliper	1.3	8.65E-10	0.04544
4. SIS, greedy 5→1	1.1	4.63E-05	0.00879
5. SIS + ESS, greedy, no caliper	1.3	1.19E-07	0.04915
6. SIS + ESS, greedy, .25 caliper	1.1	0.001345	0.04597
7. SIS + ESS, greedy, .1 caliper	1.1	0.000324	0.01719
8. SIS + ESS, greedy 5→1	1	0.02607	0.02607
9. SIS + NCS, greedy, no caliper	1.1	0.00044941	0.02189
10. SIS + NCS, greedy, .25 caliper	1	0.019986	0.01999
11. SIS + NCS, greedy, .1 caliper	1	0.029951	0.02995
12. SIS + NCS, greedy 5→1	1.1	0.00106633	0.0357

Matching Schemes	Gamma	Upper	Lower
13. SIS + ESS + NCS, greedy, no caliper	1	0.025021	0.02502
14. SIS + ESS + NCS, greedy, .25 caliper	1	0.061709	0.06171
15. SIS + ESS + NCS, greedy, .1 caliper	1	0.075619	0.07562
16. SIS + ESS + NCS, greedy 5→1	>1	0.17967	0.17967

Table 12. Sensitivity Analysis, Unobserved Covariates: SIS Model

SIS_CMATCH0			SIS_CMATCH25			SIS_CMATCH1			SIS_DMATCH		
gamma	p_lower	p_upper	gamma	p_lower	p_upper	gamma	p_lower	p_upper	gamma	p_lower	p_upper
1	0.00000013	0	1	0.0000061	0.00001	1	4.8928E-05	0.00005	1	0.0008193	0.00082
1.1	0.000000002	0.00001	1.1	0.000000151	0.00015	1.1	1.488E-06	0.00093	**1.1**	**4.631E-05**	**0.00879**
1.2	2.37E-11	0.0001	1.2	0.000000003	0.00172	1.2	3.8E-08	0.00862	1.2	2.159E-06	0.0501
1.3	2.93E-13	0.00106	1.3	6.61E-11	0.01142	**1.3**	**8.65E-10**	**0.04544**	1.3	8.8E-08	0.17604
1.4	3.55E-15	0.00666	**1.4**	**1.28E-12**	**0.0487**	1.4	1.85E-11	0.15469	1.4	3E-09	0.42712
1.5	**0**	**0.02824**	1.5	2.49E-14	0.14598	1.5	3.84E-13	0.37451	1.5	1.20E-10	0.78253
1.6	0	0.08719	1.6	4.44E-16	0.33087	1.6	7.99E-15	0.69663	1.6	4.22E-12	1.16356
1.7	0	0.20813	1.7	0	0.60143	1.7	2.22E-16	1.06042	1.7	1.48E-13	1.48916
1.8	0	0.40363	1.8	0	0.92011	1.8	0	1.39179	1.8	5.33E-15	1.72019
1.9	0	0.6623	1.9	0	1.2334	1.9	0	1.64437	1.9	2.22E-16	1.86089
2	0	0.95148	2	0	1.49814	2	0	1.81039	2	0	1.93643
2.1	0	1.23178	2.1	0	1.69516	2.1	0	1.90683	2.1	0	1.97297
2.2	0	1.47251	2.2	0	1.82691	2.2	0	1.95735	2.2	0	1.98919
2.3	0	1.65902	2.3	0	1.90742	2.3	0	1.98164	2.3	0	1.99589
2.4	0	1.79136	2.4	0	1.953	2.4	0	1.9925	2.4	0	1.9985
2.5	0	1.87846	2.5	0	1.9772	2.5	0	1.99707	2.5	0	1.99947
2.6	0	1.93223	2.6	0	1.98936	2.6	0	1.99889	2.6	0	1.99982
2.7	0	1.96363	2.7	0	1.9952	2.7	0	1.9996	2.7	0	1.99994
2.8	0	1.98112	2.8	0	1.9979	2.8	0	1.99986	2.8	0	1.99998
2.9	0	1.99049	2.9	0	1.9991	2.9	0	1.99995	2.9	0	1.99999
3	0	1.99532	3	0	1.99962	3	0	1.99998	3	0	2
3.1	0	1.99775	3.1	0	1.99984	3.1	0	1.99999	3.1	0	2
3.2	0	1.99894	3.2	0	1.99994	3.2	0	2	3.2	0	2
3.3	0	1.99951	3.3	0	1.99997	3.3	0	2	3.3	0	2
3.4	0	1.99977	3.4	0	1.99999	3.4	0	2	3.4	0	2
3.5	0	1.9999	3.5	0	2	3.5	0	2	3.5	0	2

Table 13. Sensitivity Analysis, Unobserved Covariates, SIS+ESS Model

SIS+ESS, no caliper			SIS+ESS, .25 caliper			SIS+ESS, .1 caliper			SIS+ESS, digit		
gamma	p_lower	p_upper	gamma	p_lower	p_upper	Gamma	p_lower	p_upper	gamma	p_lower	p_upper
1	0.00028815	0.00029	1	0.00930969	0.00931	1	0.00280273	0.0028	**1**	**0.02607**	**0.02607**
1.1	0.00002321	0.00252	**1.1**	**0.00134493**	**0.04597**	**1.1**	**0.00032365**	**0.01719**	1.1	0.004717	0.10408
1.2	1.708E-06	0.01346	1.2	0.00017026	0.14845	1.2	3.3224E-05	0.06708	1.2	0.00074	0.28051
1.3	**1.19E-07**	**0.04915**	1.3	1.9703E-05	0.34714	1.3	3.159E-06	0.18499	1.3	0.000105	0.56309
1.4	8E-09	0.13293	1.4	2.151E-06	0.63517	1.4	2.87E-07	0.39017	1.4	0.000014	0.90752
1.5	5.50E-10	0.28366	1.5	2.27E-07	0.96591	1.5	2.5E-08	0.66928	1.5	0.000002	1.2452
1.6	3.76E-11	0.50198	1.6	2.4E-08	1.28065	1.6	2E-09	0.98026	1.6	0	1.52318
1.7	2.62E-12	0.7666	1.7	2E-09	1.53782	1.7	1.97E-10	1.27475	1.7	0	1.72192
1.8	1.87E-13	1.04344	1.8	2.52E-10	1.72341	1.8	1.77E-11	1.5188	1.8	0	1.84869
1.9	1.38E-14	1.29978	1.9	2.65E-11	1.84441	1.9	1.61E-12	1.70006	1.9	0	1.92238
2	1.11E-15	1.51418	2	2.83E-12	1.917	2	1.51E-13	1.82302	2	0	1.96211
2.1	0	1.67886	2.1	3.11E-13	1.95767	2.1	1.47E-14	1.90041	2.1	0	1.98225
2.2	0	1.79665	2.2	3.49E-14	1.97921	2.2	1.55E-15	1.94619	2.2	0	1.99196
2.3	0	1.87599	2.3	4.00E-15	1.9901	2.3	2.22E-16	1.97191	2.3	0	1.99646
2.4	0	1.9268	2.4	4.44E-16	1.99541	2.4	0	1.98575	2.4	0	1.99847
2.5	0	1.95799	2.5	0	1.99791	2.5	0	1.99295	2.5	0	1.99935
2.6	0	1.97646	2.6	0	1.99907	2.6	0	1.99658	2.6	0	1.99973
2.7	0	1.98708	2.7	0	1.99959	2.7	0	1.99837	2.7	0	1.99989
2.8	0	1.99303	2.8	0	1.99982	2.8	0	1.99923	2.8	0	1.99995
2.9	0	1.99629	2.9	0	1.99992	2.9	0	1.99964	2.9	0	1.99998
3	0	1.99805	3	0	1.99997	3	0	1.99983	3	0	1.99999
3.1	0	1.99898	3.1	0	1.99999	3.1	0	1.99992	3.1	0	2
3.2	0	1.99948	3.2	0	1.99999	3.2	0	1.99997	3.2	0	2
3.3	0	1.99973	3.3	0	2	3.3	0	1.99998	3.3	0	2
3.4	0	1.99986	3.4	0	2	3.4	0	1.99999	3.4	0	2
3.5	0	1.99993	3.5	0	2	3.5	0	2	3.5	0	2

Table 13. Sensitivity Analysis, Unobserved Covariates, SIS+NCS Model

SIS+NCS, no caliper			SIS+NCS, .25 caliper			SIS+NCS, .1 caliper			SIS+NCS, digit		
gamma	p_lower	p_upper	gamma	p_lower	p_upper	gamma	p_lower	p_upper	gamma	p_lower	p_upper
1	0.00372925	0.00373	**1**	**0.019986**	**0.01999**	**1**	**0.029951**	**0.02995**	1	0.00723454	0.00723
1.1	**0.00044941**	**0.02189**	1.1	0.003261	0.08688	1.1	0.005331	0.11972	**1.1**	**0.00106633**	**0.0357**
1.2	4.7909E-05	0.08207	1.2	0.000458	0.24916	1.2	0.000813	0.31925	1.2	0.00013997	0.11707
1.3	4.711E-06	0.21815	1.3	0.000058	0.52272	1.3	0.000111	0.62947	1.3	1.7007E-05	0.28084
1.4	4.41E-07	0.44513	1.4	0.000007	0.86839	1.4	0.000014	0.99306	1.4	1.968E-06	0.53021
1.5	0.00000004	0.74153	1.5	0.000001	1.21537	1.5	0.000002	1.33326	1.5	2.22E-07	0.8338
1.6	4E-09	1.05901	1.6	0	1.50488	1.6	0	1.599	1.6	2.5E-08	1.14228
1.7	3.26E-10	1.34844	1.7	0	1.71292	1.7	0	1.7785	1.7	3E-09	1.41292
1.8	2.98E-11	1.57967	1.8	0	1.84538	1.8	0	1.88628	1.8	3.09E-10	1.62352
1.9	2.77E-12	1.74541	1.9	0	1.9218	1.9	0	1.94512	1.9	3.52E-11	1.77206
2	2.64E-13	1.85406	2	0	1.96249	2	0	1.97485	2	4.10E-12	1.8687
2.1	2.58E-14	1.92021	2.1	0	1.98278	2.1	0	1.98896	2.1	4.89E-13	1.92751
2.2	2.66E-15	1.9581	2.2	0	1.99238	2.2	0	1.99532	2.2	6.00E-14	1.96139
2.3	2.22E-16	1.97873	2.3	0	1.99673	2.3	0	1.99807	2.3	7.55E-15	1.98005
2.4	0	1.98951	2.4	0	1.99863	2.4	0	1.99922	2.4	8.88E-16	1.98995
2.5	0	1.99494	2.5	0	1.99944	2.5	0	1.99969	2.5	2.22E-16	1.99504
2.6	0	1.99761	2.6	0	1.99977	2.6	0	1.99988	2.6	0	1.99759
2.7	0	1.99889	2.7	0	1.99991	2.7	0	1.99995	2.7	0	1.99885
2.8	0	1.99949	2.8	0	1.99996	2.8	0	1.99998	2.8	0	1.99945
2.9	0	1.99977	2.9	0	1.99999	2.9	0	1.99999	2.9	0	1.99974
3	0	1.9999	3	0	1.99999	3	0	2	3	0	1.99988
3.1	0	1.99995	3.1	0	2	3.1	0	2	3.1	0	1.99994
3.2	0	1.99998	3.2	0	2	3.2	0	2	3.2	0	1.99997
3.3	0	1.99999	3.3	0	2	3.3	0	2	3.3	0	1.99999
3.4	0	2	3.4	0	2	3.4	0	2	3.4	0	1.99999
3.5	0	2	3.5	0	2	3.5	0	2	3.5	0	2

Table 14. Sensitivity Analysis, Unobserved Covariates: SIS+ESS+NCS Model

SIS+ESS+NCS, no caliper			SIS+ESS+NCS, .25 caliper			SIS+ESS+NCS, .1 caliper			SIS+ESS+NCS, digit		
gamma	p_lower	p_upper	gamma	p_lower	p_upper	gamma	p_lower	p_upper	gamma	p_lower	p_upper
1	**0.025021**	**0.02502**	**1**	**0.061709**	**0.06171**	**1**	**0.075619**	**0.07562**	**1**	**0.17967**	**0.17967**
1.1	0.006309	0.07982	1.1	0.017336	0.17447	1.1	0.023789	0.19583	1.1	0.07116	0.37908
1.2	0.001478	0.19148	1.2	0.004424	0.3718	1.2	0.006885	0.39339	1.2	0.02596	0.64656
1.3	0.00033	0.36931	1.3	0.001055	0.64065	1.3	0.00188	0.65275	1.3	0.00892	0.94236
1.4	0.000071	0.60232	1.4	0.000241	0.9414	1.4	0.000494	0.93809	1.4	0.00294	1.22366
1.5	0.000015	0.86354	1.5	0.000053	1.22908	1.5	0.000127	1.21094	1.5	0.00094	1.46121
1.6	0.000003	1.12186	1.6	0.000012	1.47186	1.6	0.000032	1.44419	1.6	0.0003	1.64379
1.7	0.000001	1.35258	1.7	0.000003	1.65713	1.7	0.000008	1.62641	1.7	0.00009	1.77397
1.8	0	1.54229	1.8	0.000001	1.78747	1.8	0.000002	1.75878	1.8	0.00003	1.8614
1.9	0	1.68804	1.9	0	1.87331	1.9	0.000001	1.84946	1.9	0.00001	1.91737
2	0	1.79398	2	0	1.92691	2	0	1.90868	2	0	1.95185
2.1	0	1.86755	2.1	0	1.95896	2.1	0	1.94589	2.1	0	1.97245
2.2	0	1.91674	2.2	0	1.97746	2.2	0	1.96854	2.2	0	1.98447
2.3	0	1.94864	2.3	0	1.98784	2.3	0	1.98199	2.3	0	1.99134
2.4	0	1.96881	2.4	0	1.99354	2.4	0	1.98982	2.4	0	1.99521
2.5	0	1.9813	2.5	0	1.9966	2.5	0	1.9943	2.5	0	1.99737
2.6	0	1.9889	2.6	0	1.99823	2.6	0	1.99683	2.6	0	1.99856
2.7	0	1.99347	2.7	0	1.99908	2.7	0	1.99825	2.7	0	1.99921
2.8	0	1.99618	2.8	0	1.99953	2.8	0	1.99904	2.8	0	1.99957
2.9	0	1.99778	2.9	0	1.99976	2.9	0	1.99947	2.9	0	1.99977
3	0	1.99871	3	0	1.99988	3	0	1.99971	3	0	1.99987
3.1	0	1.99925	3.1	0	1.99994	3.1	0	1.99984	3.1	0	1.99993
3.2	0	1.99957	3.2	0	1.99997	3.2	0	1.99991	3.2	0	1.99996
3.3	0	1.99975	3.3	0	1.99998	3.3	0	1.99995	3.3	0	1.99998
3.4	0	1.99986	3.4	0	1.99999	3.4	0	1.99997	3.4	0	1.99999
3.5	0	1.99992	3.5	0	2	3.5	0	1.99999	3.5	0	1.99999

Chapter Summary

This chapter provided a detailed description of the results of the analysis. Overall, the results indicate that the inclusion of additional covariates from disparate data collection efforts led to improvements in the PS models but at the expense of sample size. As covariates were added to the model, sample size was greatly reduced. Additionally, the inclusion of all covariates in the full model (SIS+ESS+NCS) led to a reversal of interpretation of the major finding when restrictions were placed on the conditioning strategy (i.e., caliper widths or digit matching). Finally, the overall treatment effect was sensitive under all conditions suggesting a weak association between the treatment, optimal credit hour enrollment, and the outcome, first year retention.

CHAPTER FIVE

DISCUSSION

This chapter outlines a summary of the study and results, along with a discussion of the findings, limitations of the study and implications for future research.

Summary of the Study Purpose

This study used existing institutional data from a large, urban, public, very high research university to compare sixteen matching schemes, built from three separate datasets, to estimate the propensity score, achieve balance between groups and test the sensitivity of the average treatment effect (ATE). For each PS model, four different conditioning strategies were applied. The first four matching schemes used commonly collected data available within a student information system (referred to as SIS dataset). The next four matching schemes combined the SIS dataset with data from an entering student survey (referred to as ESS dataset). The next four matching schemes, again, combined the SIS dataset with data gathered from a noncognitive survey (referred to as NCS dataset). The final four matching schemes included data from the SIS, ESS and the NCS datasets. Each model builds upon the next, offering additional covariates for the model building process.

Research Questions

1. To what extent do the treatment and the control groups vary naively across covariates?

2. To what extent do different PS models achieve overlap between the treatment and control groups?
3. To what extent do different PS models and conditioning strategies impact the sample size?
4. To what extent do different PS models and conditioning strategies achieve balance between groups?
5. To what extent do different PS models and conditioning strategies reach the same overall conclusions?
6. To what extent is the average treatment effect robust against unobserved covariates under different PS models and conditioning strategies?

Method

Four single-level logistic regression models were derived to estimate the propensity score using PROC LOGISTIC procedure in SAS 9.4. Data from the student information system (SIS) served as the base and these data were retained throughout the models. Two separate datasets were added to the model: entering student survey (ESS) and noncognitive survey (NCS) datasets. These datasets were combined independently with the SIS dataset (SIS+ESS, SIS+NCS) as well as together (SIS+ESS+NCS). After estimation, the region of common support was visually inspected and data were trimmed to remove extreme outliers.

Next, the propensity score from each model was conditioned in four different ways: greedy – no caliper, greedy - 0.25 caliper, greedy - 0.1 caliper width and greedy 5→1 digit matching. Greedy matching was completed using %gmatch macro developed by Bergstralh and Kosanke (1995) and greedy 5→1 digit matching was completed using the macro developed by Parsons (2000).

Following this, balance was assessed using both statistical and standardized mean differences. Next, the average treatment effect (ATE) was tested using McNemar's and the sensitivity of this effect was tested using Wilcoxons's signed rank test.

Discussion of the Study's Results

Group Differences Prior to Estimation

The data were assessed to ensure group differences existed between students who enrolled in optimal credit hours and students who did not enroll in optimal credit hours. These groups demonstrated significant differences on the outcome of interest, retention, as well as on baseline covariates. These differences allowed for propensity score methods to be used.

Following this determination, each covariate was carefully examined. It was noted that the level of association between the covariates and the selection criterion was low. Very few variables reached the anticipated inclusionary small association (r = 0.1). In theoretical research, when Monte Carlo simulation is applied, researchers have the benefit of setting different levels of association for covariates. Therefore, models typically include a mixture of association levels (Zhao, 2004). In applied research, this level of control does not exist. In reviewing applied educational studies using PS methods, detailed information about the development of the selection model is often not reported (e.g., An, 2013; Keller & Lacy, 2013; Vaughn, Lalonde & Guarnieri, 2014). Therefore, it is difficult to ascertain if the small associations found in this study are common.

Although a single study cannot provide definitive assurance, a similar study using like covariates also found low correlations with few relationships above r = .1 (Clark & Cundiff, 2011). Despite this, Clark and Cundiff's study (2011) did demonstrate a wider range of association with several covariates demonstrating a moderate association with the selection variable, enrollment in a first year college course. The low level of association between covariates and the selection variable, enrollment in optimal credit hours, is not surprising considering the lack of a theoretical model. Although there is much research on the outcome of interest, retention, there is paucity of research on the selection mechanism.

Estimation of the Propensity Score

Each of the PS models were estimated separately and demonstrated adequate fit. Overall, the concordant classification rate ranged from a low of 65.8% with the SIS model to a high of 72.2% with the full model, SIS+ESS+NCS. When examining the significance of covariates across PS models (see Table 16), it is clear that the models did not perform in an additive manner. Specifically, the full model, SIS+NCS+ESS, introduced significant covariates that were not found in the SIS+ESS model. These newly introduced significant relationships are likely the result of the changing sample. The introduction of these new datasets reduced the sample size and ultimately changed the control and treatment groups across the PS models. While this was an intended feature of this research, it resulted in PS models derived from different student samples.

Table 15. Significant Covariates across PS Models

Covariate	SIS Model	SIS+ESS Model	SIS+NCS Model	SIS +NCS +ESS Model
Academic College	*	*	*	*
Honors College	*	*	*	*
Summer College	*		*	*
HS GPA	*		*	
HS CPS			*	

Covariate	SIS Model	SIS+ESS Model	SIS+NCS Model	SIS +NCS +ESS Model
Math Placement	*		*	
ACT Math	*			
ESS Language		*		*
ESS Q1511		*		*
ESS Q156				*
ESS Q157				*
Academic Control			*	*

Prior to conditioning, each of the PS models demonstrated adequate overlap. It should be noted that the inclusion of the ESS and NCS datasets led to a shift in the mean propensity score for those models. As the prediction model improved with the inclusion of relevant covariates, the distance between the mean propensity score for the treatment and the control groups widened with those in the treatment group demonstrating a higher mean propensity score. Although this is expected, as stronger PS models would likely have a narrower range of common support, it is unclear if the current results would have been replicated had the same students been retained throughout all models.

Conditioning strategies

Although the decision to include a greater set of covariates led to more data loss than the chosen conditioning strategy, the conditioning strategy did increase data loss. Overall, 5→1 digit matching led to more data loss than the other matching strategies. This is not surprising as the matching strategies requires more precision thus leaving fewer matches that meet the requirements.

Beyond the precision that the conditioning strategy applies to the data, the same conditioning strategy performs differently across PS models. Table 17 presents a reformatted version of data provided earlier. This table demonstrates that as covariates are added to the modeling process, the data loss associated with the conditioning strategy increases. For instance, when a caliper width of .1 is applied to the SIS model, there is a 3.8% data loss but when the same conditioning strategy is applied to the full model (SIS+NCS+ESS), 7.6% of the data are dropped.

Although this is not a direct result of the conditioning strategy, it is a result of its application to more complex models. As the complexity of model increased so did the standard deviation of the propensity score. This wider spread led to fewer potential matches within the conditioning specifications.

Table 16. Percentage of Pairs Lost from Same PS Model, No Caliper

	N of the New Sample	
Matching Schemes	Pairs	% loss
1. SIS, greedy, no caliper	769	-
2. SIS, greedy, .25 caliper	751	2.3%
3. SIS, greedy, .1 caliper	740	3.8%
4. SIS, greedy 5→1	728	5.3%
7. SIS + ESS, greedy, no caliper	528	-
8. SIS + ESS, greedy, .25 caliper	505	4.4%
9. SIS + ESS, greedy, .1 caliper	499	5.5%
10. SIS + ESS, greedy 5→1	495	6.3%

Matching Schemes	N of the New Sample Pairs	% loss
11. SIS + NCS, greedy, no caliper	571	-
12. SIS + NCS, greedy, .25 caliper	548	4.0%
13. SIS + NCS, greedy, .1 caliper	540	5.4%
14. SIS + NCS, greedy 5→1	513	10.2%
15. SIS + ESS+ NCS, greedy, no caliper	407	-
16. SIS + ESS + NCS, greedy, .25 caliper	378	7.1%
17. SIS + ESS + NCS, greedy, .1 caliper	376	7.6%
18. SIS + ESS + NCS, greedy 5→1	344	15.5%

Covariate Balance

In addition to the impact conditioning strategies had on data loss, covariate balance varied across models and statistical approach. Across all PS models and both statistical approaches, balance was not achieved using greedy, no caliper. Given the nature of greedy matching, it is not surprising that balance was not achieved when no caliper width was applied. Since greedy matching grabs the nearest neighbor and does not reconsider the match, the caliper widths are necessary for ensuring reasonable matches that reduce imbalance. For the 12 remaining matching schemes, balance was achieved.

Although not an aim of the study, the two approaches used to assess covariate balance led to different conclusions. Specifically, when using standardized mean difference (SMD) to assess balance without caliper widths, more covariates were identified as not balanced than when statistical significant was used. In addition, it was not only the number of unbalanced covariates that differed but also the covariates. The same covariates were not shared across the model. Therefore, using statistical significance to assess covariate balance for PS models conditioned with greedy, no caliper width not only led to fewer imbalanced covariates but also different unbalanced covariates.

Although the SMD approach was more sensitive when the greedy, no caliper conditioning strategy, statistical significance was more sensitive as the PS model complexity increased. For the PS models that incorporated the NCS dataset, several covariates were identified as not balanced while the SMD approach found all covariates to be balanced. Although these two approaches both supported the general balance of the covariates, statistical significance was on the edge of concluding the opposite. It is evident that the way balance is assessed does impact the finding.

Treatment Effect

Although the assessment of balance did not lead to contradictory major conclusions, the findings for the major treatment effect were contradictory. Thirteen of the 16 matching schemes demonstrated a significant impact of enrolling in optimal credit hours on retention. The remaining three, nonsignificant matching occurred with the full model, SIS+ESS+NCS, when the conditioning strategy imposed limits on matching (i.e., caliper widths and digit matching). Although it appears that the introduction of critical covariates led to the reversal of this significant finding, this finding needs to be interpreted cautiously due to the sensitivity of the treatment effect.

Just as the inclusion of the ESS and NCS datasets together in the full model led to the reversal of the major finding of significance, the inclusion of more covariates could led to a reversal of the nonsignificant finding. It is important to stress that the covariates included in the model had a low correlation with enrollment in optimal credit hours. Therefore, covariates with a marginal relationship to the selection criterion led to the reversal of this finding. When considering other studies in higher education, the sensitivity of this study is not unusual. Kot (2014) found similar sensitivity when analyzing the impact of academic advising on student success. Kot's study was limited to data from the student information system but found a sensitivity parameter (Γ) parameter of 1.3 which is comparable to the range in this study >1 to 1.5. It is difficult to walk away from the analysis with a definitive answer to the contextual research question but it is evident that both the availability of covariates and the conditioning strategy influence the treatment effect.

Limitations

The first limitation of this study is generalizability. This study used data from a single institution and a single cohort of students. While it is clear that caution must be applied when trying to consider these research findings in a broader context, caution should also be made when generalizing back to the institution and future cohorts of students at that institution. The results were not robust enough to apply them to other cohorts of students, even from the same institution.

Another limitation was the development of the PS model. The development of the PS model relied on a rich set of covariates rather than an established theoretical model. Although this is similar to other research in this area, it is a significant limitation (e.g., An, 2013; Kot, 2014). An essential requirement for PS methods is ignorable treatment assignment. Although the PS model was able to be estimated and fit the data, this information does not ensure that no essential covariates were left out of the modeling process. Further the PS model did not explain much of the variance; therefore, essential covariates were likely left out of the model.

A final limitation of the study was missing data which was impacted by the decision not to impute missing data, the use of 1:1 matching as well as the combination of disparate data sources to estimate the PS model. The decision to use 1:1 matching does not maximize the use of all cases. Therefore, unmatched, eligible cases of students who did not enroll in optimal credit hours were dropped. Additionally, only students that had complete information were retained in the analysis. This decision was complicated by the survey data collection efforts occurring at different points in time. Therefore, not all students participated in each of the data collection efforts.

Practical Implications

This study highlights several implications for practice around covariate selection, PS matching schemes, assessing balance and the sensitivity of the average treatment effect (ATE).

Covariate Selection Matters

This research demonstrates the importance of having a rich set of covariates. First, the expanded covariate set led to a PS model (SIS+NCS+ESS) that accurately classified more students and explained more of the variance in enrolling in optimal credit levels than the other PS models. Additionally, the reversal of the significant impact of optimal credit enrollment on retention in the full model highlights the potential influence of having an expanded covariate set when assessing treatment effects. Although it is difficult to definitively attribute the nonsignificant findings to the addition of key covariates due to issues with missing data, the nonsignificant findings only occurred with the combined full dataset and thus warrants consideration.

The importance of covariate selection raises critical issues for practitioners. Although there is a heavy reliance on data routinely collected by institutions within the student information system (SIS), an expanded variable set will likely lead to better PS models. This means that practitioners need to consider ways to expand their datasets that not only provide a richer covariate set but also provide complete data. This study suffered from missing data due to the separation of survey efforts from the central university processes. It is important that practitioners explore ways to better incorporate critical survey efforts into routine university processes (i.e., applications, embedded questions) to bolster complete data.

Conditioning Strategy Matters

In addition to covariate availability, the conditioning strategy influences sample size, balance and the average treatment effect. When greedy, no caliper width was applied as the conditioning strategy, balance was not achieved between the groups. This conditioning strategy was not capable of creating equivalent groups. In addition, in the full model (SIS+ESS+NCS), the impact of optimal enrollment on retention was significant only when the conditioning strategy was greedy, no caliper width. Although this finding should be disregarded because the groups were not balanced, it does demonstrate the potential implication of conditioning strategies. When restrictions were applied (e.g., caliper width or digit matching), the treatment impact was not significant. The matching scheme in the full model led to different conclusion about treatment impact.

Another implication of the matching scheme was the reduction in sample size. Each of the conditioning strategies that applied restrictions to the match (e.g., caliper width or digit matching), led to the same conclusion regarding the impact of treatment. Considering that the findings were the same across these matching schemes, the reduction to the sample size becomes an issue. Practitioners will need to make decisions about how close the match needs to be. Conditioning strategies that are overly restrictive might not be required; a more relaxed strategy might suffice. In this study, the restrictions imposed on the matches did not lead to clear benefits but did demonstrate costs, sample size reduction.

Balance Assessment Strategy Matters

When assessing covariate balance, the overall conclusions in this study remained consistent across both strategies (standardized mean difference and statistical significance). Despite this, the covariates that were identified as being not balanced differed across the two strategies. The sensitivity that statistical significance demonstrated with PS models conditioned with restrictions on the match (e.g., caliper widths and digit matching) nearly led to disparate findings on balance. It seems prudent for researchers to use both strategies when assessing covariate balance. If the same findings are not reached and statistical significance demonstrated greater sensitivity, examining the effect size could help to determine the importance of the significant covariates and explain the disparate findings.

Sensitivity of the ATE Matters

A final implication for practitioners is that the sensitivity of the ATE must be assessed. It is difficult to state the impact of optimal credit hours on retention in this study. If anything can be said, it is that there is not a consistent, stable nor reliable relationship between enrolling in optimal credit hours on retention for students in this study. Across all of models and matching schemes, the findings were highly sensitive. This sensitivity is underscored by the reversal of the significant impact of optimal credit hour enrollment on retention in the full (SIS+ESS+NCS) model when restrictions were applied to the match. It is important to note that just as easily as the significant finding was reversed, this nonsignificant finding could also be reversed. The inclusion of additional covariates with a highly sensitive ATE can lead to changes in the conclusion. It is important that practitioners assessed sensitivity and do not overstate significant findings when sensitivity is a concern.

Future Research

Future research should focus on the necessary and sufficient qualities when building PS models or, at the very least, reporting the details about the PS models presented. When reviewing the research, the details about how PS models were derived and how they performed was often left out (i.e.., An, 2013). This lack of reporting makes it difficult to discern how robust the current set of covariates is in relation to previous research. Although An (2013) reported the list of covariates eligible for use in a dual enrollment PS, their relationship to dual enrollment was not reported. This information would have helped this current study by identifying other key covariates that are related to enrollment behaviors. Although this is an issue in educational research, the reporting of key features of propensity scores methods is known to be a problem in other fields as well (Ali et al., 2015).

Additionally, the development of PS models could benefit from a mixed method approach, particularly when a strong conceptual model about the selection process has not been established. Conducting focus groups might help elucidate motivations/behaviors associated with the selection process. This can either help guide data collection efforts or, if using extant data, identify potential missing covariates. Since PS methods rely on an ignorable treatment assignment more attention needs to focus on this critical step.

Finally, more research needs to be done on the implications of using different PS approaches in higher education research. Developing a deeper understanding of how these various decision points impact the overall conclusions of research will help inform both research and practice.

APPENDIX A

STUDENT INFORMATION SYSTEM (SIS)

The following definitions are quoted from the IPEDS glossary available at http://nces.ed.gov/ipeds/glossary/ and denoted with * at the end of the term.

ACT, previously known as the American College Testing program, measures educational development and readiness to pursue college-level coursework in English, mathematics, natural science and social studies. Student performance does not reflect innate ability and is influenced by a student's educational preparedness. The ACT composite score is an average of ACT English, ACT mathematics, ACT science and ACT reading. The ACT is used as part of the admission process at this institution.

Academic college, refers to the academic unit in which a student's program of study is administered. Academic college was measured during the first term of students' attendance. Students might have transferred to a new academic program within a different academic college subsequently – this would not bAe reflected in the data. For this institution the following are the academic colleges: Applied Health Sciences, Architecture, Design & the Arts, Business Administration, Education, Engineering and Liberal Arts and Sciences.

Gender, refers to students' self-identification as either male or female. There are no options for students that identify as transgendered or (cis)gender at this institution but students can elect not to respond.

Honors College, refers to a collegiate experience that is in addition to students' academic college. In addition to applying to the university, students in the honors college had to apply and be accepted to the honors college. Students are identified as honors college 'yes' if they enrolled into the honors college during their first term.

High School CPS, identifies students that graduated from a large urban public school system within the boundaries of which the institution serves.

High school GPA, refers to students unweighted high school grade point average. Students' HS GPA is used as part of the admission process in combination with students' standardized test scores.

Placement writing, refers to the entrance exam incoming students take that places them into an appropriate English course. Typically, students are either placed in college ready coursework (English 100 +) or in remedial coursework (English 090s) or below. At times students who are not native English speakers can be placed in English for Speakers of Other Languages coursework (ESL).

Placement math, refers to the entrance exam incoming students take that places them into an appropriate math course. Typically, students are either placed in college ready coursework (Math 100 +) or in remedial coursework (Math 090s) or below.

*Race/ethnicity** refers to the categories developed in 1997 by the Office of Management and Budget (OMB) that are used to describe groups to which individuals belong, identify with, or belong in the eyes of the community. The categories do not denote scientific definitions of anthropological origins. The designations are used to categorize U.S. citizens, resident aliens, and other eligible non-citizens. Individuals are asked to first designate ethnicity as: Hispanic or Latino or Not Hispanic or Latino. Second, individuals are asked to indicate all races that apply among the following: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White.

*American Indian or Alaska Native** refers to a person having origins in any of the original peoples of North America and who maintains cultural identification through tribal affiliation or community recognition.

*Asian** refers to a person having origins in any of the original peoples of the Far East, Southeast Asia or the Indian Subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand and Vietnam.

*Black or African American** refers to a person having origins in any of the black racial groups of Africa.

*Hispanic/Latino** refers to a person of Cuban, Mexican, Puerto Rican, South or Central American or other Spanish culture or origin, regardless of race.

*Native Hawaiian or Other Pacific Islander (NHPI)** refers to a person having origins in any of the original peoples of Hawaii, Guam, Samoa or other Pacific Islands.

*Nonresident alien** refers to a person who is not a citizen or national of the United States who is in this country on a visa or temporary basis and does not have the right to remain indefinitely.

*Race and ethnicity unknown** refers to the category used to report students or employees whose race and ethnicity are not known.

*Resident alien (and other eligible non-citizens)** refers to a person who is not a citizen or national of the United States but who has been admitted as a legal immigrant for the purpose of obtaining permanent resident alien status (and who holds either an alien registration card (Form I-551 or I-151), a Temporary Resident Card (Form I-688), or an Arrival-Departure Record (Form

I-94) with a notation that conveys legal immigrant status such as Section 207 Refugee, Section 208 Asylee, Conditional Entrant Parolee or Cuban-Haitian).

*White** refers to a person having origins in any of the original peoples of Europe, the Middle East or North Africa.

*Pell recipient** (Higher Education Act of 1965, Title IV, Part A, Subpart I, as amended) identifies an undergraduate postsecondary student with demonstrated financial need that has been provided grant assistance to help meet education expenses.

Retention rate refers to a measure of the rate at which students persist in their educational program at an institution, expressed as a percentage. For four-year institutions, this is the percentage of first-time bachelors (or equivalent) degree-seeking undergraduates from the previous fall who are again enrolled in the current fall. For all other institutions this is the percentage of first-time degree/certificate-seeking students from the previous fall who either re-enrolled or successfully completed their program by the current fall.

Summer college, is a summer bridge program offered by the institution to incoming students the summer prior to matriculation. Although any student can become involved with summer college, it is aimed at supporting students that have preparatory placements

APPENDIX B

ENTERING STUDENT SURVEY (ESS)

Marking Instructions

CORRECT MARK

- Use a No. 2 pencil only.
- Do not use ink, ballpoint, or felt tip pens.
- Make solid marks that fill the circle completely.
- Erase cleanly any marks you wish to change.
- Make no stray marks on this form.
- Do not fold, tear, or mutilate this form.

INCORRECT MARKS

Please enter your UIN in the boxes below and fill in the circles corresponding to the numbers in the boxes.

1. Where do you plan to live during the fall semester?

- On campus in a residence hall
- Off campus, not with parents or other relatives, within walking distance
- Off campus, not with parents or other relatives, and commute via car or mass transit
- Off campus with parents or relatives
- Other

2. What is the highest academic degree that you plan to earn at any college?

- None
- Bachelor's (B.A., B.S. etc.)
- Master's (M.A. M.S. etc.)
- Ph.D. or Ed.D.
- M.D., D.O., D.D.S. or D.V.M.
- LL.B. or J.D.(law)
- B.D. or M.Div. (Divinity)
- Other

3. Have you had or do you think you need any special tutoring or help in any of the following?

	Have Had	Will Need
Mathematics		
Science		
Writing		

4. Is English your first language? Yes No

If no, what is your first language?

5. Citizenship Status

- U.S. Citizen
- Permanent Resident (Green card)
- Neither

6. Which best describes your religious affiliation?

- Buddhist
- Hindu
- Jewish
- Muslim
- Protestant Christian
- Roman Catholic
- Other religion
- No affiliation

7. How many Advanced Placement courses or exams did you take in high school?

	None	1-2	3-5	6+
AP Courses				
AP Exams				

8. What is the highest level of education obtained by your parents?

	Father	Mother
No high school diploma/GED	F	M
High School graduate/GED	F	M
Vocational or Trade school (not college)	F	M
Some college (no degree)	F	M
Associate's Degree	F	M
Bachelor's Degree	F	M
Graduate/Professional Degree	F	M
Don't Know	F	M

9. In deciding to go to college how important was each of the following reasons? (mark *one per row*)

	Very Important	Somewhat Important	Not Important
My parents wanted me to go	V	S	N
I couldn't find a job	V	S	N
It is an opportunity to get away from home	V	S	N
To be able to get a better job	V	S	N
To gain a general education and appreciation of ideas	V	S	N
To improve my reading and study skills	V	S	N
To make me more cultured	V	S	N
To be able to make more money	V	S	N
To learn more about things that interest me	V	S	N
To prepare myself for graduate or professional school	V	S	N
A mentor/role model encouraged me to go	V	S	N
To get training for a specific career	V	S	N

10. Rate your self on the following traits as compared with the average person your age. *(mark one per row)*

H=Highest 10% 2=Above Average 3=Average 4=Below average L=Lowest 10%

- Academic Ability
- Artistic Ability
- Competitiveness
- Cooperativeness
- Creativity
- Drive to achieve
- Emotional health
- Initiative
- Leadership ability
- Mathematical ability
- Physical health
- Self Confidence (intellectual)
- Self Confidence (Social)
- Self-understanding
- Spirituality
- Time Management
- Understanding of others
- Writing ability

11. During your last year in high school, how many hours during a typical week did you spend on the following?

N=None 2=less than 2 hours 3=3-5 hours 4=6-10 hours M=More than 10 hours

- Studying/homework
- Socializing with friends
- Talking with teachers outside of class
- Exercise or sports
- Partying
- Working (for pay)
- Volunteer work
- Student clubs/groups
- Watching TV
- Reading for pleasure
- Online social networking (e.g. Facebook)
- Texting/Tweeting
- Prayer/meditation

12. Do you have any concerns about your ability to finance your college education? *(mark one)*

- None - I am confident that I will have sufficient funds
- Some, but I will probably have enough funds
- Major - I am not sure I have enough funds to complete college

13. Below are some reasons that might have influenced your decision to attend UIC. How important was each reason in your decision? *(mark one per row)*

V=Very important S=Somewhat important N=Not important

- My relatives wanted me to come here
- My teacher advised me
- ___ has a good academic reputation
- ___ has a good reputation for its social activities
- ___ has a reputation for racial and ethnic diversity
- I was offered financial assistance
- ___ has low tuition
- High school counselor advised me
- I wanted to live near home
- Not offered aid by my first choice

13. Continued *(mark one per row)*

- ___'s graduates gain admission to top grad/professional schools
- ___ graduates get good jobs
- Not accepted elsewhere
- Rankings in national magazines
- Information from web site
- My friends are attending ___
- I wanted to attend college in a city
- I was admitted to a specific program or major

14. Please indicate the importance to you personally of each of the following. *(mark one in each row)*

E=Essential V=Very important S=Somewhat Important N=Not important

- Becoming accomplished in a performing art
- Becoming an authority in my field
- Obtaining recognition from my colleagues for contributions in my field
- Influencing the political structure
- Influencing social values
- Raising a family
- Having administrative responsibility for the work of others
- Being very well off financially
- Helping others in difficulty
- Writing original works (poems, novels, etc.)
- Becoming successful in my own business
- Being involved in cleaning up the environment
- Developing a meaningful philosophy of life
- Participating in a community action program
- Helping to promote racial understanding
- Keeping up to date with political affairs
- Becoming a community leader

15. What is your best guess as to the chances you will: *(mark one per row)*

V=Very good chance S=Some chance L=Very little chance N=No chance

- Change major fields
- Change career choices
- Graduate with honors
- Participate in student government
- Get a job to help pay for college
- Work full time while in college
- Play varsity/intercollegiate athletics
- Play intramural athletics
- Make at least a "B" average
- Need extra time to complete my degree
- Get a bachelor's degree (B.A. B.S. etc)
- Drop out of ___ temporarily
- Drop out of ___ permanently
- Transfer to another college before graduating
- Be satisfied with college
- Participate in volunteer or community service
- Seek personal counseling
- Develop close friendships with other students
- Communicate regularly with my professors
- Socialize with someone of another race/ethnic group
- Participate in student clubs/groups

Thank you for your participation!

APPENDIX C

NONCOGNITIVE SURVEY (NCS) DATASET

Please provide your name and UIN on the Scantron form before you begin. All of your responses should be bubbled-in on the Scantron form.

About You and Your Family

1. **Where do you live:**
   a. At home with parents/guardians
   b. With other family members such as an older sibling or cousin
   c. Off-campus with a roommate(s)
   d. Off-campus on my own
   e. I live on-campus

2. **How long is your commute to UIC?**
   a. I live on-campus
   b. I live close by (5-10 minutes)
   c. I live somewhat close to campus (11-20 minutes)
   d. I live somewhat far away from campus (21-60 minutes)
   e. I live far from campus (more than 1 hour)

3. **Which of the following describes your current employment in a job (or jobs) outside of school?**
   a. I do not work another job while I'm in school.
   b. I work 1-10 hours per week.
   c. I work between 11 and 20 hours per week.
   d. I work between 21 and 30 hours per week.
   e. I work more than 30 hours per week.

For the follow list of organizations, please indicate if you have previously or are currently receiving support from them. Please answer yes only if you personally have received support.

4. **Bottom Line?**
   a. Yes
   b. No

5. **Goal?**
   a. Yes
   b. No

6. **ICAC?**
   a. Yes
   b. No

7. **Chicago Scholars Program?**
   a. Yes
   b. No

8. **East Village Youth Programs?**
   a. Yes
   b. No

9. **Genesys Works?**
   a. Yes
   b. No

10. Were you born in the U.S.?
a. Yes
b. No

11. Was your Mother born in the U.S.?
a. Yes
b. No

12. Was your Father born in the U.S.?
a. Yes
b. No

13. How many children do you have?
a. None
b. 1
c. 2
d. 3
e. 4 or more

14. Did you grow up in a household where a language other than English was spoken most of the time?
a. Yes
b. No

How often do you do the following things?

	Almost Never	Once in a While	Sometimes	Frequently	Almost Always
15. Translate for your parents	a.	b.	c.	d.	e.
16. Spend time with your cousins, grandparents, aunts, and uncles.	a.	b.	c.	d.	e.
17. Spend time at home with your family.	a.	b.	c.	d.	e.
18. Run errands that the family needs done.	a.	b.	c.	d.	e.
19. Help your brothers or sisters with their homework.	a.	b.	c.	d.	e.
20. Spend holidays with your family.	a.	b.	c.	d.	e.
21. Help out around the house.	a.	b.	c.	d.	e.
22. Spend time with your family on weekends.	a.	b.	c.	d.	e.
23. Help take care of your brothers and sisters.	a.	b.	c.	d.	e.
24. Eat meals with your family.	a.	b.	c.	d.	e.
25. Help take care of your grandparents.	a.	b.	c.	d.	e.
26. Do things together with your brothers and sisters.	a.	b.	c.	d.	e.

About This Class

27. How much time per week do you spend on homework for this class?
a. Less than an hour
b. 1-2 hours
c. 3-4 hours
d. 5-7 hours
e. More than 7 hours

The following statements describe learning strategies that university students may use. Think about the learning strategies you use in *this* class.

	Almost Never	Once in a While	Sometimes	Frequently	Almost Always
28. In this class, I try to determine the best approach for studying each assignment.	a.	b.	c.	d.	e.
29. In this class, I try to monitor my progress when I study.	a.	b.	c.	d.	e.
30. In this class, I make plans for how I will study.	a.	b.	c.	d.	e.
31. In this class, I use different ways to organize my thoughts, such as diagrams, charts, timetables, etc.	a.	b.	c.	d.	e.
32. In this class, I check myself to see how well I am understanding what I am studying.	a.	b.	c.	d.	e.
33. In this class, I focus on understanding the important ideas in what I am reading or studying.	a.	b.	c.	d.	e.
34. In this class, I set goals for myself which I try to accomplish.	a.	b.	c.	d.	e.

The following statements describe an instructor's ability to manage and instruct university students. Please indicate your agreement as it pertains to the instructor of *this* class.

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
35. The instructor really motivates me to do well.	a.	b.	c.	d.	e.
36. I am disappointed in the quality of the instructor.	a.	b.	c.	d.	e.
37. I am doing poorly because the instructor is not effective.	a.	b.	c.	d.	e.
38. What I learn I learn on my own.	a.	b.	c.	d.	e.
39. I would do better if the instructor were better	a.	b.	c.	d.	e.

About You and College

How confident are you that you could complete the following task:

	Not Confident 1	2	3	4	Extremely Confident 5
40. Research a term paper	a.	b.	c.	d.	e.
41. Write course papers	a.	b.	c.	d.	e.
42. Do well on your exams	a.	b.	c.	d.	e.
43. Take good class notes	a.	b.	c.	d.	e.
44. Keep up to date with your school work	a.	b.	c.	d.	e.
45. Manage time effectively	a.	b.	c.	d.	e.
46. Understand your textbooks	a.	b.	c.	d.	e.

How often do you agree with the following statements?

	Never	Occasionally	Often	Very Often
47. I feel a sense of belonging	a.	b.	c.	d.
48. I feel like a member of the campus community	a.	b.	c.	d.
49. I feel comfortable on campus	a.	b.	c.	d.
50. I would choose the same college over again	a.	b.	c.	d.
51. My college is supportive of me	a.	b.	c.	d.

How often do you do the following tasks?

	Never	Occasionally	Often	Very Often
52. Make a list of the things you have to do each day	a.	b.	c.	d.
53. Plan your day before you start it	a.	b.	c.	d.
54. Make a schedule of the activities you have to do on work/school days	a.	b.	c.	d.
55. Write a set of goals for yourself each day	a.	b.	c.	d.
56. Spend time each day planning	a.	b.	c.	d.
57. Have a clear idea of what you want to accomplish during the next week	a.	b.	c.	d.

How true are the following statements?

	Not at all true of me	Not much true of me	Somewhat true of me	Mostly true of me	Very true of me
58. I prefer class work that is challenging so I can learn new things.	a.	b.	c.	d.	e.
59. It is important for me to learn what is being taught in my classes.	a.	b.	c.	d.	e.
60. I like what I am learning in my classes.	a.	b.	c.	d.	e.
61. I think I will be able to use what I learn in this class in other classes.	a.	b.	c.	d.	e.
62. I often choose paper topics I will learn something from even if they require more work.	a.	b.	c.	d.	e.
63. Even when I do poorly on a test I try to learn from my mistakes.	a.	b.	c.	d.	e.
64. I think that what I am learning in my classes is useful for me to know.	a.	b.	c.	d.	e.
65. I think that what we are learning in my classes is interesting.	a.	b.	c.	d.	e.
66. If I run into difficulties in college, I will work harder to overcome them.	a.	b.	c.	d.	e.
67. If my educational opportunities become worse, I will try harder.	a.	b.	c.	d.	e.
68. I will work hard to be successful in college.	a.	b.	c.	d.	e.

The following are a list of things that university students might do. Please indicate whether you have <u>ever</u> done one of the following things in the past year.

	Never	Once	A few	Several times	Many times
69. Copied material and turned it in as your own work	a.	b.	c.	d.	e.
70. Copied a few sentences of material from a published source without giving the author credit	a.	b.	c.	d.	e.
71. Collaborated on an assignment when the instructor asked for individual work	a.	b.	c.	d.	e.
72. Received substantial help on an individual assignment without the teacher/instructor's permission	a.	b.	c.	d.	e.
73. Cheated on a test in any way	a.	b.	c.	d.	e.

Here are a number of statements that may or may not apply to you. Think of how you compare to most people –not just the people you know well, but most people in the world.

	Not at all like me	Not much like me	Somewh at like me	Mostly like me	Very much like me
74. I have achieved a goal that took years of work.	a.	b.	c.	d.	e.
75. I have overcome setbacks to conquer an important challenge.	a.	b.	c.	d.	e.
76. I finish whatever I begin.	a.	b.	c.	d.	e.
77. Setbacks don't discourage me.	a.	b.	c.	d.	e.
78. I am a hard worker.	a.	b.	c.	d.	e.
79. I am diligent.	a.	b.	c.	d.	e.

How well does each of these statements describe you?

	Not Well				Very Well
80. I don't feel sorry for other people when they are having problems.	1	2	3	4	5
81. When I see someone being taken advantage of, I want to help them.	1	2	3	4	5
82. It bothers me when bad things happen to **good** people.	1	2	3	4	5
83. It bothers me when bad things happen to **any** person.	1	2	3	4	5
84. When I see someone being treated unfairly, I don't feel sorry for them.	1	2	3	4	5
85. I feel sorry for other people who don't have what I have.	1	2	3	4	5
86. When I see someone being picked on, I feel sorry for them.	1	2	3	4	5
87. It makes me sad to see a person who doesn't have friends.	1	2	3	4	5
88. When I see another person who is hurt or upset, I feel sorry for them.	1	2	3	4	5

Please indicate your agreement with each item.

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
89. In most ways, my life is close to my ideal.	a.	b.	c.	d.	e.
90. The conditions of my life are excellent.	a.	b.	c.	d.	e.
91. I am satisfied with my life.	a.	b.	c.	d.	e.
92. So far I have gotten the important things I want in life.	a.	b.	c.	d.	e.
93. If I could live my life over, I would change almost nothing.	a.	b.	c.	d.	e.
94. If I stop attending class, no one will notice.	a.	b.	c.	d.	e.
95. Sometimes I feel overwhelmed at the size of the university.	a.	b.	c.	d.	e.
96. I have a hard time knowing who to go to for help at college.	a.	b.	c.	d.	e.
97. I have found it easy to navigate the campus.	a.	b.	c.	d.	e.
98. If I have a question about college, I know who to talk to for help.	a.	b.	c.	d.	e.

Survey of Non-Cognitive Factors Influencing Student Success

NO	Factor	Reference for Scale Items	Scale Name
1 – 14, 27	Demographic Factors	Commute Time Housing Choice Employment Older Sibling as Mentor Born in U.S. Parents Born in U.S. Parenting Class Attendance Language Background	
15- 26	Attitudes Toward Family Obligations	Fuligni, A. J., Tseng, V., & Lam, M. (1999). Attitudes toward family obligations among American adolescents with Asian, Latin American, and European backgrounds. *Child Development*, 70, 1030-1044.	11-items Family Obligations – Current Assistance Subscale*
28- 34	Self-Regulated Learning	Shell, D. F., Husman, J., Turner, J. E., Cliffel, D. M., Nath, I., & Sweany, N. (2005). The impact of computer supported collaborative learning communities on high school student's knowledge building strategic leaning, and perceptions of the classroom. *Journal of Educational Computing Research*, *33*(3), 327-349.	7-items Student Perception of Classroom Knowledge Building Scale (SPOCK)-- Self-Regulated Strategy Use subscale
35- 39	Perceived Efficacy of the Instructor	Prevatt, F., Li, H., Welles, T., Festa-Dreher, D., Yelland, S., & Lee, J. (2011). The Academic Success Inventory for College Students: Scale Development and Practical Implications for Use with Students. *Journal of College Admission*, *211*, 26-31.	5-items Academic Success Inventory for College Students (ASICS)--Perceived Instructor Efficacy subscale

40-46	Perceived Self-efficacy	Solberg, V. S., O'Brien, K., Villareal, P., Kennel, R., & Davis, B. (1993). Self-efficacy and Hispanic college students: Validation of the college self-efficacy instrument. *Hispanic Journal of Behavioral Sciences, 15*, 80-95.	7-items College Self-efficacy – Course Efficacy Subscale*
47-51	Perceived Sense of Belonging	Johnson, D. R., Soldner, M., Leonard, J. B., Alvarez, P., Inkelas, K. K., Rowan-Kenyon, H. T., & Longerbeam, S. D. (2007). Examining sense of belonging among first-year undergraduates from different racial/ethnic groups. *Journal of College Student Development*, 48, 525-542.	5-items Overall Sense of Belonging Subscale
52-57	Time Management	Britton, B. K., & Tesser, A. (1991). Effects of time-management practices on college grades. *Journal of Educational Psychology*, 83, 405-410.	6-items Time Management - Short-Range Planning Subscale*
58-65	Academic Motivation	Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33-40.	9-items Academic Motivation - Intrinsic Value Subscale*
66-68	Academic Control Striving Behaviors	Heckhausen, J., Wrosch, C., & Schulz, R. (2010). A motivational theory of life-span development. *Psychological review, 117*(1), 32.	3-items Optimization in Primary and Secondary Control (OPS)– Goal Engagement: Primary Selective Control subscale*
69-73	Problem Behaviors	McCabe, D. L., & Trevino, L. K. (1997). Individual and contextual influences on academic dishonesty: A multicampus investigation. *Research in Higher Education, 38*(3), 379-396.	5-items Academic Dishonesty scale*
74-79	Grit	Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92, 1087-1101.	6-items Grit - Perseverance of Effort Subscale*

80-88	Caring	Lerner, R. M., Lerner, J. V., Almerigi, J. B., Theokas, C., Phelps, E., Gestsdottir, S.,& von Eye, A. (2005). Positive Youth Development, Participation in community youth development programs, and community contributions of fifth-grade adolescents findings from the first wave Of the 4-H study of Positive Youth Development. *The Journal of Early Adolescence*, *25*(1), 17-71.	9-item scale Positive Youth Development scale—Caring subscale
89-93	Subjective Well-Being	Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71-75.	5-items Satisfaction with Life Scale (SWLS)*
94-98	Feeling Lost in the System	Farruggia, S.& Watson, L. (2014) Measure made for purpose of study	
99 100	Social Status	Goodman, E., Adler, N. E., Kawachi, I., Frazier, A. L., Huang, B., & Colditz, G. A. (2001). Adolescents' perceptions of social status: development and evaluation of a new indicator. Pediatrics, 108(2), e31-e31.	2-items Subjective Social Status Scale

*A modified version of the original survey was used.

APPENDIX D

SAS CODE

Step One: Descriptive Analysis

```
/*dissertation*/
libname diss "C:\Users\jdwren\Desktop\Dissertation";

PROC IMPORT
DATAFILE="C:\Users\jdwren\Desktop\Dissertation\Julie_UIC_Data_160503.xls"
OUT=diss.base
DBMS=xls REPLACE;
RUN;

DATA DISS.BASE_WF;
SET DISS.BASE (rename=(instructor2=instructor2r instructor3=instructor3r
instructor4=instructor4r instructor5=instructor5r
                        CARING1=CARING1R CARING5=CARING5R LOST4=LOST4R
LOST5=LOST5R));

/*ADJUST FOR REVERSE CODING*/
%MACRO VAR(VAR);m
&VAR=5-&VAR.R;
%MEND VAR;
%VAR (INSTRUCTOR2);
%VAR (INSTRUCTOR3);
%VAR (INSTRUCTOR4);
%VAR (INSTRUCTOR5);
%VAR (CARING1);
%VAR (CARING5);
%VAR (LOST4);
%VAR (LOST5);


IF CREDATTEMPT220148 >= 12;/*KEEP ONLY FULLTIME STUDENTS*/
IF CREDATTEMPT220148 <15 THEN F1_15 = 0; ELSE F1_15=1;
IF CREDATTEMPT220158 >=1 THEN F2_REG = 1; ELSE F2_REG = 0;

/* SCALE SCORES OF NCS VARIABLES*/
IF NMISS(of selfeff1-selfeff7) > 0 THEN selfeff_total = . ; ELSE
selfeff_total = SUM(of selfeff1-selfeff7);
IF NMISS(of TimeManage1-TimeManage6) > 0 THEN TimeManage_total = . ; ELSE
TimeManage_total = SUM(of TimeManage1-TimeManage6);
IF NMISS(of Belong1-Belong5) > 0 THEN Belong_total = . ; ELSE Belong_total =
SUM(of Belong1-Belong5);
IF NMISS(of swb1-swb5) > 0 THEN swb_total = . ; ELSE swb_total = SUM(of swb1-
swb5);
IF NMISS(of Motiv1-motiv8) > 0 THEN motiv_total = . ; ELSE motiv_total =
SUM(of motiv1-motiv8);
IF NMISS(of FamilyOb1-familyob12) > 0 THEN familyob_total = . ; ELSE
familyob_total = SUM(of familyob1-familyob12);
IF NMISS(of Grit1-Grit6) > 0 THEN grit_total = . ; ELSE grit_total = SUM(of
grit1-grit6);
```

```
IF NMISS(of srl1-srl7) > 0 THEN srl_total = . ; ELSE srl_total = SUM(of srl1-
srl7);
IF NMISS(of instructor1-instructor5) > 0 THEN instructor_total = . ; ELSE
instructor_total = SUM(of instructor1-instructor5);
IF NMISS(of academiccontrol1-academiccontrol3) > 0 THEN academiccontrol_total
= . ; ELSE academiccontrol_total = SUM(of academiccontrol1-academiccontrol3);
IF NMISS(of cheating1-cheating5) > 0 THEN cheating_total = . ; ELSE
cheating_total = SUM(of cheating1-cheating5);
IF NMISS(of caring1-caring9) > 0 THEN caring_total = . ; ELSE caring_total =
SUM(of caring1-caring9);
IF NMISS(of lost1-lost5) > 0 THEN lost_total = . ; ELSE lost_total = SUM(of
lost1-lost5);

/*recoding variables*/
/*SIS VARIABLES*/
IF PLACEMENTWRITING = 'ESL 060' THEN WRITING_RANK = 1;
            ELSE IF PLACEMENTWRITING = 'ENGL 070' THEN WRITING_RANK = 2;
            ELSE IF PLACEMENTWRITING = 'ENGL 071' THEN WRITING_RANK = 3;
            ELSE IF PLACEMENTWRITING = 'ENGL 160' THEN WRITING_RANK = 4;
            ELSE IF PLACEMENTWRITING = 'ENGL 161' THEN WRITING_RANK = 5;

      IF PLACEMENTMATH = 'Math 075' THEN MATH_RANK = 1;
            ELSE IF PLACEMENTMATH = 'Math 090' THEN MATH_RANK = 2;
            ELSE IF PLACEMENTMATH = 'MATH 121, 160, 165 and STAT 101' THEN
MATH_RANK = 3;
            ELSE IF PLACEMENTMATH = 'MATH 180 and STAT 130' THEN MATH_RANK =
4;

/*ESS VARIABLES*/
      IF ESS_LIVE IN (3,5) THEN ESS_LIVER = 2; ELSE ESS_LIVER = ESS_LIVE;
/*OFF CAMPUS*/
      IF ESS_DEGREE IN (7,8) THEN ESS_DEGREER = 9; ELSE ESS_DEGREER = ESS_
DEGREE; /*OTHER*/


run;


/*SIGNIFICANT DIFFERENCE BETWEEN GROUPS ON RETENTION -
RETENTION IS LOWER AMONG INDIVIDUALS WHO DO NOT ENROLL IN 15
CREDITS DURING THEIR FIRST TERM*/

/*GROUPING VARIABLE AND OUTCOME VARIABLE*/
PROC FREQ DATA = DISS.BASE_WF;
TABLE F1_15 * F2_REG /chisq measures
plots=(freqplot(twoway=groupvertical scale=percent));
RUN;

/*DESCRIPTIVES*/

/*Expected Cell Size Considerations
The validity of the chi-square test depends on both the sample
size and
```

```
the number of cells. Several rules of thumb have been suggested to indicate
whether the chi-square approximation is satisfactory. One such rule
sug-
gested by Cochran (1954) says that the approximation is adequate if
no
expected cell frequencies are less than one and no more than 20% are less
than five.*/

proc sort data=diss.base_wf;
by f1_15;
run;

proc freq data = diss.base_wf;
tables F1_15 * (ETHNIC GENDER HONCOLL PELL HSCPS SUMMCOLL FGENCOLLNEW
PLACEMENTWRITING PLACEMENTMATH
            /*ESS*/ ESS LIVE /*RECODE 20% RULE*/ ESS LIVER ESS degree ESS
DEGREER /*RECODE 20% RULE*/ ESS mathhaD ESS mathneed ESS scihad ESS scineed
ESS writehad ESS writewiL
            ESS lang ESS religion ESS apcourse ESS apexam)/MISSING;
run;

proc univariate data = DISS.BASE_WF;
      var /*SIS*//* FYAGE SACTE SACTM SHSGPAR
            /*ess*/ /*Q91     Q92   Q93   Q94   Q95   Q96   Q97   Q98   Q99
      Q910  Q911  Q912  Q101  Q102  Q103  Q104  Q105  Q106  Q107
                        Q108  Q109  Q1010 Q1011 Q1012 Q1013 Q1014 Q1015 Q1016
      Q1017 Q1018 Q111  Q112  Q113  Q114
                        Q115  Q116  Q117  Q118  Q119  Q1110 Q1111 Q1112 Q1113
      Q12   Q131  Q132  Q133  Q134  Q135
                        Q136  Q137  Q138  Q139  Q1310 Q1311 Q1312 Q1313 Q1314
      Q1315 Q1316 Q1317 Q1318 Q141  Q142
                        Q143  Q144  Q145  Q146  Q147  Q148  Q149  Q1410 Q1411
      Q1412 Q1413 Q1414 Q1415 Q1416 Q1417
                        Q151  Q152  Q153  Q154  Q155  Q156  Q157  Q158  Q159
      Q1510 Q1511 Q1512 Q1513 Q1514 Q1515
                        Q1516 Q1517 Q1518 Q1519 Q1520 Q152
            /*NCS*/ selfeff1 selfeff2 selfeff3 selfeff4 selfeff5 selfeff6
selfeff7
                        TimeManage1 TimeManage2 TimeManage3 TimeManage4
      TimeManage5 TimeManage6
                        SWB1  SWB2  SWB3  SWB4  SWB5
                        FamilyOb1   FamilyOb2   FamilyOb3   FamilyOb4
      FamilyOb5   FamilyOb6   FamilyOb7   FamilyOb8   FamilyOb9   FamilyOb10
      FamilyOb11  FamilyOb12
                        Grit1 Grit2 Grit3 Grit4 Grit5 Grit6
                        AcademicControl1  AcademicControl2  AcademicControl
                        CARING1     CARING2     CARING3     CARING4
      CARING5     CARING6     CARING7     CARING8     CARING9;
      BY F1_15;
RUN;

/*internal consistency of scales - decision to use scales except for caring*/
ods graphics on;
```

```
%macro corr (corr);
proc corr data=diss.base_wf nomiss nocorr alpha plots;
   var &corr;
 run;
%mend corr;
%corr (selfeff1 selfeff2 selfeff3 selfeff4 selfeff5 selfeff6 selfeff7);
%corr (TimeManage1      TimeManage2 TimeManage3 TimeManage4 TimeManage5
      TimeManage6 );
%corr (SWB1 SWB2  SWB3  SWB4  SWB5);
%corr (FamilyOb1  FamilyOb2   FamilyOb3   FamilyOb4   FamilyOb5   FamilyOb6
      FamilyOb7   FamilyOb8   FamilyOb9   FamilyOb10  FamilyOb11
      FamilyOb12);
%corr (Grit1      Grit2 Grit3 Grit4 Grit5 Grit6);
%corr (AcademicControl1 AcademicControl2  AcademicControl3);
%corr (CARING1    CARING2     CARING3     CARING4     CARING5     CARING6
      CARING7     CARING8     CARING9);

/*CORRELATIONS*/
/*interval_dichotmous data*/
PROC CORR data=DISS.BASE_WF OUTP=DISS.BASE_CORR;
VARIABLE F1_15
            /*SIS VARIABLES*/
            GENDER HONCOLL PELL HSCPS SUMMCOLL SACTC SACTE SACTM SHSGPAR
            /*ESS VARIABLES*/
            Q91 Q92 Q93 Q94 Q95 Q96 Q97 Q98 Q99 Q910 Q911 Q912 Q101 Q102 Q103
Q104 Q105 Q106
            Q107 Q108 Q109 Q1010 Q1011 Q1012 Q1013 Q1014 Q1015 Q1016 Q1017
Q1018 Q111 Q112 Q113
            Q114 Q115 Q116 Q117 Q118 Q119 Q1110 Q1111 Q1112 Q1113 Q12 Q131
Q132 Q133 Q134 Q135
            Q136 Q137 Q138 Q139 Q1310 Q1311 Q1312 Q1313 Q1314 Q1315 Q1316
Q1317 Q1318 Q141 Q142
            Q143 Q144 Q145 Q146 Q147 Q148 Q149 Q1410 Q1411 Q1412 Q1413 Q1414
Q1415 Q1416 Q1417
            Q151 Q152 Q153 Q154 Q155 Q156 Q157 Q158 Q159 Q1510 Q1511 Q1512
Q1513 Q1514 Q1515 Q1516
            Q1517 Q1518 Q1519 Q1520 Q1521
            /*NCS VARIABLES*/
            SelfEff_Total TimeManage_total swb_total familyob_total
grit_total academiccontrol_total caring_total
            CARING1     CARING2     CARING3     CARING4     CARING5
      CARING6     CARING7     CARING8     CARING9;
RUN;
/*tested NCS correlations for items - not any better than the scale thus
maintained the scale*/

/*correlations categorical*/
%LET VAR = (GENDER HONCOLL PELL HSCPS SUMMCOLL PLACEMENTWRITING PLACEMENTMATH
ESS mathhad ESS mathneed ESS scihad ESS scineed ESS writehad ESS writewil ESS
lang ESS apcourse ESS apexam);
PROC FREQ DATA = DISS.BASE_WF;
TABLE &VAR * (GENDER HONCOLL PELL HSCPS SUMMCOLL FGENCOLLNEW PLACEMENTWRITING
PLACEMENTMATH
```

```
ESS mathhad ESS mathneed ESS scihad ESS scineed ESS writehad ESS writewil ESS
lang ESS apcourse ESS apexam )/CHISQ;
RUN;
 /*DROP SACTC ESS APEXAM*/

/*BUILD LOGISTIC REGRESSION MODEL FOR GROUPING VARIABLE - ENROLLING IN 15+*/
title 'Logistic Regression on Optimal Credit Enrollment';
   proc logistic data=DISS.BASE_WF outest=betas covout;
        class
            /*SIS VARIABLES*/
            GENDER ETHNIC (REF='White') HONCOLL PELL COLLEGE (REF = 'Liberal
Arts & Sciences') SUMMCOLL
            HSCPS PLACEMENTMATH(REF='MATH 121, 160, 165 and STAT
101')PLACEMENTWRITING (REF = 'ENGL 160')
            /*ESS VARIABLES*/
            ESS liver (REF='1') ESS degreer (REF='3') ESS lang (REF='1') ESS
religion (REF='8') ESS apcourse (REF='1')
            ESS mathhad ESS mathneed ESS scihad ESS scineed ESS writehad ESS
writewil/ param=ref ref=last;

      model F1_15(event='1')=
            /*SIS VARIABLES*/
            GENDER ETHNIC HONCOLL PELL COLLEGE HSCPS SUMMCOLL PLACEMENTMATH
PLACEMENTWRITING
            SACTE SACTM SHSGPAr
            /*ESS VARIABLES*/
            ESS liver ESS degreer ESS mathhad ESS mathneed ESS scihad ESS
scineed ESS writehad ESS writewil
            ESS lang ESS religion ESS apcourse
            Q101 Q106 Q108 Q1010 Q111 Q115 Q137 Q1311 Q149 Q1417 Q153 Q156
Q157 Q159 Q1511 Q1516 Q1520 Q1521
            /*NCS VARIABLES*/
            SelfEff_Total TimeManage_total swb_total familyob_total
grit_total academiccontrol_total
            CARING1     CARING2     CARING3     CARING4     CARING5
      CARING6     CARING7     CARING8     CARING9
             / lackfit rsquare;
   run;
```

Step Two: Estimate Propensity Score (SIS Model)

```
/*STEP TWO ESTIMATE THE PROPENSITY SCORE*/
/*SIS MODEL*/

/*
https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer
.htm#statug_logistic_sect052.htm*/

/*MAKE SIS DATASET*/
DATA DISS.S2_SISMODEL;
```

```
      SET DISS.BASE_WF (KEEP=euin F2_REG F1_15 GENDER ETHNIC HONCOLL PELL
COLLEGE PLACEMENTWRITING PLACEMENTMATH SUMMCOLL SACTE SHSGPAR SACTM HSCPS) ;
             if nmiss(of _NUMERIC_)=0;
             if cmiss(of _ALL_)=0;
RUN;

/*MULTICOLLINEARITY*/
PROC REG DATA=DISS.S2_SISMODEL;
      MODEL F1_15 = SACTE SACTM SHSGPAR /*PLACEMENTWRITING PLACEMENTMATH*/
GENDER /*ETHNIC*/ SUMMCOLL HONCOLL HSCPS PELL/ TOL VIF COLLIN;
RUN;

/*DESCRIPTIVES*/
PROC FREQ data = diss.S2_sismodel;
      tables f1_15 * (GENDER ETHNIC HONCOLL HSCPS PELL COLLEGE
PLACEMENTWRITING PLACEMENTMATH SUMMCOLL);
RUN;

PROC SORT DATA = DISS.S2_SISMODEL;
      BY F1_15;
RUN;

PROC MEANS DATA = DISS.S2_SISMODEL MEAN STD;
      VAR  SACTE SHSGPAR SACTM;
      BY F1_15;
RUN;


/*SIG TESTING - CHECKED FOR INTERACTIONS*/
title 'Logistic Regression on Optimal Credit Enrollment';
   proc logistic data=DISS.S2_SISMODEL outest=betas covout;
        class GENDER ETHNIC (REF='White') HONCOLL PELL COLLEGE (REF =
'Liberal Arts & Sciences') SUMMCOLL HSCPS
            PLACEMENTWRITING (REF = 'ENGL 160') PLACEMENTMATH(REF='MATH 121,
160, 165 and STAT 101') / param=ref ref=first;
      model F1_15(event='1')=GENDER ETHNIC HONCOLL PELL HSCPS COLLEGE
PLACEMENTWRITING PLACEMENTMATH  SUMMCOLL
            SACTE SHSGPAR SACTM

            /*GENDER| ETHNIC| HONCOLL| PELL| HSCPS| COLLEGE|
PLACEMENTWRITING| PLACEMENTMATH|  SUMMCOLL|
            SACTE| SHSGPAR| SACTM @ 2 - INTERACTIONS NOT SIGNIFICANT*/
                   / lackfit
                             rsquare;
      output out=diss.S2_sismodel_pred prob=prob lower=lcl upper=ucl
prob=prob
             predprob=(individual crossvalidate);
   run;
```

Step Two: Estimate Propensity Score (SIS + ESS Model)

```
/*STEP TWO ESTIMATE THE PROPENSITY SCORE*/
```

```
/*SIS + ESS MODEL*/

/*MAKE SIS_ESS DATASET*/
DATA DISS.S2_SISESS MODEL;
      SET DISS.BASE_WF (KEEP=EUIN F1_15 F2_REG GENDER ETHNIC HONCOLL HSCPS
PELL COLLEGE PLACEMENTWRITING PLACEMENTMATH SUMMCOLL SHSGPAR SACTM SACTE
      ESS LIVER ESS DEGREER ESS mathhaD ESS mathneed ESS scihad ESS scineed
ESS writehad ESS writewiL ESS lang ESS religion ESS apcourse
      Q106 Q108 Q111 Q153 Q156 Q157 Q159 Q1511 Q1516 Q1520 Q1521);
             if nmiss(of _NUMERIC_)=0;
             if cmiss(of _ALL_)=0;
RUN;

/*MULTICOLLINEARITY*/
PROC REG DATA=DISS.S2_SISESS MODEL;
      MODEL F1_15 = GENDER /*ETHNIC*/ HONCOLL HSCPS PELL /*COLLEGE*/ SUMMCOLL
/*PLACEMENTWRITING PLACEMENTMATH*/ SACTE SHSGPAR SACTM
      ESS LIVER ESS DEGREER ESS mathhaD ESS mathneed ESS scihad ESS scineed
ESS writehad ESS writewiL ESS lang ESS religion ESS APCOURSE
      Q106 Q108 Q111 Q153 Q156 Q157 Q159 Q1511 Q1516 Q1520 Q1521/ VIF TOL
COLLIN;
RUN;


/*FINAL DESCRIPTIVES*/
PROC FREQ data = DISS.S2_SISESS MODEL;
      tables (GENDER ETHNIC HONCOLL HSCPS PELL COLLEGE PLACEMENTWRITING
PLACEMENTMATH SUMMCOLL
                  ESS LIVER ESS DEGREER ESS mathhaD ESS mathneed ESS scihad
ESS scineed ESS writehad
                  ESS writewiL ESS lang ESS religion ESS APCOURSE)*f1_15;
RUN;

proc means data=DISS.S2_SISESS MODEL mean STD;
CLASS F1_15;
var SACTE SHSGPAR SACTM Q106 Q108 Q111 Q153 Q156 Q157 Q159 Q1511 Q1516 Q1520
Q1521;
run;


title 'Logistic Regression on Optimal Credit Enrollment';
   proc logistic data=DISS.S2_SISESS MODEL outest=betas covout;
        class
            /*SIS VARIABLES*/
            GENDER ETHNIC (REF='White') HONCOLL PELL COLLEGE (REF = 'Liberal
Arts & Sciences') SUMMCOLL
            HSCPS PLACEMENTMATH(REF='MATH 121, 160, 165 and STAT
101')PLACEMENTWRITING (REF = 'ENGL 160')
            /*ESS VARIABLES*/
            ESS liver (REF='1') ESS degreer (REF='3') ESS lang (REF='1') ESS
religion (REF='8') ESS apcourse (REF='1')
            ESS mathhad ESS mathneed ESS scihad ESS scineed ESS writehad ESS
writewil/ param=ref ref=last;
```

```
      model F1_15(event='1')=
            /*SIS VARIABLES*/
            GENDER ETHNIC HONCOLL PELL COLLEGE HSCPS SUMMCOLL PLACEMENTMATH
PLACEMENTWRITING
            SACTE SACTM SHSGPAr
            /*ESS VARIABLES*/
            ESS liver ESS degreer ESS mathhad ESS mathneed ESS scihad ESS
scineed ESS writehad ESS writewil
            ESS lang ESS religion ESS apcourse
            Q106 Q108 Q111 Q153 Q156 Q157 Q159 Q1511 Q1516 Q1520 Q1521

             / lackfit rsquare;
      output out=DISS.S2_SISESS MODEL_PRED prob=prob lower=lcl upper=ucl
prob=prob
             predprob=(individual crossvalidate);
   run;
```

Step Two: Estimate Propensity Score (SIS + NCS Model)

```
/*STEP TWO ESTIMATE THE PROPENSITY SCORE*/
/*SIS + NCS MODEL*/

/*
https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer
.htm#statug_logistic_sect052.htm*/

/*MAKE SIS_NCS DATASET*/
DATA DISS.S2_SISNCS_MODEL;
      SET DISS.BASE_WF (KEEP=EUIN F2_REG F1_15 EUIN F2_REG F1_15 GENDER
ETHNIC HONCOLL COLLEGE PELL
            PLACEMENTWRITING PLACEMENTMATH HSCPS SUMMCOLL SACTE SHSGPAR SACTM
            SelfEff_Total TimeManage_total swb_total familyob_total
grit_total academiccontrol_total
            CARING1     CARING2     CARING3     CARING4     CARING5
      CARING6     CARING7     CARING8     CARING9);
             if nmiss(of _NUMERIC_)=0;
             if cmiss(of _ALL_)=0;
RUN;

/*MULTICOLLINEARITY*/
PROC REG DATA=DISS.S2_SISNCS_MODEL;
      MODEL F1_15 =  GENDER /*ETHNIC*/ HONCOLL PELL HSCPS SUMMCOLL
/*PLACEMENTWRITING PLACEMENTMATH*/ SACTE SACTM  SHSGPAr
      SelfEff_Total TimeManage_total swb_total familyob_total grit_total
academiccontrol_total
      CARING1     CARING2     CARING3     CARING4     CARING5     CARING6
      CARING7     CARING8     CARING9 /VIF COLLIN;
RUN;

/*FINAL DESCRIPTIVES*/
```

```
PROC FREQ data = DISS.S2_SISNCS_MODEL;
      tables f1_15 * ( GENDER ETHNIC HONCOLL COLLEGE PELL PLACEMENTWRITING
PLACEMENTMATH HSCPS SUMMCOLL );
RUN;

proc means data=DISS.S2_SISNCS_MODEL mean STD;
CLASS F1_15;
var SelfEff_Total TimeManage_total swb_total familyob_total grit_total
academiccontrol_total
      CARING1     CARING2     CARING3     CARING4     CARING5     CARING6
      CARING7     CARING8     CARING9;
run;


title 'Logistic Regression on Optimal Credit Enrollment';
   proc logistic data=DISS.S2_SISNCS_MODEL outest=betas covout;
        class
            /*SIS VARIABLES*/
            GENDER ETHNIC (REF='White') HONCOLL PELL COLLEGE (REF = 'Liberal
Arts & Sciences') SUMMCOLL
            HSCPS PLACEMENTMATH(REF='MATH 121, 160, 165 and STAT
101')PLACEMENTWRITING (REF = 'ENGL 160')/ param=ref ref=last;

      model F1_15(event='1')=
            /*SIS VARIABLES*/
            GENDER ETHNIC HONCOLL PELL COLLEGE HSCPS SUMMCOLL PLACEMENTMATH
PLACEMENTWRITING
            SACTE SACTM SHSGPAr

            /*NCS VARIABLES*/
            SelfEff_Total TimeManage_total swb_total familyob_total
grit_total academiccontrol_total
            CARING1     CARING2     CARING3     CARING4     CARING5
      CARING6     CARING7     CARING8     CARING9
                   / lackfit
                               rsquare;
      output out=diss.S2_SISNCS_model_pred prob=prob lower=lcl upper=ucl
prob=prob
             predprob=(individual crossvalidate);
   run;
```

Step Two: Estimate Propensity Score (SIS + NCS + ESS Model)

```
/*STEP TWO ESTIMATE THE PROPENSITY SCORE*/
/*SIS + ESS + NCS MODEL*/

/*
https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer
.htm#statug_logistic_sect052.htm*/

/*MAKE SIS_ESS NCS_ DATASET*/
DATA DISS.s2_SISESSNCS_MODEL;
```

```
      SET DISS.BASE_WF (KEEP=EUIN F1_15 F2_REG GENDER ETHNIC HONCOLL HSCPS
PELL COLLEGE SUMMCOLL FULL_WRITING PLACEMENTMATH SACTE SHSGPAR SACTM
      ESS LIVER ESS DEGREER ESS mathhaD ESS mathneed ESS scihad ESS scineed
ESS writehad ESS writewiL ESS lang ESS religion ESS APCOURSE
      Q106 Q108 Q111 Q153 Q156 Q157 Q159 Q1511 Q1516 Q1520 Q1521
SelfEff_Total TimeManage_total swb_total familyob_total grit_total
      academiccontrol_total CARING1 CARING2     CARING3     CARING4
      CARING5     CARING6     CARING7     CARING8     CARING9 );
             if nmiss(of _NUMERIC_)=0;
             if cmiss(of _ALL_)=0;
RUN;

/*MULTICOLLINEARITY*/
PROC REG DATA=DISS.s2_SISESSNCS_MODEL;
      MODEL  F1_15 = GENDER /*ETHNIC*/ HONCOLL HSCPS PELL /*COLLEGE*/
SUMMCOLL /*FULL_WRITING PLACEMENTMATH*/ SACTE SHSGPAR SACTM
      ESS LIVER ESS DEGREER ESS mathhaD ESS mathneed ESS scihad ESS scineed
ESS writehad ESS writewiL ESS lang ESS religion ESS APCOURSE
      Q106 Q108 Q111 Q153 Q156 Q157 Q159 Q1511 Q1516 Q1520 Q1521
SelfEff_Total TimeManage_total swb_total familyob_total grit_total
      academiccontrol_total CARING1 CARING2     CARING3     CARING4
      CARING5     CARING6     CARING7     CARING8     CARING9/ VIF TOL
COLLIN;
RUN;

/*FINAL DESCRIPTIVES*/
PROC FREQ data = DISS.S2_SISESSNCS_MODEL;
      tables (GENDER ETHNIC HONCOLL HSCPS PELL COLLEGE FULL_WRITING
PLACEMENTMATH SUMMCOLL
                              ESS LIVER ESS DEGREER ESS mathhaD ESS mathneed
ESS scihad ESS scineed ESS writehad
                              ESS writewiL ESS lang ESS religion ESS APCOURSE
) *f1_15;
RUN;

proc means data=DISS.S2_SISESSNCS_MODEL mean STD;
CLASS F1_15;
var SACTE SHSGPAR SACTM ESS APCOURSE
      Q106 Q108 Q111 Q153 Q156 Q157 Q159 Q1511 Q1516 Q1520 Q1521
      SelfEff_Total TimeManage_total swb_total familyob_total grit_total
academiccontrol_total
      CARING1     CARING2     CARING3     CARING4     CARING5     CARING6
      CARING7     CARING8     CARING9 ;
run;


title 'Logistic Regression on Optimal Credit Enrollment';
   proc logistic data=DISS.S2_SISESSNCS_MODEL outest=betas covout;
        class
            /*SIS VARIABLES*/
            GENDER ETHNIC (REF='White') HONCOLL PELL COLLEGE (REF = 'Liberal
Arts & Sciences') SUMMCOLL
```

```
            HSCPS PLACEMENTMATH(REF='MATH 121, 160, 165 and STAT
101')FULL_WRITING (REF = 'ENGL 160')
            /*ESS VARIABLES*/
            ESS liver (REF='1') ESS degreer (REF='3') ESS lang (REF='1') ESS
religion (REF='8') ESS apcourse (REF='1')
            ESS mathhad ESS mathneed ESS scihad ESS scineed ESS writehad ESS
writewil/ param=ref ref=last;

      model F1_15(event='1')=
            /*SIS VARIABLES*/
            GENDER ETHNIC HONCOLL PELL COLLEGE HSCPS SUMMCOLL FULL_WRITING
PLACEMENTMATH
            SACTE SACTM SHSGPAr
            /*ESS VARIABLES*/
            ESS liver ESS degreer ESS mathhad ESS mathneed ESS scihad ESS
scineed ESS writehad ESS writewil
            ESS lang ESS religion ESS apcourse
            Q106 Q108 Q111 Q153 Q156 Q157 Q159 Q1511 Q1516 Q1520 Q1521
            /*NCS VARIABLES*/
            SelfEff_Total TimeManage_total swb_total familyob_total
grit_total academiccontrol_total
            CARING1     CARING2     CARING3     CARING4     CARING5
      CARING6     CARING7     CARING8     CARING9
             / lackfit
                  rsquare;
      output out=diss.S2_SISESSNCS_model_pred prob=prob lower=lcl upper=ucl
prob=prob
             predprob=(individual crossvalidate);
run;
```

Step Three: Assess Region of Common Support

```
/*STEP THREE - ASSESS THE REGION OF COMMON SUPPORT*/
/*
https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer
.htm#statug_logistic_sect052.htm*/

%MACRO CAT(FILE);
/*http://www.basug.org/downloads/2011q3/Scott.pdf*/

proc sort data=&file;
by f1_15;
run;

proc univariate data= &FILE plot;
title 'Histograms of Propensity Scores by Treatment Group';
var prob;
class F1_15;
histogram prob / ctext=purple cfill=blue
kernel (k=normal color=green w=3 l=1)
normal (color = red w=3 l=2)
ncols= 1
nrows= 2;
inset n='N' (comma6.0) mean='Mean' (6.2)
```

```
median='Median' (6.2)
mode='Mode'(6.2)
normal kernel(type) /
position=NW;
run;

proc boxplot data=&file;
symbol width = 2;
plot prob*f1_15 /  cboxes=black  cframe = white  idsymbol= circle  idcolor=
black
font='times new roman'
height=3.5 boxwidth=6
boxstyle=schematic
waxis= 2;
run;

%MEND CAT;
%CAT (diss.s2_sismodel_pred);
%CAT (diss.s2_sisESS model_pred);
%CAT (diss.s2_sisncs_model_pred);
%CAT (diss.s2_sisessncs_model_pred);

/*trim data set*/

%macro cat (file, nfile, lval, hval);
data &nfile;
      set &file;
      if prob > &lval;
      if prob < &hval;
run;
%mend cat;
%CAT (diss.s2_sismodel_pred, diss.s2_sismodel_predt,0.325470,0.935863);
%CAT (diss.s2_sisESS model_pred, diss.s2_sisESS
model_predt,0.1657254,0.960410);
%CAT (diss.s2_sisncs_model_pred,
diss.s2_sisncs_model_predt,0.295664,0.950206);
%CAT (diss.s2_sisessncs_model_pred,
diss.s2_sisessncs_model_predt,0.1615458,0.971048);
```

Step Four: Greedy Matching

```
/*Greedy Match with Caliper*/
  /*-----------------------------------------------------------------*
   | The documentation and code below is supplied by HSR CodeXchange.
   |
   *-----------------------------------------------------------------*/



  /*-----------------------------------------------------------------*
   | MACRO NAME  : gmatch
   | SHORT DESC  : Match 1 or more controls to cases using the
```

```
|               GREEDY algorithm
*------------------------------------------------------------------*
| CREATED BY  : Kosanke, Jon                  (04/07/2004 16:32)
|             : Bergstralh, Erik
*------------------------------------------------------------------*
| PURPOSE
|
| GMATCH Macro to match 1 or more controls for each of N cases
| using the GREEDY algorithm--REPLACES GREEDY option of MATCH macro.
| Changes:
| --cases and controls in same dataset
| --not mandatory to randomly pre-ort cases and controls, but recommended
| --options to transform X's and to choose distance metric
| --input parameters consistent with %DIST macro for optimal matching
|
| *******
|
| Macro name: %gmatch
|
| Authors: Jon Kosanke and Erik Bergstralh
|
| Date: July 23, 2003
|       October 31, 2003...tweaked print/means based on "time" var
|
| Macro function:
|
| Matching using the GREEDY algorithm
|
| The purpose of this macro is to match 1 or more controls(from a total
| of M) for each of N cases.  The controls may be matched to the cases by
| one or more factors(X's).  The control selected for a particular
| case(i) will be the control(j) closest to the case in terms of Dij.
| Dij can be defined in multiple ways. Common choices are the Euclidean
| distance and the weighted sum of the absolute differences between the
| case and control matching factors.  I.e.,
|
|     Dij= SQRT [SUM { W.k*(X.ik-X.jk)**2} ],  or
|
|     Dij= SUM { W.k*ABS(X.ik-X.jk) },
|
|                                      where the sum is over the number
|                                      of matching factors X(with index
|                                      k) and W.k = the weight assigned
|                                      to matching factor k and X.ik =
|                                      the value of variable X(k) for
|                                      subject i.
|
| The control(j) selected for a case(i) is the one with the smallest Dij
| (subject to constraints DMAX and DMAXK, defined below). In the case of
| ties, the first one encountered will be used. The higher the user-
defined
| weight, the more likely it is that the case and control will be matched
| on the factor.  Assign large weights (relative to the other weights) to
```

```
   | obtain exact matches for two-level factors such as gender. An option to
   | using weights might be to standarize the X's in some fashion. The macro
   | has options to standardize all X's to mean 0 and variance 1 and to use
   | ranks.
   |
   | The matching algorithm used is the GREEDY method. Using the greedy
method,
   | once a match is made it is never broken.  This may result in
inefficiencies
   | if a previously matched control would be a better match for the current
   | case than those controls currently available. (An alternative method is
to
   | do optimal matching using the VMATCH & DIST macros. This method
guarantees
   | the best possible matched set in terms of minimizing the total Dij.)
   | The GREEDY method generally produces very good matches, especially if
the
   | control pool is large relative to the number of cases. When  multiple
   | controls/case are desired, the algorithm first matches 1 control to all
   | cases and then proceeds to select second controls.
   |
   |
   | The gmatch macro checks for missing values of matching variables and the
   | time variable(if specified) and deletes those observations from the
input
   | dataset.
   |
   | Call statement:
   |
   |
   | %gmatch(data=,group=,id=,
   |       mvars=,wts=,dmaxk=,dmax=,transf,
   |       time=, dist=,
   |       ncontls=,seedca=,seedco=,
   |       out=,outnmca=,outnmco=,print=);
   |
   | Parameter definitions(R=required parameter):
   |
   |
   |  R    data  SAS data set containing cases and potential controls. Must
   |             contain the ID, GROUP, and the matching variables.
   |
   |  R    group SAS variable defining cases. Group=1 if case, 0 if control.
   |
   |  R     id   SAS CHARACTER ID variable for the cases and controls.
   |
   |
   |  R   mvars  List of numeric matching variables common to both case and
   |             control data sets.  For example, mvars=male age birthyr.
   |
   |  R     wts  List of non-negative weights corresponding to each matching
   |             variable.  For example wts=10 2 1 corresponding to male, age
   |             and birthyr as in the above example.
```

```
|
|       dmaxk  List of non-negative values corresponding to each matching
|              variable.  These numbers are the largest possible absolute
|              differences compatible with a valid match.  Cases will
|              NOT be matched to a control if ANY of the INDIVIDUAL
|              matching factor  differences are >DMAXK.  This optional
|              parameter allows one to form matches of the type male+/-0,
|              age+/-2, birth year+/-5 by specifying DMAXK=0 2 5.
|
|       dmax   Largest value of Dij considered to be a valid match.  If
|              you want to match exactly on a two-level factor(such as
|              gender coded as 0 or 1) then assign DMAX to be less than
|              the weight for the factor.  In the example above, one could
|              use wt=10 for male and dmax=9.  Leave DMAX blank if any
|              Dij is a valid match.  One would typically NOT use both
|              DMAXK and DMAX.  The only advantage to using both, would be
|              to further restrict potential matches that meet the
|              DMAXK criteria.
|
|        dist  Indicates type of distance to calculate.
|
|              1=weighted sum(over matching vars) of
|              absolute case-control differences(default)
|
|              2=weighted Euclidean distance
|
|        time  Time variable used for risk set matching.  Matches are only
|              valid if the control time > case time. May need to
|
|      transf  Indicates whether all matching vars are to be transformed
|              (using the combined case+control data) prior to computing
|              distances.  0=no(default),
|                          1=standardize to mean 0 and variance 1,
|                          2=use ranks of matching variables.
|
|     ncontls  Indicates the number of controls to match to each case.  The
|              default is 1.  With multiple controls per case, the
algorithm
|              will first match every case to one control and then again
|              match each case to a second control, etc.  Controls selected
|              on the first pass will be stronger matches than those
selected in
|              later rounds.  The output data set contains a variable
(cont_n)
|              which indicates on which round the control was selected.
|
|     seedca   Seed value used to randomly sort the cases prior to
|              matching. This positive integer will be used as input to
|              the RANUNI function.  The greedy matching algorithm is
|              order dependent which, among other things means that
|              cases matched first will be on average more similar to
|              their controls than those matched last(as the number of
|              control choices will be limited).  If the matching order
```

```
|              is related to confounding factors (possibly age or
|              calendar time) then biases may result.  Therefore it is
|              generally considered good practice when using the GREEDY
|              method to randomly sort both the cases and controls
|              before beginning the matching process.
|
|     seedco   Seed value used to randomly sort the controls prior to
|              matching using the GREEDY method.  This seed value must
|              also be a positive integer.
|
|
| print= Option to print data for matched cases. Use PRINT=y to
|        print data and PRINT=n or blank to not print.  Default is y.
|
|        out=name of SAS data set containing the results of the matching
|            process.  Unmatched cases are not included.  See outnm
|            below.  The default name is __out.  This data set will have
|            the following layout:
|
|          Case_id  Cont_id  Cont_n  Dij  Delta_caco MVARS_ca  MVARS_co
|             1        67      1     5.2  (Differences & actual
|             1        78      2     6.1   values for matching factors
|             2        52      1     2.9   for cases & controls)
|             2        92      2     3.1
|             .        .       .      .
|             .        .       .      .
|
|        outnmca=name of SAS data set containing NON-matched cases.
|                Default name is __nmca .
|
|        outnmco=name of SAS data set containing NON-matched controls.
|                Default name is __nmco .
|
|
|  References:  Bergstralh, EJ and Kosanke JL(1995).  Computerized
|               matching of controls.  Section of Biostatistics
|               Technical Report 56.  Mayo Foundation.
|
|
|  Example: 1-1 matching by male(exact), age(+-2) and year(+-5).
|           The wt for male is not relevant, as only exact matches
|           on male will be considered.  The weight for age(2) is
|           double that for year(1).
|
|
|       %gmatch(data=all, group=ca_co,id=clinic,
|              mvars=male age_od yr_od,
|              wts=2 2 1, dmaxk=0 2 5,out=mtch,
|              seedca=87877,seedco=987973);
|
*------------------------------------------------------------------*
| OPERATING SYSTEM COMPATIBILITY
|
```

```
   | UNIX SAS v8    :   YES
   | UNIX SAS v9    :
   | MVS SAS v8     :
   | MVS SAS v9     :
   | PC SAS v8      :
   | PC SAS v9      :
   *------------------------------------------------------------------*
   | EXAMPLES
   |
   | Another example is located at the bottom of the code.
   *------------------------------------------------------------------*
   | Copyright 2004 Mayo Clinic College of Medicine.
   |
   | This program is free software; you can redistribute it and/or
   | modify it under the terms of the GNU General Public License as
   | published by the Free Software Foundation; either version 2 of
   | the License, or (at your option) any later version.
   |
   | This program is distributed in the hope that it will be useful,
   | but WITHOUT ANY WARRANTY; without even the implied warranty of
   | MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU
   | General Public License for more details.
   *------------------------------------------------------------------*/
 /*reverse control and treatment groups for matching*/

/*MAKE REVERSE FILE FOR CONDITIONING*
%macro CAT (file, file2);
data &FILE2;
set &FILE;
if F1_15 = 1 then F1_15r = 0;
if F1_15 = 0 then F1_15r = 1;
run;
%mend CAT;
%CAT (diss.s2_SISMODEL_PREDT, diss.s2_sismodel_rev);
%CAT (diss.s2_SISESS MODEL_PREDT, diss.s2_sisESS model_rev);
%CAT (diss.s2_SISNCS_MODEL_PREDT, diss.s2_sisNCS_model_rev);
%CAT (diss.s2_SISESSNCS_MODEL_PREDT, diss.s2_sisESSNCS_model_rev);



/*SD =0.1134223*
PROC MEANS DATA = DISS.S2_SISMODEL_PREDT STD;
      VAR PROB;
RUN;

/*SD = 0.1361091 *
PROC MEANS DATA = DISS.S2_SISESS MODEL_PREDT STD;
      VAR PROB;
RUN;

/*SD = 0.1221295 *
PROC MEANS DATA = DISS.S2_SISNCS_MODEL_PREDT STD;
      VAR PROB;
```

```
RUN;

/*SD = 0.1498787*
PROC MEANS DATA = DISS.S2_SISESSNCS_MODEL_PREDT STD;
      VAR PROB;
RUN;

/*GREEDY MATCHING - CALIPER*/


%MACRO GMATCH(DATA=,GROUP=,ID=,
             MVARS=,WTS=,DMAXK=,DMAX=,DIST=1,
             NCONTLS=1, TIME=,TRANSF=0,
             SEEDCA=,SEEDCO=,PRINT=y,
             OUT=,OUT2=,OUTNMCA=__NMCA,OUTNMCO=__NMCO);

   %LET BAD=0;

   %IF %LENGTH(&DATA)=0 %THEN %DO;
      %PUT ERROR: NO DATASET SUPPLIED;
      %LET BAD=1;
   %END;

   %IF %LENGTH(&ID)=0 %THEN %DO;
      %PUT ERROR: NO ID VARIABLE SUPPLIED;
      %LET BAD=1;
   %END;

   %IF %LENGTH(&GROUP)=0 %THEN %DO;
      %PUT ERROR: NO CASE(1)/CONTROL(0) GROUP VARIABLE SUPPLIED;
      %LET BAD=1;
   %END;

   %IF %LENGTH(&MVARS)=0 %THEN %DO;
      %PUT ERROR: NO MATCHING VARIABLES SUPPLIED;
      %LET BAD=1;
   %END;

  %IF %LENGTH(&WTS)=0 %THEN %DO;
      %PUT ERROR: NO WEIGHTS SUPPLIED;
      %LET BAD=1;
   %END;

   %LET NVAR=0;
   %DO %UNTIL(%SCAN(&MVARS,&NVAR+1,' ')= );
      %LET NVAR=%EVAL(&NVAR+1);
   %END;
   %LET NWTS=0;
   %DO %UNTIL(%QSCAN(&WTS,&NWTS+1,' ')= );
      %LET NWTS=%EVAL(&NWTS+1);
   %END;
   %IF &NVAR^= &NWTS %THEN %DO;
      %PUT ERROR: #VARS MUST EQUAL #WTS;
```

```
    %LET BAD=1;
 %END;

%LET NK=0;
 %IF %QUOTE(&DMAXK)^=  %THEN %DO %UNTIL(%QSCAN(&DMAXK,&NK+1,' ')= );
    %LET NK=%EVAL(&NK+1);
 %END;
 %IF &NK>&NVAR %THEN %LET NK=&NVAR;
 %DO I=1 %TO &NVAR;
    %LET V&I=%SCAN(&MVARS,&I,' ');
 %END;

%IF &NWTS>0 %THEN %DO;
      DATA _NULL_;
      %DO I=1 %TO &NWTS;
           %LET W&I=%SCAN(&WTS,&I,' ');
           IF &&W&I<0 THEN DO;
                PUT 'ERROR: WEIGHTS MUST BE NON-NEGATIVE';
                CALL SYMPUT('BAD','1');
           END;
      %END;
      RUN;
 %END;

%IF &NK>0 %THEN %DO;
      DATA _NULL_;
      %DO I=1 %TO &NK;
           %LET K&I=%SCAN(&DMAXK,&I,' ');
           IF &&K&I<0 THEN DO;
                PUT 'ERROR: DMAXK VALUES MUST BE NON-NEGATIVE';
                CALL SYMPUT('BAD','1');
           END;
      %END;
      RUN;
 %END;

  %MACRO MAX1;
    %IF &DMAX^= %THEN %DO;
       & __D<=&DMAX
    %END;
    %DO I=1 %TO &NK;
       & ABS(__CA&I-__CO&I)<=&&K&I
    %END;
  %MEND MAX1;

 %macro greedy;
  %GLOBAL BAD2;

    data __CHECK; set &DATA;
        __id=&id;
        if __id="" then delete;
        %DO I=1 %TO &NVAR;
              IF %scan(&mvars,&i)=. THEN DELETE;
```

```
     %END;
     %IF &TIME^= %THEN %DO;
          IF &TIME=. THEN DELETE;
     %END;
 run;

*** transform data if requested/separate cases & controls;
%if &transf=1 %then %do;
proc standard data=__check m=0 s=1 out=_stdzd; var &mvars;
data _caco;
  set _stdzd;
%end;

%if &transf=2 %then %do;
proc rank data=__check out=_ranks; var &mvars;
data _caco;
  set _ranks;
%end;

%if &transf=0 %then %do;
data _caco;
  set __check;
%end;


DATA __CASE; SET _caco;
     if &group=1;
DATA __CASE; SET __CASE END=EOF;
 KEEP __IDCA __CA1-__CA&NVAR __R &mvars
   %if &time^= %then %do;
       __catime
   %end;
    ;
   __IDCA=&ID;
   %if &time^= %then %do;
      __catime=&time;
   %end;
   %DO I=1 %TO &NVAR;
      __CA&I=&&V&I;
   %END;
   %if &seedca^= %then %do;
   SEED=&SEEDCA;
   __R=RANUNI( SEED  );
   %end;
   %else %do;
   __R=1;
   %end;

   IF EOF THEN CALL SYMPUT('NCA',_N_);
PROC SORT; BY __R __IDCA;

DATA __CONT; SET _caco;
   if &group=0;
```

```
DATA __CONT; SET __CONT END=EOF;
 KEEP __IDCO __CO1-__CO&NVAR __R &mvars
  %if &time^= %then %do;
     __cotime
  %end;
  ;
   __IDCO=&ID;
   %if &time^= %then %do;
      __cotime=&time;
   %end;
   %DO I=1 %TO &NVAR;
      __CO&I=&&V&I;
   %END;
   %if &seedco^= %then %do;
   SEED=&SEEDCo;
   __R=RANUNI( SEED  );
   %end;
   %else %do;
   __R=1;
   %end;

   IF EOF THEN CALL SYMPUT('NCO',_N_);
RUN;
%LET BAD2=0;
%IF &NCO < %EVAL(&NCA*&NCONTLS) %THEN %DO;
   %PUT ERROR: NOT ENOUGH CONTROLS TO MAKE REQUESTED MATCHES;
   %LET BAD2=1;
%END;

%IF &BAD2=0 %THEN %DO;
   PROC SORT; BY __R __IDCO;
   DATA __MATCH;
    KEEP __IDCA __CA1-__CA&NVAR __DIJ __MATCH __CONT_N
    %if &time^= %then %do;
       __catime __cotime
    %end;
    ;
    ARRAY __USED(&NCO) $ 1 _TEMPORARY_;
      DO __I=1 TO &NCO;
         __USED(__I)='0';
      END;
      DO __I=1 TO &NCONTLS;
         DO __J=1 TO &NCA;
            SET __CASE POINT=__J;
            __SMALL=.;
            __MATCH=.;
            DO __K=1 TO &NCO;
               IF __USED(__K)='0' THEN DO;
                  SET __CONT POINT=__K;

                 %if &dist=2 %then %do;
                  **wtd euclidian dist;
                   __D= sqrt(
```

```
                         %do k=1 %to &nvar;
                         %scan(&wts,&k)*(__ca&k - __co&k)**2
                         %if &k<&nvar %then + ;
                        %end;
                         );
                       %end;
                       %else %do;
                        **wtd sum absolute diff;
                         __D=
                        %do k=1 %to &nvar;
                        %scan(&wts,&k)*abs(__ca&k - __co&k )
                        %if &k<&nvar %then + ;
                        %end;
                          ;
                       %end;

                        IF __d^=. & (__SMALL=. | __D<__SMALL) %MAX1
                        %if &time^= %then %do;
                           & __cotime > __catime
                        %end;
                        THEN DO;
                           __SMALL=__D;
                           __MATCH=__K;
                           __DIJ=__D;
                           __CONT_N=__I;
                        END;
                     END;
                  END;
                  IF __MATCH^=. THEN DO;
                     __USED(__MATCH)='1';
                     OUTPUT;
                  END;
               END;
            END;
            STOP;
         DATA &OUT;
          SET __MATCH;
          SET __CONT POINT=__MATCH;
          KEEP __IDCA __IDCO __CONT_N __DIJ __CA1-__CA&NVAR
               __CO1-__CO&NVAR __d1-__d&nvar __absd1-__absd&nvar  __WT1-
__WT&NVAR
                  __catime __cotime __dtime;

          %if &time= %then %do;
              __cotime=.; __catime=.;
          %end;
          LABEL
                   __catime="&time/CASE"
                   __cotime="&time/CONTROL"
                   __dtime="&time/ABS. DIFF"
                __CONT_N='CONTROL/NUMBER'
                __DIJ='DISTANCE/D_IJ'
               %DO I=1 %TO &NVAR;
```

```
            __CA&I="&&V&I/CASE"
            __CO&I="&&V&I/CONTROL"
            __absd&I="&&V&I/ABS. DIFF "
            __d&I="&&V&I/DIFF "
            __WT&I="&&V&I/WEIGHT"
          %END;
            ;
         %DO I=1 %TO &NVAR;
            __d&i= (__CA&I-__CO&I);      **raw diff;
            __absd&I=abs(__CA&I-__CO&I); **abs diff;
            __WT&I=&&W&I;
         %END;
            __dtime=__cotime-__catime;

     PROC SORT DATA=&OUT; BY __IDCA __CONT_N;
     proc sort data=__case; by __IDCA;
     data &outnmca; merge __case
          &out(in=__inout where=(__cont_n=1)); by __idca;
          if __inout=0; **non-matches;

     proc sort data=__cont; by __IDCO;
     proc sort data=&out; by __IDCO;
     data &outnmco; merge __cont
          &out(in=__inout); by __idco;
          if __inout=0; **non-matched controls;
     proc sort data=&out; by __IDCA; **re-sort by case id;

   %if %upcase(&print)=Y %then %do;
     PROC PRINT data=&out LABEL SPLIT='/';
      VAR __IDCA __IDCO __CONT_N

       __DIJ
      %DO I=1 %TO &NVAR;
       __absd&I
      %END;
      %if &time^= %then %do;
       __dtime
      %end;
      %DO I=1 %TO &NVAR;
       __CA&I __CO&I
      %END;
      %if &time^= %then %do;
       __catime __cotime
      %end;
       ;
      sum __dij;

     title9'Data listing for matched cases and controls';
     footnote"Greedy matching(gmatch) macro: data=&data group=&group
id=&id    ";
     footnote2"   mvars=&mvars  wts=&wts dmaxk=&dmaxk dmax=&dmax
ncontls=&ncontls";
```

```
        footnote3"   transf=&transf dist=&dist time=&time seedca=&seedca
seedco=&seedco";
        footnote4"   out=&out   outnmca=&outnmca  outnmco=&outnmco";
        run;
        title9'Summary data for matched cases and controls--one
obs/control';
         %if &sysver ge 8 %then %do;
        proc means data=&out  maxdec=3 fw=8
          n mean median min p10 p25 p75 p90 max sum;
        %end;
        %else %do;
        proc means data=&out maxdec=3
         n mean min max sum;
        %end;
        class __cont_n;
         var __dij

             %do I=1 %TO &NVAR;
                 __absd&I
             %end;
             %if &time^= %then %do;
                 __dtime
             %end;
             %do I=1 %TO &NVAR;
                 __ca&I
             %end;
             %if &time^= %then %do;
                 __catime
             %end;
             %do I=1 %TO &NVAR;
                 __co&I
             %end;
             %if &time^= %then %do;
                 __cotime
             %end;
                ;
        run;
        *** estimate matching var means within matched sets for controls;
        proc means data=&out  n mean noprint; by __idca;
         var __dij
        %do i=1 %to &nvar;
           __co&i
        %end;
             __cotime
           ;
        output out=_mcont n=n_co mean=__dijm
        %do i=1 %to &nvar;
          __com&i
        %end;
            __tcom
          ;
        data _onecase; set &out; by __idca; if first.__idca;
        data __camcon; merge _onecase _mcont; by __idca;
```

```
    keep __idca n_co __dijm
        __dtime __catime  __tcom
     %do i=1 %to &nvar;
      __ca&i __com&i  __actd&i __absd&i
     %end;
    ;


    %do i=1 %to &nvar;
    __absd&i=abs(__ca&i - __com&i);
    __actd&i=(__ca&i - __com&i);
   %end;
    __dtime=__tcom-__catime
     ;

  label
   n_co="No./CONTROLS"
   __dijm="Average/Dij"
   __dtime="&time/Mean Time DIFF"
   __tcom="&time/Mean CONT TIME"

  %do i=1 %to &nvar; %let vvar=%scan(&mvars,&i);
    __absd&i="&vvar/Mean ABS. DIFF"
    __com&i="&vvar/Mean CONTROL"
  %end;
    ;
 title9'Summary data for matched cases and controls--one obs/case(using
average control value)';
 %if &sysver ge 8 %then %do;
 proc means data=__camcon maxdec=3 fw=8
   n mean median min p10 p25 p75 p90 max sum;
 %end;
 %else %do;
 proc means data=__camcon maxdec=3
   n mean min max sum;
 %end;
 var n_co __dijm
 %do i=1 %to &nvar;
  __absd&i
 %end;
 %if &time^= %then %do;
  __dtime
 %end;
 %do i=1 %to &nvar;
 __ca&i
 %end;
 %if &time^= %then %do;
  __catime
 %end;
 %do i=1 %to &nvar;
 __com&i
 %end;
```

```
      %if &time^= %then %do;
      __tcom
      %end;
            ;
     %end; **end of print=y loop**;
    %END;  **end of bad2=0 loop**;
    run;
    title9; footnote;
    run;

    %mend greedy;

    %IF &BAD=0 %THEN %DO;
          %GREEDY
    %END;

    PROC SQL;
             CREATE TABLE CASES AS
             SELECT *
       FROM &DATA
       INNER JOIN &OUT
       ON __IDCA=EUIN;
       QUIT;

       PROC SQL;
             CREATE TABLE CONTROL AS
             SELECT *
       FROM &DATA
       INNER JOIN &OUT
       ON __IDCO=EUIN;
       QUIT;

       DATA &OUT2;
             SET CASES CONTROL;
       RUN;

       PROC PRINT DATA=&OUT2;
RUN;

%MEND GMATCH;


/*SIS MODELS*
%gmatch(data=diss.S2_SISMODEL_REV, group=f1_15R, id=euin, mvars=prob,wts = 0,
dmaxk=, dist=2,
ncontls=1,seedca=2546, seedco=679, OUT=S4_SIS_CMATCH,
OUT2=DISS.S4_SIS_CMATCH0, print=Y);
run;

%gmatch(data=diss.S2_SISMODEL_REV, group=f1_15r, id=euin, mvars=prob,wts = 0,
dmaxk=(.25*0.1134223), dist=2,
ncontls=1,seedca=2546, seedco=679, OUT=S4_SIS_CMATCH25,
out2=DISS.S4_SIS_CMATCH25, print=Y);
```

```
run;
%gmatch(data=diss.S2_SISMODEL_REV, group=f1_15r, id=euin, mvars=prob,wts = 0,
dmaxk=(.1*0.1134223), dist=2,
ncontls=1,seedca=2546, seedco=679, OUT=S4_SIS_CMATCH1,
out2=DISS.S4_SIS_CMATCH1, print=Y);
run;

/*SIS ESS MODELS*
%gmatch(data=diss.S2_SISESS MODEL_REV, group=f1_15r, id=euin, mvars=prob,wts
= 0, dmaxk=, dist=2,
ncontls=1,seedca=2546, seedco=679, OUT=S4_SISESS CMATCH0, out2=DISS.S4_SISESS
CMATCH0, print=Y);
run;
%gmatch(data=diss.S2_SISESS MODEL_REV, group=f1_15r, id=euin, mvars=prob,wts
= 0, dmaxk=(.25*0.1361091), dist=2,
ncontls=1,seedca=2546, seedco=679, OUT=S4_SISESS CMATCH25,
out2=DISS.S4_SISESS CMATCH25, print=Y);
run;
%gmatch(data=diss.S2_SISESS MODEL_REV, group=f1_15r, id=euin, mvars=prob,wts
= 0, dmaxk=(.1*0.1361091), dist=2,
ncontls=1,seedca=2546, seedco=679, OUT=S4_SISESS CMATCH1, out2=DISS.S4_SISESS
CMATCH1, print=Y);
run;

/*SIS NCS MODELS*
%gmatch(data=diss.S2_SISNCS_MODEL_REV, group=f1_15r, id=euin, mvars=prob,wts
= 0, dmaxk=, dist=1,
ncontls=1,seedca=2546, seedco=679, OUT=S4_SISNCS_CMATCH0,
out2=DISS.S4_SISNCS_CMATCH0, print=Y);
run;

%gmatch(data=diss.S2_SISNCS_MODEL_REV, group=f1_15r, id=euin, mvars=prob,wts
= 0, dmaxk=(.25*0.1221295), dist=1,
ncontls=1,seedca=2546, seedco=679, OUT=S4_SISNCS_CMATCH25,
out2=DISS.S4_SISNCS_CMATCH25, print=Y);
run;

%gmatch(data=diss.S2_SISNCS_MODEL_REV, group=f1_15r, id=euin, mvars=prob,wts
= 0, dmaxk=(.1*0.1221295), dist=1,
ncontls=1,seedca=2546, seedco=679, OUT=S4_SISNCS_CMATCH1,
out2=DISS.S4_SISNCS_CMATCH1, print=Y);
run;

/*SIS ESS NCS MODELS*/

%gmatch(data=diss.S2_SISESSNCS_MODEL_REV, group=f1_15r, id=euin,
mvars=prob,wts = 0, dmaxk=, dist=1,
ncontls=1,seedca=2546, seedco=679, OUT=S4_SISESSNCS_CMATCH0,
OUT2=DISS.S4_SISESSNCS_CMATCH0, print=Y);
run;

%gmatch(data=diss.s2_SISESSNCS_MODEL_REV, group=f1_15r, id=euin,
mvars=prob,wts = 0, dmaxk=(.25*0.1498787), dist=1,
```

```
ncontls=1,seedca=2546, seedco=679, OUT=S4_SISESSNCS_CMATCH25,
OUT2=DISS.S4_SISESSNCS_CMATCH25, print=Y);
run;

%gmatch(data=diss.s2_SISESSNCS_MODEL_REV, group=f1_15r, id=euin,
mvars=prob,wts = 0, dmaxk=(.1*0.1498787), dist=1,
ncontls=1,seedca=2546, seedco=679, OUT=S4_SISESSNCS_CMATCH1,
OUT2=DISS.S4_SISESSNCS_CMATCH1, print=Y);
run;
```

Step Four: Greedy 5->1 Digit Matching

```
/*http://www.citymatch.org/sites/default/files/documents/MCHEPITraining/Ranki
n_PropensityScoreMatching_WedsLateAfternoon.pdf*/
/*http://www2.sas.com/proceedings/sugi26/p214-26.pdf*/

/* ************************************** */
/* Greedy 5->1 Digit Matching Macro       */
/* ************************************** */
/*error in parsons code see
http://www2.sas.com/proceedings/sugi25/25/po/25p225.pdf*/
%MACRO GREEDMTCH
(
Lib,       /* Library Name            */
Dataset,   /* Data set of all         */
depend,    /* Dependent variable      */
/* that indicates          */
/* Case or Control;   matches
*/
/* Code 1 for Cases,       */
/*      0 for Controls     */
matches   /* Output file of matched  */
);

/* Macro to sort the Cases and Controls dataset */
%MACRO SORTCC;
proc sort data=tcases out=Scase;
by prob; run;
proc sort data=tctrl out=Scontrol;
by prob randnum;run;
%MEND SORTCC;

/* Macro to Create the initial Case and
Control Data Sets */
%MACRO INITCC (digits);
data tcases (drop=cprob) tctrl  (drop=aprob) ;
set &LIB..&dataset.;
/* Create the data set of Controls*/
if &depend. = 0 and prob ne . then do;
      cprob  = Round(prob,&digits.);
```

```
      Cmatch = 0;
      Length RandNum 8;
      RandNum=ranuni(1234567);
      Label RandNum= 'Uniform Randomization Score';
      output tctrl;
      end;
/* Create the data set of Cases   */
else if &depend. = 1 and prob ne . then do;
      Cmatch = 0;
      aprob =Round(prob,&digits.);
      output tcases;
      end;
run;
%sortcc;
%MEND INITCC;

/*  Macro to Perform the Match */
%MACRO MATCH (MATCHED,DIGITS);
data &matched. (drop=Cmatch randnum aprob cprob start oldi curctrl matched);
/* select the cases data set */
set SCase ;
curob + 1;
matchto = curob;
if curob = 1 then do;
start = 1;
oldi = 1;
end;
/* select the controls data set */
DO i = start to n;
set Scontrol point= i nobs = n;
if i gt n then goto startovr;
if _Error_ = 1 then abort;
curctrl = i;
/* output control if match found */
if aprob = cprob then do;
Cmatch = 1;
output &matched.;
matched = curctrl;
goto found;
end;
/* exit do loop if out of potential
matches */
else if cprob gt aprob then
goto nextcase;
startovr: if i gt n then
goto nextcase;
END; /* end of DO LOOP  */
/* If no match was found, put pointer
Posters
back*/
nextcase:
if Cmatch=0 then start = oldi;
/* If a match was found, output case and
```

```
increment pointer */
found:
if Cmatch = 1 then do;
oldi = matched + 1;
start = matched + 1;
set SCase point = curob;
output &matched.;
end;
retain oldi start;
if _Error_=1 then _Error_=0;
run;

/* Get files of unmatched cases and     */
/* controls.  Note that in the example  */
/* data, the patient identifiers are HID*/
/* (Hospital ID) and PATIENTN (Patient  */
/* identifier.  All cases have complete */
/* data for these two fields.  Modify   */
/* these fields with the appropriate    */
/* patient identifier field(s)          */
proc sort data=scase out=sumcase;
by euin;
run;
proc sort data=scontrol
out=sumcontrol;
by euin;
run;
proc sort data=&matched. out=smatched
(keep= euin matchto);
by euin;
run;
data tcases (drop=matchto);
merge sumcase(in=a) smatched;
by euin;
if a and matchto=.;
cmatch = 0;
aprob =Round(prob,&digits.);
run;
data tctrl (drop=matchto);
merge sumcontrol(in=a) smatched;
by euin;
if a and matchto=.;
cmatch = 0;
cprob  = Round(prob,&digits.);
run;
%SORTCC
%MEND MATCH;

/* Note:  This section can be        */
/* modified to try variations of the */
/* basic algorithm.                  */
/* Create file of cases and controls */
%INITCC(.00001);
```

```
/* Do a 5-digit match */
%MATCH(Match5,.00001);
/* Do a 4-digit match on remaining
unmatched */
%MATCH(Match4,.0001);
/* Do a 3-digit match on remaining
unmatched */
%MATCH(Match3,.001);
/* Do a 2-digit match on remaining
unmatched */
%MATCH(Match2,.01);
/* Do a 1-digit match on remaining
unmatched */
%MATCH(Match1,.1);

/* Merge all the matches into one file */
/* The purpose of the marchto variable */
/* is to identify matched pairs for the*/
/* matched pair anlayses.  matchto is  */
/* initially assigned the observation  */
/* number of the case.  Since there    */
/* would be duplicate numbers after the*/
/* individual files were merged,       */
/* matchto is incremented by file.     */
/* Note that if the controls file      */
/* contains more than N=100,000 records*/
/* and/or there are more than 1,000    */
/* matches made at each match level,   */
/* then the incrementation factor must */
/* be changed.                         */
data matches;
set match5(in=a) match4(in=b) match3(in=c) match2(in=d) match1(in=e);
if b then matchto=matchto + 100000;
if c then matchto=matchto + 10000000;
if d then matchto=matchto + 1000000000;
if e then matchto=matchto + 100000000000;
run;
/* Sort file -- Need sort for Univariate
analysis in tables
*/
proc sort data=matches out = &lib..&matches.;
by &depend.;
run;

%MEND GREEDMTCH;
/*
%GREEDMTCH (diss,s2_sismodel_predT,F1_15,s4_sis_dmatch);
%GREEDMTCH (diss,s2_sisESS model_predT,F1_15,s4_sisESS dmatch);
%GREEDMTCH (diss,s2_sisncs_model_predT,F1_15,s4_sisncs_dmatch);*/
%GREEDMTCH (diss,s2_sisessncs_model_predT,F1_15,s4_sisessncs_dmatch);
```

Step Five: Balance (Statistical)

```
/*balance statistical*/

/*sis models*/
%macro cat (file);
   proc logistic data=&file ;
        class GENDER ETHNIC (REF='White') HONCOLL PELL COLLEGE (REF =
'Liberal Arts & Sciences') SUMMCOLL HSCPS
            PLACEMENTWRITING (REF = 'ENGL 160') PLACEMENTMATH(REF='MATH 121,
160, 165 and STAT 101') / param=ref ref=first;
      model F1_15(event='1')=GENDER ETHNIC HONCOLL PELL HSCPS COLLEGE
PLACEMENTWRITING PLACEMENTMATH  SUMMCOLL
            SACTE SHSGPAR SACTM
                   / lackfit
                              rsquare;
    run;

%mend cat;
%cat (DISS.S4_SIS_CMATCH0);
%cat (DISS.S4_SIS_CMATCH25);
%cat (DISS.S4_SIS_CMATCH1);
%cat (DISS.S4_SIS_DMATCH);


/*sis+ess models*/
%macro cat (file);
proc logistic data=&file ;
        class
            /*SIS VARIABLES*/
            GENDER ETHNIC (REF='White') HONCOLL PELL COLLEGE (REF = 'Liberal
Arts & Sciences') SUMMCOLL
            HSCPS PLACEMENTMATH(REF='MATH 121, 160, 165 and STAT
101')PLACEMENTWRITING (REF = 'ENGL 160')
            /*ESS VARIABLES*/
            ESS liver (REF='1') ESS degreer (REF='3') ESS lang (REF='1') ESS
religion (REF='8') ESS apcourse (REF='1')
            ESS mathhad ESS mathneed ESS scihad ESS scineed ESS writehad ESS
writewil/ param=ref ref=last;

      model F1_15(event='1')=
            /*SIS VARIABLES*/
            GENDER ETHNIC HONCOLL PELL COLLEGE HSCPS SUMMCOLL PLACEMENTMATH
PLACEMENTWRITING
            SACTE SACTM SHSGPAr
            /*ESS VARIABLES*/
            ESS liver ESS degreer ESS mathhad ESS mathneed ESS scihad ESS
scineed ESS writehad ESS writewil
            ESS lang ESS religion ESS apcourse
            Q106 Q108 Q111 Q153 Q156 Q157 Q159 Q1511 Q1516 Q1520 Q1521

             / lackfit rsquare;
   run;
```

```
%mend cat;
%cat (DISS.S4_SISESS CMATCH0);
%cat (DISS.S4_SISESS CMATCH25);
%cat (DISS.S4_SISESS CMATCH1);
%cat (DISS.S4_SISESS DMATCH);


/*sis+ncs models*/
%macro cat (file);
proc logistic data=&FILE;
        class
            /*SIS VARIABLES*/
            GENDER ETHNIC (REF='White') HONCOLL PELL COLLEGE (REF = 'Liberal
Arts & Sciences') SUMMCOLL
            HSCPS PLACEMENTMATH(REF='MATH 121, 160, 165 and STAT
101')PLACEMENTWRITING (REF = 'ENGL 160')/ param=ref ref=last;

      model F1_15(event='1')=
            /*SIS VARIABLES*/
            GENDER ETHNIC HONCOLL PELL COLLEGE HSCPS SUMMCOLL PLACEMENTMATH
PLACEMENTWRITING
            SACTE SACTM SHSGPAr

            /*NCS VARIABLES*/
            SelfEff_Total TimeManage_total swb_total familyob_total
grit_total academiccontrol_total
            CARING1     CARING2     CARING3     CARING4     CARING5
      CARING6     CARING7     CARING8     CARING9
                   / lackfit
                               rsquare;
   run;
%mend cat;
%cat (DISS.S4_SISNCS_CMATCH0);
%cat (DISS.S4_SISNCS_CMATCH25);
%cat (DISS.S4_SISNCS_CMATCH1);
%cat (DISS.S4_SISNCS_DMATCH);




/*sis+ess+ncs models*/
%macro cat (file);
proc logistic data=&FILE;
         class
            /*SIS VARIABLES*/
            GENDER ETHNIC (REF='White') HONCOLL PELL COLLEGE (REF = 'Liberal
Arts & Sciences') SUMMCOLL
            HSCPS PLACEMENTMATH(REF='MATH 121, 160, 165 and STAT
101')FULL_WRITING (REF = 'ENGL 160')
            /*ESS VARIABLES*/
            ESS liver (REF='1') ESS degreer (REF='3') ESS lang (REF='1') ESS
religion (REF='8') ESS apcourse (REF='1')
```

```
            ESS mathhad ESS mathneed ESS scihad ESS scineed ESS writehad ESS
writewil/ param=ref ref=last;

      model F1_15(event='1')=
            /*SIS VARIABLES*/
            GENDER ETHNIC HONCOLL PELL COLLEGE HSCPS SUMMCOLL FULL_WRITING
PLACEMENTMATH
            SACTE SACTM SHSGPAr
            /*ESS VARIABLES*/
            ESS liver ESS degreer ESS mathhad ESS mathneed ESS scihad ESS
scineed ESS writehad ESS writewil
            ESS lang ESS religion ESS apcourse
            Q106 Q108 Q111 Q153 Q156 Q157 Q159 Q1511 Q1516 Q1520 Q1521
            /*NCS VARIABLES*/
            SelfEff_Total TimeManage_total swb_total familyob_total
grit_total academiccontrol_total
            CARING1     CARING2     CARING3     CARING4     CARING5
      CARING6     CARING7     CARING8     CARING9
             / lackfit
                  rsquare;
   run;
%mend cat;
%cat (DISS.S4_SISESSNCS_CMATCH0);
%cat (DISS.S4_SISESSNCS_CMATCH25);
%cat (DISS.S4_SISESSNCS_CMATCH1);
%cat (DISS.S4_SISESSNCS_DMATCH);
```

Step Five: Balance (Standard Mean Difference)

```
/*BALANCE*/

/****************************************************************************
**/
/* Program : stddiff.sas
/* Purpose : SAS macro to calculate the Standardized Difference
/* Usage : %stddiff(inds = Studydata, groupvar = dex,
/* numvars = age bmi/r glucose,
/* charvars = female surgtype,
/* stdfmt = 8.5,
/* outds = std_result);
/****************************************************************************
***/
/* NOTE: All binary variables must be coded as 0 and 1 in the dataset
/* PARAMETERS:
/* inds:       input dataset
/* groupvar:   a binary variable, must be coded as 0 and 1
/* numvars:    a list of continuous variables.
/*             "/r" denotes to use the rank-based mean and SD to calculate
Stddiff
/* charvars:   a list of categorical variables. If a variable is a binary
categorical variable,
```

```
/*              it must be coded as 0 and 1 since we use the level = 0 as the
reference level.
/* stdfmt = 8.5 the format of Standardized Difference
/* outds output result dataset
/****************************************************************************
*****/

options  symbolgen mlogic mprint;
%macro      stddiff( inds,
                     groupvar,
                     numvars,
                     charvars,
                     wtvar,
                     stdfmt,
                     outds );

/* create a table to store stddiff */
proc sql;
   create table &outds.
       (VarName char(32),
            Stddiff char (10)
       );
quit;

/* delete records if the group variable is missing */

data base_data;
      set &inds.;
      where &GroupVar. ne .;
run;

/* remove leading or tailing blanks */
%let groupvar = %sysfunc(strip(&GroupVar.));

                              /****************************************/
                              /* part 1: compare continuous variables */
                              /****************************************/

%if %length(&numvars.) > 0 %then %do;

/* remove multiple blanks and get the total number of continuous variables */
      %let numvar = %sysfunc(compbl(&numvars.));
      %let numvar = %sysfunc(strip(&numvar.));
      %let n_convar = %sysfunc(countc(&numvar.,' '));
      %let n_convar = %eval(&n_convar. + 1);

/* summarize variables one-by-one */
      %do ii = 1 %to &n_convar.;
      %let convar = %sysfunc(scan(&numvar.,&ii.,' '));

    /* if requires rank-based mean and std for skewed variables */
            %if %index(&convar., /r) > 0 %then %do;
            %let convar = %sysfunc(scan(&convar.,1,'/'));
```

```
        %let convar = %sysfunc(strip(&convar.));

        data temp_1;
              set base_data (keep = &groupvar. &convar. &wtvar.);
        run;

/* rank a variable */
        proc rank data=temp_1 out=temp_2;
              var &convar.;
              ranks rank_&convar.;
        run;

/* get ranked-mean and sd */

              proc means data = temp_2;
                    class &groupvar.;
                    var rank_&convar.;
                    weight &wtvar.;
                    output out = temp_3 mean = _mean_  std = _std_;
              run;

              data  temp_3;
                    set temp_3;
                    where _type_ = 1;
              run;

              proc sort data = temp_3;
                    by &groupvar.;
              run;
        %end;

  /* for normal-distributed variable */

        %else %do;
        %let convar = %sysfunc(strip(&convar.));
        data temp_1;
              set base_data (keep = &groupvar. &convar. &wtvar.);
        run;
        data temp_2;
              set temp_1;
        run;

      /* get mean and sd */

              proc means data = temp_2;
                    class &groupvar.;
                    var &convar.;
                    weight &wtvar.;
                    output out = temp_3 mean = _mean_  std = _std_;
              run;

              data  temp_3;
                    set temp_3;
```

```
                        where _type_ = 1;
                  run;

                  proc sort data = temp_3;
                        by &groupvar.;
                  run;

            %end;

/* calculate stddiff */
         proc sql;
            create table temp_4 as
                  select (a._mean_ - b._mean_)/
                        sqrt((a._std_**2 + b._std_**2)/2) as d
                  from temp_3(where = (&groupvar = 1)) as a,
                               temp_3(where = (&groupvar = 0)) as b;
         quit;

         data temp_5;
                  set temp_4;
              stddiff = compress(put(d,&stdfmt.));
              keep stddiff;
          run;

      /* insert into std table */
         proc sql noprint;
            select stddiff into: std_value from temp_5;
            insert into &outds.  values("&convar.", "&std_value.");
         quit;

      /* delete temporary data sets */

         proc datasets lib = work nodetails nolist;
          delete  temp_1 - temp_5;
         quit;
         %end;
%end;

            /********************************************/
            /* part 2: compare categorical variables      */
            /********************************************/

%if %length(&charvars.) > 0 %then %do;
      %let n_charvar = %sysfunc(countw(&charvars.));

/* get column percents for each levels of the variable by the group */
      %do jj = 1 %to &n_charvar.;
            %let char_var = %scan(&charvars., &jj.);
            %let char_var = %sysfunc(strip(&char_var.));
              data temp_1;
                  set base_data (keep = &groupvar. &char_var. &wtvar.);
              run;
```

```
      proc sql;
          create table temp_2 as
          select distinct &char_var. as &char_var.
          from temp_1
          where &char_var. is not missing;
      quit;

      proc sql noprint;
          select count(*) into :_mylevel_ from temp_2;
      quit;

    %let _mylevel_ = %sysfunc(strip(&_mylevel_.));

      data temp_3;
          set temp_2;
                do &groupvar. = 0,1 ;
          output;
                end;
      run;

    ods output CrossTabFreqs = temp_4;
    proc freq data = temp_1;
          table &char_var. * &groupvar.;
          %if %length(&wtvar.) > 0 %then %do;
                weight &wtvar.;
                %end;
    run;

    proc sql;
          create table  temp_5 as
          select a.*, b.ColPercent
          from temp_3 as a
          left join temp_4 as b
          on    a.&groupvar. = b.&groupvar. and
          a.&char_var. = b.&char_var.;
    quit;

    data temp_6;
          set temp_5;
          if ColPercent = . then ColPercent = 0;
    run;

    proc sort data = temp_6 out = catfreq;
          by &groupvar. &char_var.;
    run;

    proc datasets lib = work nodetails nolist;
          delete  temp_1 - temp_6;
    quit;

/* if a categorical variable only has one level: 0 or 1 */
/* stddiff = 0 */
    %if &_mylevel_. = 1 %then %do;
```

```
                  proc sql noprint;
                        insert into &outds.  values("&char_var.", "0");
                  quit;
            %end;

/* if a categorical variable  has two level: 0 and 1 */
/* it is a binary variable, using two sample proportion formula */
            %else %if &_mylevel_. = 2 %then %do;

                  data temp_7;
                        set catfreq;
                        where &char_var. = 1;
                        ColPercent = ColPercent/100;
                  run;

                  proc sql;
                        create table temp_8 as
                        select (a.ColPercent -
b.ColPercent)/(sqrt((a.ColPercent*(1-
                                    a.ColPercent) +
                              b.ColPercent*(1-b.ColPercent))/2)) as d
                        from temp_7(where = (&groupvar = 1)) as a,
                   temp_7(where = (&groupvar = 0)) as b;
                  quit;

                  data temp_9;
                  set temp_8;
                  stddiff = compress(put(d,&stdfmt.));
                keep stddiff;
            run;

                  proc sql noprint;
                        select  stddiff into: std_value from temp_9;
                              insert into &outds.  values("&char_var.",
"&std_value.");
                  quit;

                  proc datasets lib = work nodetails nolist;
                        delete  temp_7 temp_8 temp_9;
                  quit;
      %end;
/* if a categorical variable  has more than two level such as a, b and c */
            %else %if &_mylevel_. > 2 %then %do;
                  %let _k_ = %eval(&_mylevel_. - 1);
                  %let _k_ = %sysfunc(strip(&_k_.));
                  data temp_7;
                        set catfreq;
                        by &groupvar.;
                        if last.&groupvar. then delete;
                        ColPercent = ColPercent/100;
                  run;

                  proc sql noprint;
```

```
                        select ColPercent into :tlist separated by ' '
                        from temp_7 where &groupvar. = 1;

                        select ColPercent into :clist separated by ' '
                        from temp_7 where &groupvar. = 0;
                  quit;

/* vector T, C and T-C */
                  data t_1;
                        array t{*}  t1- t&_k_.   (&tlist.);
                        array c{*}  c1- c&_k_.   (&clist.);
                        array tc{*} tc1 - tc&_k_. ;
                        do i = 1 to dim(t);
                        tc{i} = t{i} - c{i};
                        end;
                  drop i;
                  run;

/* each column has one element of a S covariance matrix (k x k) */

                  %let _dm = ;
                  %let _dm = %eval(&_k_.*&_k_.);
                  data covdata;
                        array t{*}  t1- t&_k_.  (&tlist.);
                        array c{*}  c1- c&_k_.   (&clist.);
                        array cv{&_k_.,&_k_.} x1 -x&_dm.;
                        do i = 1 to &_k_.;
                        do j = 1 to &_k_.;
                              if i = j then do;
                                          cv{i,j} = 0.5*(t{i}*(1-t{i}) +
c{i}*(1-c{i}));
                                          end;
                              else do;
                                          cv{i,j} = -0.5 * (t[i] * t[j] +
c[i] * c[j]);
                                          end;
                              if cv{&_k_.,&_k_.] ne . then output;
                        end;
                        end;
                  run;

                  proc transpose data = covdata(keep = x1 -x&_dm.) out =
covdata_1;
                  run;

                  data covdata_2;
                        set covdata_1;
                        retain id gp 1;
                        if mod(_n_ - 1,&_k_.) = 0 then gp = gp + 1;
                  run;

                  proc sort data = covdata_2 ;
                        by gp id;
```

```
                    run;

                    data covdata_3;
                          set covdata_2;
                          by gp id;
                          retain lp;
                          if first.gp then lp = 0;
                          lp = lp+1;
                    run;

/* transpose to a S variance-covariance matrix format */

                    data covdata_4;
                          set covdata_3;
                          retain y1-y&_k_.;
                          array cy{1:&_k_.} y1-y&_k_.;
                          by gp id;
                          if first.gp then do;
                          do k = 1 to &_k_.;
                                cy{k} = .;
                              end;
                          end;
                          cy{lp} = col1;
                          if last.gp then output;
                          keep y:;
                    run;

/* get inverse of S matrix */
                data A_1;
                 set covdata_4;
                 array _I{*} I1-I&_k_.;
                 do j=1 to &_k_.;
                  if j=_n_ then _I[j]=1;
                  else _I[j]=0;
                 end;
                 drop j;
                run;

/* solve the inverse of the matrix */

  %macro inv;
      %do j=1 %to &_k_.;
            proc orthoreg data=A_1 outest=A_inv_&j.(keep=y1-y&_k_.)
                  noprint singular=1E-16;
                  model I&j=y1-y&_k_. /noint;
            run;
            quit;
      %end;

            data A_inverse;
            set %do j=1 %to &_k_.;
            A_inv_&j
      %end;;
```

```
        run;
%mend;
%inv;

        proc transpose data=A_inverse out=A_inverse_t;
        run;

 /* calculate the mahalanobis distance */
        data t_2;
              set A_inverse_t;
              array t{*}  t1- t&_k_.  (&tlist.);
              array c{*}  c1- c&_k_.  (&clist.);
              i = _n_;
              trt = t{i};
              ctl = c{i};
              tc = t{i} - c{i};
        run;

        data t_3;
              set t_2;
              array aa{&_k_.} col1 - col&_k_.;
              array bb{&_k_.} bb1- bb&_k_.;
              do i = 1 to &_k_.;
              bb{i} = aa{i}*tc;
              end;
        run;

        proc summary data = t_3 ;
              var bb1-bb&_k_.;
              output out = t_4 sum =;
        run;

        data t_5;
              merge t_1 t_4;
              array d1{*} tc1- tc&_k_. ;
              array d2{*} bb1-bb&_k_.;
              array d3{*} y1-y&_k_.;
              do i = 1 to &_k_.;
                    d3{i} = d1{i}*d2{i};
              end;
              d = sqrt(sum(of y1-y&_k_.));
              stddiff = compress(put(d,&stdfmt.));
              keep stddiff;
        run;

        proc sql noprint;
              select  stddiff into: std_value from t_5;
              insert into &outds.  values("&char_var.", "&std_value.");
        quit;

        proc datasets lib = work nodetails nolist;
              delete  covdata covdata_1 covdata_2 covdata_3 covdata_4
              A_1 A_inverse A_inverse_t t_1 t_2 t_3 t_4 t_5
```

```
            A_inv_:;
            quit;
         %end;
      %end;
%end;

proc datasets lib = work nodetails nolist;
  delete Catfreq  Base_data temp_7;
quit;

proc print data = &outds.;
      title 'Calculated Standardized Difference';
run;

title;

%mend stddiff;
/*SIS MODELS*
%stddiff(diss.s4_sis_cmatch0,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH,
                        stdfmt=8.4,
                        outds=diss.s5_SIS_CMATCH0_SMD);

%stddiff(diss.s4_sis_cmatch25,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH,
                        stdfmt=8.4,
                        outds=diss.s5_SIS_CMATCH25_SMD);

%stddiff(diss.s4_sis_cmatch1,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH,
                        stdfmt=8.4,
                        outds=diss.s5_SIS_CMATCH1_SMD);

%stddiff(diss.s4_sis_dmatch,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH,
                        stdfmt=8.4,
                        outds=diss.s5_SIS_dMATCH_SMD);



/*SIS ESS MODELS*
```

```
%stddiff(diss.s4_sisESS cmatch0,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR Q106 Q108 Q111 Q153 Q156
Q157 Q159 Q1511 Q1516 Q1520 Q1521,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH
                        ESS liver ESS degreer ESS mathhad ESS mathneed ESS
scihad ESS scineed ESS writehad ESS writewil
                        ESS lang ESS religion ESS apcourse ,
                        stdfmt=8.4,
                        outds=diss.s5_SISESS CMATCH0_SMD);


%stddiff(diss.s4_sisESS cmatch25,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR Q106 Q108 Q111 Q153 Q156
Q157 Q159 Q1511 Q1516 Q1520 Q1521,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH
                        ESS liver ESS degreer ESS mathhad ESS mathneed ESS
scihad ESS scineed ESS writehad ESS writewil
                        ESS lang ESS religion ESS apcourse ,
                        stdfmt=8.4,
                        outds=diss.s5_SISESS CMATCH25_SMD);


%stddiff(diss.s4_sisESS cmatch1,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR Q106 Q108 Q111 Q153 Q156
Q157 Q159 Q1511 Q1516 Q1520 Q1521,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH
                        ESS liver ESS degreer ESS mathhad ESS mathneed ESS
scihad ESS scineed ESS writehad ESS writewil
                        ESS lang ESS religion ESS apcourse ,
                        stdfmt=8.4,
                        outds=diss.s5_SISESS CMATCH1_SMD);


%stddiff(diss.s4_sisESS dmatch,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR Q106 Q108 Q111 Q153 Q156
Q157 Q159 Q1511 Q1516 Q1520 Q1521,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH
                        ESS liver ESS degreer ESS mathhad ESS mathneed ESS
scihad ESS scineed ESS writehad ESS writewil
                        ESS lang ESS religion ESS apcourse ,
                        stdfmt=8.4,
                        outds=diss.s5_SISESS dMATCH_SMD);
```

```
/*SIS NCS Models*
%stddiff(diss.s4_SISNCS_cmatch0,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR SelfEff_Total
TimeManage_total swb_total familyob_total grit_total academiccontrol_total
                        CARING1     CARING2     CARING3     CARING4
      CARING5     CARING6     CARING7     CARING8     CARING9  ,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH,
                        stdfmt=8.4,
                        outds=diss.s5_SISNCS_CMATCH0_SMD);

%stddiff(diss.s4_SISNCS_cmatch25,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR SelfEff_Total
TimeManage_total swb_total familyob_total grit_total academiccontrol_total
                        CARING1     CARING2     CARING3     CARING4
      CARING5     CARING6     CARING7     CARING8     CARING9  ,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH,
                        stdfmt=8.4,
                        outds=diss.s5_SISNCS_CMATCH25_SMD);

%stddiff(diss.s4_SISNCS_cmatch1,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR SelfEff_Total
TimeManage_total swb_total familyob_total grit_total academiccontrol_total
                        CARING1     CARING2     CARING3     CARING4
      CARING5     CARING6     CARING7     CARING8     CARING9  ,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH,
                        stdfmt=8.4,
                        outds=diss.s5_SISNCS_CMATCH1_SMD);

%stddiff(diss.s4_SISNCS_dmatch,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR SelfEff_Total
TimeManage_total swb_total familyob_total grit_total academiccontrol_total
                        CARING1     CARING2     CARING3     CARING4
      CARING5     CARING6     CARING7     CARING8     CARING9  ,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS PLACEMENTWRITING PLACEMENTMATH,
                        stdfmt=8.4,
                        outds=diss.s5_SISNCS_dMATCH_SMD);

/*SIS ESS NCS MODELS*

%stddiff(diss.s4_SISESSNCS_cmatch0,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR Q106 Q108 Q111 Q153 Q156
Q157 Q159 Q1511 Q1516 Q1520 Q1521 SelfEff_Total TimeManage_total swb_total
familyob_total grit_total academiccontrol_total
```

```
                        CARING1     CARING2     CARING3     CARING4
      CARING5     CARING6     CARING7     CARING8     CARING9,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS FULL_WRITING PLACEMENTMATH
                        ESS liver ESS degreer ESS mathhad ESS mathneed ESS
scihad ESS scineed ESS writehad ESS writewil
                        ESS lang ESS religion ESS apcourse ,
                        stdfmt=8.4,
                        outds=diss.s5_SISESSNCS_CMATCH0_SMD);

%stddiff(diss.s4_SISESSNCS_cmatch25,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR Q106 Q108 Q111 Q153 Q156
Q157 Q159 Q1511 Q1516 Q1520 Q1521 SelfEff_Total TimeManage_total swb_total
familyob_total grit_total academiccontrol_total
                        CARING1     CARING2     CARING3     CARING4
      CARING5     CARING6     CARING7     CARING8     CARING9,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS FULL_WRITING PLACEMENTMATH
                        ESS liver ESS degreer ESS mathhad ESS mathneed ESS
scihad ESS scineed ESS writehad ESS writewil
                        ESS lang ESS religion ESS apcourse ,
                        stdfmt=8.4,
                        outds=diss.s5_SISESSNCS_CMATCH25_SMD);

%stddiff(diss.s4_SISESSNCS_cmatch1,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR Q106 Q108 Q111 Q153 Q156
Q157 Q159 Q1511 Q1516 Q1520 Q1521 SelfEff_Total TimeManage_total swb_total
familyob_total grit_total academiccontrol_total
                        CARING1     CARING2     CARING3     CARING4
      CARING5     CARING6     CARING7     CARING8     CARING9,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS FULL_WRITING PLACEMENTMATH
                        ESS liver ESS degreer ESS mathhad ESS mathneed ESS
scihad ESS scineed ESS writehad ESS writewil
                        ESS lang ESS religion ESS apcourse ,
                        stdfmt=8.4,
                        outds=diss.s5_SISESSNCS_CMATCH1_SMD);*/


%stddiff(diss.s4_SISESSNCS_dmatch,
                        groupvar=f1_15,
                        numvars=SACTE SACTM SHSGPAR Q106 Q108 Q111 Q153 Q156
Q157 Q159 Q1511 Q1516 Q1520 Q1521 SelfEff_Total TimeManage_total swb_total
familyob_total grit_total academiccontrol_total
                        CARING1     CARING2     CARING3     CARING4
      CARING5     CARING6     CARING7     CARING8     CARING9,
                        charvars=GENDER ETHNIC HONCOLL PELL COLLEGE SUMMCOLL
HSCPS FULL_WRITING PLACEMENTMATH
                        ESS liver ESS degreer ESS mathhad ESS mathneed ESS
scihad ESS scineed ESS writehad ESS writewil
                        ESS lang ESS religion ESS apcourse ,
```

```
                        stdfmt=8.4,
                        outds=diss.s5_SISESSNCS_dMATCH_SMD);
```

Step Six: Average Treatment Effect

```
/*http://www.stat.purdue.edu/~tqin/system101/method/method_mcnemar_sas.htm*/
/*http://www.sascommunity.org/mwiki/images/9/9a/Propensity_Score_Methods_in_S
AS.pdf*/
/*need to restructure dataset so that the items are paired*/

/*DIGIT MATCHING MACRO*/

%macro cat (inds);
*Restructure your data first!;
data OPTIMAL NOTOPTIMAL;
  set &inds;
  if f1_15 = 1 then output OPTIMAL;
  if f1_15 = 0 then output NOTOPTIMAL;
run;

proc sort data=OPTIMAL;
by    matchto;
run;

proc sort data=NOTOPTIMAL;
by matchto;
run;

data &inds._matched;
merge       optimal(rename = (f2_reg = retT))
            notoptimal(rename = (f2_reg = retC)) ;
by matchto;
run;


proc freq data=&inds._matched;
  tables  retT*retC /agree expected ;
  title "McNemar'stest for comparing outcomes among matched pairs &INDS";
run;

%mend cat;
%cat (DISS.S4_SIS_dMATCH);
%cat (DISS.S4_SISESS dMATCH);
%cat (DISS.S4_SISNCS_dMATCH);
%cat (DISS.S4_SISESSNCS_dMATCH);

/*greedy matching caliper macro*/
%macro cat (inds);
*Restructure your data first!;
data OPTIMAL NOTOPTIMAL;
  set &inds;
```

```
  if f1_15 = 1 then output OPTIMAL;
  if f1_15 = 0 then output NOTOPTIMAL;
run;

proc sort data=OPTIMAL (RENAME=(__IDCA=MATCHTO));
by MATCHTO;
run;

proc sort data=NOTOPTIMAL (RENAME=(__IDCA=MATCHTO));
by MATCHTO;
run;

data &inds._matched;
merge       optimal(rename = (f2_reg = retT))
            notoptimal(rename = (f2_reg = retC)) ;
by matchto;
run;

proc freq data=&inds._matched;
  tables  retc*rett /agree expected ;
  title "McNemar'stest for comparing outcomes among matched pairs &INDS";
run;

%mend cat;

%cat (DISS.S4_SIS_CMATCH0);
%cat (DISS.S4_SIS_CMATCH25);
%cat (DISS.S4_SIS_CMATCH1);

%cat (DISS.S4_SISESS CMATCH0);
%cat (DISS.S4_SISESS CMATCH25);
%cat (DISS.S4_SISESS CMATCH1);

%cat (DISS.S4_SISNCS_CMATCH0);
%cat (DISS.S4_SISNCS_CMATCH25);
%cat (DISS.S4_SISNCS_CMATCH1);

%cat (DISS.S4_SISESSNCS_CMATCH0);
%cat (DISS.S4_SISESSNCS_CMATCH25);
%cat (DISS.S4_SISESSNCS_CMATCH1);
```

Step Seven: Sensitivity

```
/*sensitivity test
%let a= # of matched pairs in which exactly one has the outcome (AKA
DISCORDANT PAIRS);
%let b= # of discordant pairs where Treated has outcome;*/

%macro sens(a,b,title);
data g;
do gamma_init= 0 to 50;
```

```
gamma = 1 + gamma_init/10;
p_plus = gamma/(1 + gamma);
p_neg = 1/(1 + gamma);
p_upper = 2*(1 - probbnml(p_plus,&a, &b) );
p_lower = 2*(1 -probbnml(p_neg,&a,&b )  ) ;
output;  end; run;
proc print data=g noobs;
var gamma p_lower p_upper;
title "Sensitivity analysis for McNemar's test &title";
run;
%mend sens;
/*sis matches*/
%sens(240,160,DISS.S4_SIS_CMATCH0);
%sens(236,152,DISS.S4_SIS_CMATCH25);
%sens(251,157,DISS.S4_SIS_CMATCH1);
%sens(234,142,DISS.S4_SIS_DMATCH);

/*sis ess matches*/
%sens(163,104,DISS.S4_SISESS_CMATCH0);
%sens(162,97,DISS.S4_SISESS_CMATCH25);
%sens(163,100,DISS.S4_SISESS_CMATCH1);
%sens(159,93,DISS.S4_SISESS_DMATCH);

/*sis ncs matches*/
%sens(164,100,DISS.S4_SISNCS_CMATCH0);
%sens(167,98,DISS.S4_SISNCS_CMATCH25);
%sens(167,97,DISS.S4_SISNCS_CMATCH1);
%sens(152,92,DISS.S4_SISNCS_DMATCH);

/*sis ess ncs matches*/
%sens(106,64,DISS.S4_SISESSNCS_CMATCH0);
%sens(115,67,DISS.S4_SISESSNCS_CMATCH25);
%sens(103,60,DISS.S4_SISESSNCS_CMATCH1);
%sens(94,53,DISS.S4_SISESSNCS_DMATCH);
```

REFERENCE LIST

Ali, M. S., Groenwold, R. H., Belitser, S. V., Pestman, W. R., Hoes, A. W., Roes, K. C. B., Klungel, O. H. (2015). Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *Journal of Clinical Epidemiology*, *68*(2), 112.

An, B. P. (2013). The Impact of Dual Enrollment on College Degree Attainment Do Low-SES Students Benefit? *Educational Evaluation and Policy Analysis*, *35*(1), 57–75. http://doi.org/10.3102/0162373712461933

Angrist, P.D. & Pischke, J.S. (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton University Press.

Austin, P. C. (2007). The performance of different propensity score methods for estimating marginal odds ratios. *Statistics in Medicine*, *26*(16), 3078–3094. http://doi.org/10.1002/sim.2781

Austin, P. C. (2008). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine, 27*(12), 2037–2049. http://doi.org/10.1002/sim.3150

Austin, P. C. (2009). Type I Error Rates, Coverage of Confidence Intervals, and Variance Estimation in Propensity-Score Matched Analyses. *The International Journal of Biostatistics*, *5*(1). http://doi.org/10.2202/1557-4679.1146

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399–424. http://doi.org/10.1080/00273171.2011.568786

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, *33*(6), 1057–1069. http://doi.org/10.1002/sim.6004

Bergstralh, E.J., & Kosanke, J.L. (1995). Computerized matching of controls. *Technical Report Series No. 56*, (Department of Health Science Research, Mayo Clinic, Rochester).

Cochran, W. G. (1968). Errors of measurement in statistics. *Technometrics*, *10*(4), 637–666. http://doi.org/10.2307/1267450

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: a review. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, *35*(4), 417–446.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis & Management*, *27*(4), 724–750. http://doi.org/10.1002/pam.20375

Cox, D. R. (1958). *Planning of experiments*. New York, Wiley.

Daniel F. McCaffrey, Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, *9*(4), 403–425. http://doi.org/0.1037/1082-989X.9.4.403

Dehejia, R. H., & Wahba, S. (1999). Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, *94*(448), 1053–1062. http://doi.org/10.2307/2669919

Doyle, W. R. (2011). Effect of increased academic momentum on transfer rates: An application of the generalized propensity score. *Economics of Education Review*, *30*(1), 191–200. http://doi.org/10.1016/j.econedurev.2010.08.004

Fisher, R. A. (1925). *Statistical methods for research workers* (1st ed.). Edinburgh: Oliver and Boyd.

Guo, S., & Fraser, W. M. (2015). *Propensity score analysis: statistical methods and applications*. Thousand Oaks, CA: Sage Publications, Inc.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity Score Techniques and the Assessment of Measured Covariate Balance to Test Causal Associations in Psychological Research. *Psychological Methods, 15*(3), 234–239.

Heil, S., Reisel, L., & Attewell, P. (2014). College Selectivity and Degree Completion. *American Educational Research Journal, 51*(5), 913–935. http://doi.org/10.3102/0002831214544298

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945–960. http://doi.org/10.2307/2289064

Keller, R. R., & Lacy, M. G. (n.d.). Propensity score Analysis of an Honors Program's Contribution to students' Retention and Graduation outcomes. *Journal of the National Collegiate Honors Council*.

Kot, F. C. (2014). The Impact of Centralized Advising on First-Year Academic Performance and Second-Year Enrollment Behavior. *Research in Higher Education, 55*(6), 527–563. http://doi.org/10.1007/s11162-013-9325-4

Leon, A. C., & Hedeker, D. (2007). Quantile Stratification Based on a Misspecified Propensity Score in Longitudinal Treatment Effectiveness Analyses of Ordinal Doses. *Computational Statistics & Data Analysis*, *51*(12), 6114–6122. http://doi.org/10.1016/j.csda.2006.12.021

Little, R. J., & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annual Review of Public Health*, *21*(1), 121.

Luellen, J. K., Shadish, W. R., & Clark, M. H. (2005). Propensity scores: an introduction and experimental test. *Evaluation Review*, *29*(6), 530–558. http://doi.org/10.1177/0193841X05275596

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. *Journal of Higher Education*, *82*(3), 265–291.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: methods and principles for social research*. Cambridge University Press.

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: improving causal inference in educational and social science research*. Oxford; New York: Oxford University Press.

Parsons, L. S. (2000.). SUGI 26: Reducing Bias in a Propensity Score Matched-Pair Sample Using Greedy Matching Techniques - p214-26.pdf. Retrieved July 10, 2016, from http://www2.sas.com/proceedings/sugi26/p214-26.pdf

Pascarella, E. T. (2005). *How college affects students: a third decade of research* (1st ed. .). San Francisco: San Francisco: Jossey-Bass.

Pearl, J. (2009). Causal inference in statistics: an overview. *Statistical. Survey. 3 (2009), 96-146.*

Pearl, J. (2010). The foundations of causal inference. *Sociological Methodology*, *40*, 75–149.

Peikes, D. N., Moreno, L., & Orzol, S. M. (2008). Propensity Score Matching: A Note of Caution for Evaluators of Social Programs. *The American Statistician*, *62*(3), 222–231.

Rosenbaum, P. R. (2002). Attributing Effects to Treatment in Matched Observational Studies. *Journal of the American Statistical Association*, *97*(457), 183–192.

Rosenbaum, P. R., & Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, *45*(2), 212–218.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. http://doi.org/10.2307/2335942

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*(387), 516–524. http://doi.org/10.2307/2288398

Rosenbaum, P.R. (2005). Sensitivity analysis in observational studies. In B.S. Everitt & D.C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1809–1814). New York, NY: John Wiley.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*, *6*(1), 34–58.

Rubin, D. B. (1980). Randomization analysis of experimental data: the fisher randomization test comment. *Journal of the American Statistical Association*, *75*(371), 591–593. http://doi.org/10.2307/2287653

Rubin, D. B. (1990). [On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.] Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science*, *5*(4), 472–480.

Schafer, M. H., Wilkinson, L. R., & Ferraro, K. F. (2013). Childhood (mis)fortune, educational attainment, and adult health: contingent benefits of a college degree? *Social Forces*, *91*(3), 1007–1034.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? a randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, *103*(484), 1334–1344. http://doi.org/10.1198/016214508000000733

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston [u.a.]: Houghton Mifflin.

Shah, B. R., Laupacis, A., Hux, J. E., & Austin, P. C. (2005). Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *Journal of Clinical Epidemiology*, *58*(6), 550–559. http://doi.org/10.1016/j.jclinepi.2004.10.016

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, *15*(3), 250–267. http://doi.org/10.1037/a0018719

Steiner, Peter M., & Cook, David L. (2013). Matching and propensity scores. In Little, T. D. (Ed.), *The Oxford Handbook of Quantitative Methods* (Vol. Volume I, Foundations). New York, NY: Oxford University Press.

Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science*, *25*(1), 1–21.

Stürmer, T., Joshi, M., Glynn, R. J., Avorn, J., Rothman, K. J., & Schneeweiss, S. (2006). A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *Journal of Clinical Epidemiology*, *59*(5), 437–447. http://doi.org/10.1016/j.jclinepi.2005.07.004

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, *51*(6), 309–317.

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, *46*(1), 90–118. http://doi.org/10.1080/00273171.2011.540475

Ting, S. R. (1998). Predicting first-year grades and academic progress of college students of first-generation and low-income families. *Journal of College Admission*, (158), 14–23.

Tinto, V. (1975). Dropout from higher education: a theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), 89–125. http://doi.org/10.2307/1170024

Tinto, V. (1993). *Leaving college: rethinking the causes and cures of student attrition* (2nd ed.). Chicago; London: University of Chicago Press.

Vaughan, A. L., Lalonde, T. L., & Jenkins-Guarnieri, M. A. (2014). Assessing student achievement in large-scale educational programs using hierarchical propensity scores. *Research in Higher Education*, *55*(6), 564–580. http://doi.org/10.1007/s11162-014-9329-8

Watkins, S., Jonsson-Funk, M., Brookhart, M. A., Rosenberg, S. A., O'Shea, T. M., & Daniels, J. (2013). An Empirical Comparison of Tree-Based Methods for Propensity Score Estimation. *Health Services Research*, *48*(5), 1798–1817. http://doi.org/10.1111/1475-6773.12068

Westreich, D., Lessler, J., & Funk, M. J. (2010). Propensity score estimation: machine learning and classification methods as alternatives to logistic regression. *Journal of Clinical Epidemiology*, *63*(8), 826–833. http://doi.org/10.1016/j.jclinepi.2009.11.020

Zhao, Z. (2004). Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence. *The Review of Economics and Statistics*, *86*(1), 91–100.

VITA

Julie Wren received her bachelor degree in Psychology from Saint Ambrose University. She went on to complete her master's degree in Psychology with a Clinical Science emphasis from the University of Northern Iowa. After completing her training in clinical psychology, Julie began working in higher education. During this time, she decided to continue her graduate studies and pursued a doctoral degree from Loyola University Chicago in Research Methodology. Julie's research interest centers on improving casual claims in non-experimental research and she continues to work in higher education. In her applied work, she focuses on assessment and measurement.