What is the Impact of the Underserved Pathway program on

Entering an Underserved Family Medicine Residency?

A Comparison of Three Approaches for Estimating the Average Treatment Effect

Amanda Kost

A thesis submitted in partial fulfillment

of the requirements

for the degree of

Master of Education

University of Washington

2016

Committee:

Elizabeth A. Sanders, Chair

Robert D. Abbott

Program Authorized to Offer Degree:

College of Education

University of Washington

Abstract

What is the Impact of the Underserved Pathway program on

Entering an Underserved Family Medicine Residency?

A Comparison of Three Approaches for Estimating the Average Treatment Effect

Amanda Kost

Chair of the Supervisory Committee:

Dr. Elizabeth A. Sanders, Assistant Professor

Educational Psychology

Introduction: It is well known that more primary care physicians are needed in underserved areas of the U.S. While prior research has shown that medical student experiences in underserved settings helps to increase the likelihood that they will ultimately choose to become physicians serving underserved areas, none have controlled for selection bias. In addition, none have examined intermediate steps in medical training experiences, such as how specific experiences relate to residency choices (the pipeline to eventual clinical practice). One manner of strengthening causal claims includes controlling for covariates that are correlated with treatment (program) and outcome variables using regression approaches, but this can have dimensionality or multicollinearity problems. Another approach to estimating less biased treatment effects involves propensity score (PS) methods, which are reduce the problems in the regression approach by employing a single score that captures the correlations between treatment status and all covariates. The primary aim of the present study was to use a set of methods to estimate whether the University of Washington (UW) School of Medicine's longitudinal extracurricular experience, *Underserved Pathway*, impacts graduates' choice in completing their residency in an underserved area. A secondary aim was to evaluate which of the analytic methods (and results) is most defensible.

Methods: Extant data from one cohort of $N = 158$ UW medical students from matriculation surveys conducted between 2004 and 2011 were used for this project. Three popular approaches to estimating the *Underserved Pathway* program's effect on underserved setting residency choice were employed, including multiple logistic regression, PS 1:n Matching with replacement, and inverse weighted probability (IWP) regression.

Results: Average treatment (program) effects from the three approaches ranged from 12.0% of *Underserved Pathway* graduates choosing an underserved area for their residency (using the IPW method), 17.2% (using logistic regression), to 23.4% (using PS matching); the latter two were statistically significantly greater than zero by a Wald test. Tests of the covariate balance (i.e., the extent to which the ignorability assumption held) showed that PS matching offered better covariate balance than IPW for metrical covariates but that no other differences between methodologies on covariate balance were found.

Discussion: Completion of the *Underserved Pathway* resulted in a significant (17.2% - 23.4%) increase in program graduates matching to a residency in an underserved setting according to the logistic regression and PS matching approaches; these methods are preferred given that neither differed in covariate balance from each other, and further, PS matching was superior in balance across groups over IPW on one covariate. Additionally, given that the logistic regression approach does not delete any cases, it seems likely that the logistic regression approach is the method that is most defensible in reducing selection bias in this situation. Selection of a method should address covariate balance with simplicity of approach with robust and transparent reporting to allow for assessment of any causal claims.

TABLE OF CONTENTS

Chapter 1. Introduction 1

1.1 Primary Care Physicians in Underserved Settings 1
1.2 Career Choice Models of U.S. Medical Students 2
1.3 Existing Research on Links Between Residency Characteristics and Eventual Practice .. 3
1.4 Challenges in Outcomes Research in Medical Education 4
1.5 The University of Washington School of Medicine and the Underserved Pathway 5
1.6 UP Participation and Causality 6
1.6.1 Rubin's Causal Model 6
1.6.2 Assumptions – Stable Unit Treatment Value Assumption 8
1.6.3 Assumptions – Ignorability 9
1.7 Propensity Analysis Methods 10
1.8 Goals of The Present Study 13

Chapter 2. Methods 15

2.1 Subjects 15
2.2 Data Sources 15
2.2.1 Underserved Pathway Records 15
2.2.2 UWSOM Records 15
2.2.3 AAFP Residency Database and Critical Access Hospital List 16
2.3 Covariates 17
2.3.1 Covariates used in Multiple Logistic Regression and to Create Propensity Score ... 17
2.3.2 Treatment Variable 20
2.3.3 Outcome Variable 21
2.4 Analytic Plan 21
2.4.1 Sample Description 21
2.4.2 Statistical Approaches 21
2.4.3 Covariate Balance Checks 22
2.4.4 Sensitivity Analysis 23

Chapter 3. Results 24

3.1 Sample Description 24
3.2 Approach Outcomes 25
3.3 Covariate Balance Checks for PSM and IPW 26
3.4 Sensitivity Analysis 27

Chapter 4. Discussion 29

4.1 Comparison of Different Approaches 29
4.2 Underserved Pathway Completion and Underserved Residency Match 31

Tables and Figures 34

Bibliography 47

Appendix A – Survey Instrument 51

Appendix B – Stata Code 57

ACKNOWLEDGEMENTS

I would like to thank my former chair, Tom Norris, who encouraged me down this path and supported me in undertaking this degree. I also wish to express my appreciation to Doug Schaad and the Underserved Pathway faculty and staff for providing the data and support for this thesis project. I am grateful for the warm welcome I received from the College of Education, as an unusual student taking a long time to complete my degree. I am deeply indebted to my thesis advisor, Dr. Liz Sanders, who took me under her wing and shared her wisdom not only about statistics but also about parenting and life in general. Last but not least, I thank my husband, who always believed I could do the math.

DEDICATION

To Lonan and Jaelithe, who still greeted me with smiles even when I got home late from class.

Chapter 1. INTRODUCTION

The United States suffers from shortages of primary care physicians, most acutely those caring for underserved populations (Doescher, M. Fordyce, M. Skillman, S. Jackson, J. Rosenblatt, R., 2009). Expansion of the federal programs of community health centers and increased demand for primary care through the Affordable Care Act has exacerbated these workforce deficits. Prior research suggests medical student participation in programs focusing specifically on rural and urban underserved populations, as well as residency training in underserved settings such as community health centers, are both associated with increasing those who practice in underserved settings. However, prior research has not yet examined the intermediate steps in training (i.e., training prior to the residency experience) that might later yield increases in number of physicians ultimately choosing to serve in rural, underserved settings. Hence, the present study examines the estimated impact of a specific training experience for medical students, called the *Underserved Pathway* program, on students' future residency choice.

1.1 PRIMARY CARE PHYSICIANS IN UNDERSERVED SETTINGS

The United States continues to face substantial shortages in the primary care workforce, as reported in the Council of Graduate Medicine Education's 20th report. (Council on Graduate Medical Education, 2010) This deficit is particularly acute for primary care physicians (PCP) caring for vulnerable and underserved populations both in rural and urban areas in federal programs collectively known as community health centers (CHCs) (Doescher, M. Fordyce, M. Skillman, S. Jackson, J. Rosenblatt, R., 2009). Even prior to the Affordable Care Act (ACA), Rosenblatt et al. warned that the expansion of the federal programs targeting these populations, including federally qualified health centers (FQHCs) and rural health centers (RHCs) would face extensive challenges in finding enough physicians to staff these clinics (Rosenblatt, Andrilla, Curtin, & Hart, 2006). Given these pre-existing shortages of physicians and increased demand for primary care services through the Medicaid expansion of the ACA, the US needs more physicians to pursue primary care careers caring for underserved patients. The pipeline of primary care physicians that ultimately end up caring for underserved populations begins prior to matriculation to medical school, stretches across medical school and residency training, and ends when those physicians pursue jobs and careers that offer direct patient care to those underserved populations.

1.2 Career Choice Models of U.S. Medical Students

Several medical schools have specialized educational tracks or experiences for students who matriculate with interest or attributes that make them more likely to pursue an underserved career. Longitudinal rural tracks are the best studied and participation in these is linked to eventual rural practice (Rabinowitz, Diamond, Markham, & Wortman, 2008). A track for focusing on the care of urban underserved populations has also been associated with eventual careers caring for underserved populations (Ko, Heslin, Edelstein, & Grumbach, 2007). Laudably, these studies have focused long-term practice outcomes of students as opposed to more proximal outcomes such as residency characteristics. However, given the urgent need for increased number of underserved PCPs, waiting more than a decade to have some determination of the outcome of medical student programs is not currently feasible.

1.3 Existing Research on Links Between Residency Characteristics and Eventual Practice

Several studies over the past decade have examined the association of training in a residency that exposes physicians to underserved practice settings and their eventual clinical practice in a similar setting. A study of family medicine residencies in a network of four states (Washington, Alaska, Montana, and Idaho), found that residents who trained at CHCs compared to non-CHCs were nearly three times more likely to report practicing in a CHC after residency (Morris, Johnson, Kim, & Chen, 2008). This landmark study included 17 residencies but was limited by selection bias and did not control for any confounders – in other words, *residents selecting a CHC training environment may be more likely to practice at a CHC when done, simply because they prefer this environment and not because of the impact of training in a CHC during residency*. Subsequent similar research looking at the practice outcomes for a single residency program with three different continuity clinics found a similar association between the CHC residency clinic site and underserved practice after residency (Ferguson, Cashman, Savageau, & Lasser, 2009). This more recent study did use regression to control for covariates such as the demographics of the subjects and their initial interest and experience in underserved practice. These associations are consistent with a more recent Graham Center report on family medicine residencies located at community health centers, which found that graduates of teaching health center residencies were more likely to plan to practice in underserved settings 33% compared to graduates of traditional residencies 18% (Bazemore et al., 2015)However, this study also did not control for selection bias.

Finally, a nationwide study of all residency training programs found that residency exposure to any underserved setting (including FQHCs, RHCs, and critical access hospitals

(CAHs)) was associated with physicians being more likely to have insurance (Medicare billing) claims from underrepresented areas after residency (Phillips, Petterson, & Bazemore, 2013). Despite these hopeful findings, the study findings are limited by use of administrative data location (Medicare billing) as a proxy outcome for being a practitioner in an underserved setting; further, there was an unknown quantity and quality of training exposure to the different settings for the residents (i.e., we can't know the "dosage" of the exposure).

1.4 Challenges in Outcomes Research in Medical Education

While the research on links between medical school, residency experiences and practice area choice was occurring, a discussion had commenced on what medical education research (both medical school and residency) should focus on, particularly with respect to outcomes. Highlighting the fact that much of the published medical education research was limited to learner perceptions or knowledge, some have called for outcomes-based research that investigates the ultimate clinical care these programs are preparing learners to practice (Chen, Bauchner, & Burstin, 2004). Nevertheless, it is very difficult to identify causal or even correlational relationships between these variables given the length and complexity of physician training. By the time a physician begins practice, she or he will have completed four years of medical school and a minimum of three years of residency. These years often occur in different locations throughout the country, and under a model of training that divides experiences into 4- to 12-week blocks across the multiple clinical years. Recently, an alternative for this particular outcome model recommended investigating linkages among stages of training, as illustrated in Figure 1 (Cook & West, 2013)The argument is this: if causal or correlational relationships can be identified for smaller training components of shorter durations, then these components can then be linked together into a larger theoretical whole, with the ultimate endpoint being clinical care. Given the urgent need to increase primary care physicians in underserved areas, one tactic for identifying experiences within this "pipeline" that links to subsequent stages of training may be important information going forward.

1.5 The University of Washington School of Medicine and the Underserved Pathway

The University of Washington's School of Medicine (UWSOM) is the only allopathic medical school for the 5-state region of Washington, Wyoming, Alaska, Montana, and Idaho. Students spend their preclinical time at one of six regional campuses in these states, and then experience their required and elective clerkships in communities across the five states. The UWSOM offers a 4-year, extracurricular longitudinal experience designed to support student interest in caring for underserved populations called *Underserved Pathway* (UP). Approximately 240 students matriculate each year to the UWSOM; of those enrolled, in 2016 approximately 20% participated in UP.

UP began in 2006 and includes eight components – an online curriculum, preclinical preceptorships at an underserved site, electives on underserved topics, mentorship with a physicians with experience caring for underserved populations, service learning, a scholarly project with an underserved focus, and clinical clerkships at underserved sites. Students enroll voluntarily in UP in their first two years of medical school. Students must complete all UP components to be recognized with a certificate of completion at graduation; there are a handful of students who enroll in UP but do not complete all components (approximately 5-10% each year). One goal of UP is to maintain medical student interest in caring for underserved populations such that they feed into residency programs that offer training to support a career actually caring for these populations.

1.6 UP Participation and Causality

Because most the present study's students self-selected into UP, and all students who apply are accepted, there is strong concern regarding selection bias when attempting to estimate the true career outcomes of UP participants compared to non-participants. Interestingly, most medical education research on extracurricular programs has focused largely on student perceptions of the experience, rather than observable behaviors. Even when outcomes, such as medical student specialty choice, are examined, known confounders, such as the age or gender of the student, are frequently not accounted for in statistical models. However, even when these known confounders are included, they do not necessarily fully address the issue of selection bias. Concerns about remediating selection bias in quasi-experiments (i.e., when subjects are not randomized to conditions), particularly for studies with a large number of possible confounders (covariates that are correlated with both the treatment and the outcome), has led to the development of propensity score analysis methods.

1.6.1 *Rubin's Causal Model*

The gold standard of research design to supports claims of causality is the "true" experimental approach in which subjects are randomized to conditions in which every subject has an equal probability of being in any condition (often to "treatment" and no-treatment "control"). The result of randomization is that the groups formed should be probabilistically the same across all possible confounders (covariates) both observed and unobserved. In other words, the only difference between the groups is the treatment each received, and thus differences between groups detected in the outcome can be causally linked to differences between the treatments. However, in medical education research, this approach is often not feasible due to the resources needed to conduct such an experiment. Thus, we are left with a quasi-experimental approach, sometimes-coined "observational" studies, in which subjects do experience a particular treatment that they have chosen or assigned to ad-hoc (perhaps by an administrator), and thus the mechanism that caused the self-selection or assignment may also be causing the outcome. In other words, there is no way to untangle the effects of individual attributes (confounders) and the treatment condition itself on a given outcome.

Rubin's Causal Model uses the concept of counterfactuals to better articulate how causal effects manifest (Guo & Fraser, 2014). Briefly, it is assumed that each subject that is a control actually has two *potential* outcomes – the outcome they would have if they received the control condition (observable) and the outcome they would have if they received the treatment condition (not observable, they did not receive it). The same holds true for treatment subjects – their unobserved potential outcome is the outcome they would have if they had received the control condition (they did not receive it) and their observed potential outcome is having received the treatment condition. However, we cannot know the "counterfactual" for every subject; instead, we have only the observed outcome for each. For example, in the present study, medical school graduates cannot experience both not completing the UP program (control) and completing it (treatment); nevertheless, we can observe the *average* effect of those receiving treatment and those receiving control, and the difference between these averages is statistically expected, in the long run, to be the caused by the treatment so long as two assumptions are met. These assumptions are the *stable unit treatment value assumption* and *ignorability of treatment* *assignment* (Guo & Fraser, 2014), and it is the latter of the two assumptions that is always problematic for quasi-experiments.

1.6.2 *Assumptions – Stable Unit Treatment Value Assumption*

The stable unit treatment value assumption (SUTVA) is tenable if the outcome of treatment on one subject is not impacted by the treatment assignment of other subjects. This should not be confused with the statistical assumption of independence, which states that treatment must be experienced independently by each subject without any sort of cluster effects. SUTVA indicates that, over time, the treatment prescribed within each condition should be the same for all subjects within that group, and that group assignment should not have been modified by what other subjects in other conditions have received (in many randomized studies, great effort is made to "blind" subjects to their treatment, as well as the research assistants in the case of "double-blind" studies). More simply stated, treatment should be defined the same way for all subjects, and subjects who participate in one group (such as UP, treatment) should not contaminate the experience of the other group (control). In the present study, the treatment is completion of UP and the requirements to complete UP have not changed since it began in 2006. To ensure that all students received the same UP treatment (i.e., to satisfy the first part of the SUTVA requirement), we only include subjects that graduated after 2010 who would had been exposed to four full years of UP treatment during their time in medical school. Because the structure of graduate medical education involves a national process that matches a new cohort of medical school graduates to the first year of the training program each year), the treatment of prior students will not have had an impact on the matching of incoming students.

In fact, the only way that SUTVA would not be satisfied in this case would be if the control group subjects became aware of and interested in the UP experience, particularly if it was perceived by peers as providing certain rewards or other advantages, and then these students (not just one, but many) found a way to have their own UP-like experience (not necessarily directly the UP program). If that was the case, then the difference between groups on who chooses an underserved residency (the outcome) would no longer be a true comparison of treatment and no-treatment because the control group would be said to have experienced interference. Given the notoriously demanding schedules medical students experience, it would be remarkably unlikely for this to be any sort of problem in the current study.

1.6.3 *Assumptions – Ignorability*

The second and more difficult assumption to meet in observational studies is ignorability. Ignorability holds that the treatment assignment process is independent of (uncorrelated with) the outcome. In experimental design, where subjects are randomly distributed to treatment or control conditions, this assumption holds since every subject has an equal probability in any condition, and therefore there are no expected pre-existing differences among the groups on any variable, probabilistically speaking. The subject *assignment mechanism* is completely unrelated to the outcome. This translates to the idea of "balance" which indicates that the groups are balanced in their distributions across all measurable covariates. In other words, there should be no statistically significant differences between groups on any characteristic prior to study onset, including demographic characteristics or previous aptitude measures. However, in observational studies, subjects with certain characteristics, such as minority status or undergraduate major, may be more likely to self-select into a certain condition, such as participating in the UP program. As such, the two conditions (participating or not participating in UP) are likely to be unbalanced across certain covariates, and in all likelihood, these covariates that exhibit imbalance are likely also related to the outcome: students from minority backgrounds or certain majors may be more likely to choose a residency in an underserved area, regardless of their participation in UP. Of course, this is a violation of the ignorability assumption, unless we somehow control for the covariate differences.

Multiple regression is the most often used method for estimating unbiased treatment effects via direct statistical control of covariate imbalance. However, these models are not estimable if (a) there is too a large number of covariates (potential confounders) relative to the sample size, which is called a dimensionality problem, or (b) the covariates selected for the model are too highly correlated with each other causing unstable standard error estimates, also known as a multicollinearity problem. Hence, a better way to reduce selection bias may be to use propensity score (PS) methods. This class of methods employs various approaches to achieve covariate balance in an attempt to compute average treatment effect estimates in which the ignorability assumption is more tenable just as the traditional multiple regression method does, but without the worry of dimensionality or multicollinearity problems.

1.7 Propensity Analysis Methods

Propensity Score (PS) methods encompass a variety of approaches to estimating unbiased (or reduced bias) treatment effects. To be clear, for brevity this paper focuses solely on average treatment effects (ATE), rather than its cousin, the average treatment effects on the treated (ATT; involves a set of methods with only one treated group, rather than a treated and an untreated group). Specifically, there are two primary PS approaches: one that employs some form of "matching" on a propensity score (PS), and one that uses the PS as a "weight" in the multiple regression model (similar to weights used in complex survey samples). In both cases, a propensity score itself is the predicted probability (or natural log of the odds, known as a logit) that a subject should have been assigned to the treatment condition (however treatment is defined), given a set of covariates. In other words, the PS is a kind of "composite" score across all the covariates that corrects for covariate imbalance leading to selection bias. Finally, in both PS matching and PS weighting approaches, a 2-step approach is employed in which (1) a propensity score must be estimated using logistic regression, and (2) the PS is used to conduct an analysis to test the treatment effect on the outcome as described by Stuart (2010) and later by Guo and Fraser (2014). Below I describe each of the two methods in more detail.

Although propensity score (PS) analyses that employ matching can vary in procedure depending on the goals of the study and the conditions of the data available, most follow five basic steps (Guo & Fraser, 2014)The sample used must include two groups of subjects (a "treatment" condition and a "control" condition). In the present study, the treatment will be defined as UP program participation, and the control as no-UP participation.

1. As many pre-treatment variables as possible (covariates) are used to predict the probability the subjects are from the treatment group; in other words, their "propensity" for being in the treatment condition, and this predicted value (a propensity score) is saved for each subject. Recall that this PS score is much like a composite covariate that takes into account group differences across all of the measured covariates.
2. Subjects from the treatment condition are then matched to one or more control subjects with a similar PS (the PS is often converted into logits a priori). Note that various algorithms exist to perform matching, including matching with or without replacement of control subjects. Further, the matching algorithms can result in a handful of treatment subjects not having a control match, and some

control subjects not being matched to a treatment subject (most algorithms focus on maximizing the number of treatment subjects matched, meanwhile discarding some control subjects).. The propensity score matching algorithm used in the present was “nearest neighbor” matching with replacement, which meant that control subjects were matched to a treatment subject if they were within 0.3 logits (caliper width) on the PS, and a control subject was used more than once if it fit with more than one treatment subject (this is the 1:n part of the algorithm).

3. After matching, comparison analyses are conducted to ensure that the matched groups are equally balanced on the covariates to keep the ignorability assumption tenable (which is what matching hopes to achieve).
4. Statistical analyses are performed to compare the matched groups on outcomes.
5. A final sensitivity analysis is performed to examine any likelihood that there are possible confounders that were not identified in the initial propensity model (that might cause omitted variable bias leading to non-ignorability).

Inverse probability weighting (IPW) takes a somewhat different approach. This method also uses the first step described for the PS matching method above via logistic multiple regression. The range of PS scores of those who received treatment is compared to those who experienced the control condition visually (graphically); it is hoped that there is sufficient “common support” or “overlap” among treatment and control groups that the ignorability assumption of causal theory will be tenable. Next, each subject in the treatment group is weighted by the probability of the inverse of the *treatment* PS (1/PS), and each subject in the control group is weighted by the probability of the inverse of the *control* PS (1/(1-PS).

Recall that the treatment PS is the likelihood the subject is from the treatment group given their scores on the covariates. If a treatment subject has a high treatment PS (according to their covariate pattern), such as 0.90, then their selection bias would be considered high and the weight would be 1/0.90, which is fairly close to a weight of 1. If the treatment subject however has a low PS, such as 0.30, then the weight would be 1/0.30, which is quite a bit higher than 1. Similarly (but in an opposite direction), if a control subject has a high *treatment* PS, such as 0.90, then their *control* PS is 1- 0.90, which is 0.10, and their weight will be 1/0.10, which much greater than 1. In other words, subjects who have high propensity scores for their particular condition will get less weight due to their exhibiting more selection bias on the covariates (after all, if the groups aren't biased, then there shouldn't be any difference on the covariates and the probability of being in one of two groups would be close to 0.50).

Finally, a regression analysis is performed on the weighted sample (which adjusts for covariate imbalance across groups) to test the effect of treatment status on the outcome.

The goal of both PS matching and IPW is to achieve groups that are as equal as possible on potential confounders, as would be expected if subjects had been randomized to conditions, so that causal claims about the treatment effects may be strengthened. Nevertheless, the lingering question remains: which method should be believed if they result in different answers?

1.8 GOALS OF THE PRESENT STUDY

The goals of this study are twofold: 1) evaluate the effectiveness of UP, which was developed as a "linkage model" for learning, and 2) compare the UP effectiveness estimates from three types of bias reduction approaches (multiple logistic regression, PS matching, and IPW).

Students with pre-existing interest in caring for underserved populations or certain attributes, demographic or experiential, are more likely to practice later in underserved family medicine as outlined in a 2009 Graham Center monograph. Programs in medical school could nurture this interest and feed these students into residency programs that offer exposure to underserved practice settings to prepare for this career. How specialized tracks for students interested in underserved care in medical schools feed into family medicine residencies that train at underserved sites is unknown. Until now, no study has specifically examined the pipeline from an underserved track to a residency that exposes students to underserved settings such as FQHC, RHC or CAH. In addition, prior research has largely lacked any control for the selection bias of students who choose to enroll in programs oriented toward underserved populations. The present study employed specific bias reduction techniques for estimating the effectiveness of the UP program.

The second aim of this study is to examine three typical different approaches to controlling for selection bias for estimating the treatment effect. Multiple logistic regression, the first approach, directly incorporates covariates in the model for predicting the treatment effect on outcome. However, as discussed before, one of the problems with a multiple regression approach is that the model estimates can be unstable or inestimable due to dimensionality or multicollinearity issues. Propensity score matching, a second technique for estimating an unbiased treatment effect, attempts to achieve covariate balance (eliminating selection bias on the measured covariates) by matching subjects from each group using the PS, a single value that incorporates selection bias across all covariates simultaneously. Finally, a third approach, the inverse weighted probability model, attempts to achieve covariate balance by using the inverse of the PS as a weight, much like complex survey sampling does, for estimating the treatment effect.

Chapter 2. METHODS

2.1 SUBJECTS

Participants included all $N = 204$ UWSOM graduates who chose to match to a family medicine residency between 2010 and 2015. Of these graduates, $N = 158$ (77%) had complete data available for analyses.

2.2 DATA SOURCES

2.2.1 *Underserved Pathway Records*

Records from the UW Department of Family Medicine (which administers the *Underserved Pathway* (UP) program) were used to identify family medicine matched graduates who did not participate in UP (75% of the 158 students with complete data), who enrolled but did not complete UP (6%), and who enrolled and completed UP (25%). Importantly, only students who completed all UP steps were designated as the UP group (all other subjects were defined as non-UP).

2.2.2 *UWSOM Records*

The first UWSOM record used was data from the American Medical College Application Service (AMCAS) application that all students submit to apply to UWSOM. This includes self-reported demographic information (age, gender, race/ethnicity) and dichotomous measures of whether the student came from a home background that was disadvantaged, medically underserved, or included federal or state assistance.

The second data source was a survey that all UWSOM students complete upon matriculation into medical school. Survey items included biographical data, career preferences, language background, and current career plans. Biographical data collected were categorical and included the highest level of education obtained by family members, employment (including health fields) of family members, number of siblings, and state and size of community the student grew up in, including distance and accessibility to nearest metropolitan area. Career preferences (for future practice) were rated on a 5-point Likert-type scale (ranging from 1 = strongly avoid, to 5 = strongly inclined), including 28 medical specialties, 32 items about level of specialization, practice setting, insurance payment models, and types of medical problems seen, and 23 items about preferred geographic setting of practice. Language background was measured with four categorical items about first language spoken, age at which first spoke English, primary household language, and language known best. Finally, current career plans were assessed using three open-ended spaces in which students were asked to write in their top choice of medical specialties, their certainty of those choices using a 5-point rating scale, and their rating of 13 descriptions of practice attributes. All items from the survey are shown in Appendix A.

2.2.3 *AAFP Residency Database and Critical Access Hospital List*

The third source of data for this study came from the American Academy of Family Physicians (AAFP) residency database, which is self-reported information about whether residents have a continuity clinic or rotate through a federally qualified health center (FQHC) and the primary admitting hospital for the residency. These admitting hospitals were compared to a national list of critical access hospitals to identify whether the residencies occurred in critical access hospitals (Flex Monitoring Team, 2016).

2.3 Covariates

2.3.1 *Covariates used in Multiple Logistic Regression and to Create Propensity Score*

Not all items from the questionnaire (see Appendix A) were used in the multiple regression or PS approaches to estimating the treatment effect. Instead, a subset of items were included that were known from prior research to correlate, or were observed in the sample data to correlate, with both (a) UP completion status (the treatment) as well as (b) choice to select a residency in an underserved area (the outcome). In addition, preliminary exploratory factor analyses were conducted to determine which variables were redundant with others, and composites were created among variables that were highly correlated to avoid extreme multicollinearity (for estimating the PS itself). This pared the covariate data down to 26 variables that are mutually exclusive (i.e., none of the composites involved items used in other composites), defined as follows.

- Age – a metrical variable measured in years calculated by subtracting the year of the student's residency match from their year of birth
- Gender – dichotomized as male or female (where 1=female 0 = otherwise)
- Race/ethnicity – dichotomized into minority status (where 1=person of color 0 = otherwise)
- Underprivileged background – a dichotomized composite variable indicating whether a student reported any of the following: rural background, medically underserved background, disadvantaged background, or receipt of federal or state assistance (1 = yes, 0 = otherwise)

- English language learner – dichotomized into whether the student spoke English as a second language at any time during childhood (1= English language learner, 0 = native English speaker)
- Low parental education – a composite variable that indicates the number of primary parents/caregivers with *less than* a college level of education[1] (ranging from 0 to 2)
- High parental education – a composite variable that indicates the number of parents with *more than* a college education (e.g., master's degree, Ph.D., J.D., or medical degree)
- Size of community <10,000 – dichotomized as a small community home living environment (1 = ever lived in a community with a population <10,000 before age 18, 0 = otherwise)
- Distance from nearest metropolitan area >200 miles – dichotomized as growing up in a rural community (1 = ever lived >200 miles from the nearest metropolitan area before age 18, 0 = otherwise)
- Low access to metropolitan area – dichotomized as growing up in a community with low accessibility to urban areas, meaning it was difficult to travel to the metropolitan area due to geographical constraints even if it was close by (1 = ever lived in a community with low accessibility to metropolitan area before age 18, 0 = otherwise)
- Private practice interest – metrical composite variable that is the mean of interest on a 5-point rating scale across four items: solo private practice, shared private practice, private group practice, and mixed specialty practice

[1] A composite variable that indicated the number of parents with a 4-year college education was omitted due to multicollinearity in estimating the propensity scores.

- Employed practice interest – metrical composite variable that is the mean of interest on a 5-point scale across two items: clinic and hospital practice
- Military practice interest – metrical composite variable that is the mean of interest on a 5-point scale across two items: military and VA practice
- Public practice interest – metrical composite variable that is the mean of interest on a 5-point scale across two items: public health and public hospital practice
- Academic practice interest – metrical composite variable that is the mean of interest on a 5-point rating scale across five items: basic science, clinical teacher, research organization, research MD, and administration
- Industry practice interest – metrical composite variable that is the mean of interest on a 5-point rating scale across three items: industry, pharmaceutical companies, and insurance companies
- Interest in other practice – metrical composite variable that is the mean of interest on a 5-point rating scale across two items: church and other settings
- Metropolitan practice interest – metrical variable that is the student's interest in practicing in a metropolitan area on a 5-point rating scale
- City practice interest – metrical variable that is the student's interest in practicing in a city on a 5-point rating scale
- Town practice interest– metrical variable that is the student's interest in practicing in town on a 5-point rating scale
- Low access practice interest – metrical composite variable that is the mean of interest on a 5-point rating scale across four items: community over 100 miles from nearest

metropolitan area, community over 200 miles from the nearest metropolitan area, low accessibility to a metropolitan area, and moderate accessibility to a metropolitan area

- Rural practice – metrical composite variable that is the mean of interest on a 5-point rating scale across two items: a rural environment and a rural deprived environment
- Underserved practice – metrical composite variable that is the mean of interest in on a 5-point rating scale across 3 items: urban deprived environment, foreign deprived environment, and preference for working with underserved population
- Outpatient practice – metrical composite variable that is the mean of interest on a 5-point rating scale across five items: patient education, continuity of care, clinic based care, chronic care, and care in an HMO environment
- Tertiary care/specialty practice – metrical composite variable that is the mean of interest on a 5-point scale across 26 items: tertiary care, and all specialties except internal medicine, pediatrics, and family medicine
- Family medicine first – a dichotomous variable indicating family medicine practice would be in their top choice of specialty (1= listed as a top choice, 0 = not listed as top choice)

2.3.2 *Treatment Variable*

The independent variable is completion of the Underserved Pathway (UP) project, dichotomized to completion or no completion. Graduates who enrolled but did not complete where coded as no completion.

2.3.3 *Outcome Variable*

The outcome variable, Underserved Residency, is a binary variable indicating the residency that the student chose offered exposure to a FQHC (either as a continuity clinic or through a rotation), or a CAH. This variable is dummy coded with 1 = any exposure to an underserved area/population, and 0 = no exposure.

2.4 ANALYTIC PLAN

2.4.1 *Sample Description*

Descriptive statistics, including means, standard deviations, and zero-order correlations for the outcome and covariates were computed for the combined sample; means and standard deviations were also computed and compared for UP (treatment) and non-UP (control) groups separately. Groups were also compared on each variable to determine baseline differences (selection bias) among the groups: for metrical variables, 2-group *t*-tests were used with an alpha = 0.05, 2-tailed; for binary outcomes, 2-group χ^2 tests were used with an alpha = 0.05.

2.4.2 *Statistical Approaches*

Four approaches were used to calculate the average treatment effect of completing UP on students' choice to serve in an underserved residency, all of which were conducted in STATA 14. Across all approaches, the outcome is binary and hence logistic regression was employed. The first approach was a simple logistic regression (naïve) that only included UP as a covariate. This approach is considered naïve since it does not account for any covariate differences (selection bias). The second approach used a multiple logistic regression (MLR) in which UP status as well as all covariates were used in the analysis to predict the outcome. The third and fourth approaches included propensity score (PS) approaches. Specifically, the third approach used PS matching (PSM) using 1:n nearest neighbor matching with replacement, a caliper of 0.3 logits minimum distance for a potential match. The average treatment effect (ATE) on the population was calculated using robust standard errors. As described by both Stuart and Luo in 2010, this method was selected as a frequently used and defensible manner of matching to reduce selection bias (Luo, Gardiner, & Bradley, 2010; Stuart, 2010). (As a check on matching an overlap plot of the density of PS values was inspected to evaluate common support between the two groups; in addition, covariate balance post-matching was also assessed using *t*-tests and χ^2 tests.) Finally, the final approach to estimating the treatment effect employed inverse probability weighting (IPW) in which each subject is weighted for the level of bias their covariate pattern exhibited. A minimum propensity score of 0.00001 was set as tolerance for common support. There was no trimming of the sample and again the average treatment effect on the population was calculated using robust standard errors.

2.4.3 *Covariate Balance Checks*

A number of post-estimation analyses were conducted to explore covariate balance in the naïve, PSM, and IPW approaches (not possible for multiple regression as multiple regression uses the semi-partial correlation between the treatment and outcome after controlling for relationships with other variables, not changes in subjects included, or weighted, in the analyses). First, a list of matched subjects was generated for the PSM approached. Then, as suggested by Austin and Stuart, standardized differences were performed for each metrical variable while variance ratios were calculated for each categorical variable (Austin & Stuart, 2015). A within-subjects ANOVA was performed on the 18 standardized differences and then on the 8 variance ratios to determine if there were any significant differences in covariate balance across these three approaches.

2.4.4 *Sensitivity Analysis*

Finally, for the PSM approach only, a sensitivity analysis was employed to assess whether there were plausible unidentified confounders that were still inducing selection bias; this test uses McNemar's matched samples χ^2 test based on the calculation of a gamma statistic (Rosenbaum, 2010).

Chapter 3. RESULTS

3.1 SAMPLE DESCRIPTION

Table 1 reports descriptive statistics for the entire sample and bivariate correlations (Pearson's *r*) between the outcome, treatment, and 26 covariates. Of all family medicine graduates, 58% matched to an underserved residency, and almost a quarter (25%) completed the Underserved Pathway (UP) program. The average age of the graduates was 30.45 years, 70% were female, 22% were people of color, 10% reported being an English language learner as a child, and 12% reported having grown up in an underserved background.

Three covariates had significant but small correlations with the outcome: interest in public practice, other practice, and metro practice (rs < 0.20). Ten covariates had significant correlations with the treatment: interest in employed practice, military practice, academic practice, city practice, town practice, low access area practice, and outpatient practice, as well as listing family medicine as a top residency choice, and interest in tertiary care.

In Table 2, the outcome variable and covariates were compared using *t*-tests for metrical variables and χ^2tests for binary (categorical) variables. Note that an experimental design using randomization of subjects to treatment and control conditions would be expected to have 5% of covariates to exhibit significant group differences (assuming an alpha level of 0.05). In the present sample, 10 out of 26 covariates were unbalanced (38% of the variables) which does indicate selection bias exists. The direction of the differences showed that UP graduates had higher ratings for town, low access, rural, and outpatient practice, and were also more likely to list family medicine as a top choice whereas non-UP graduates had higher ratings for employed, military, academic, city, and tertiary care for their future practice.

3.2 APPROACH OUTCOMES

The naïve model results (Approach 1) are given in Table 3, and show that the simple logistic regression results indicate that UP was not a significant predictor of selecting an underserved residency (regression coefficient *b*-value $p > 0.05$). The estimated treatment effect using predicted probabilities (albeit non-significant), which are also equal to the difference between the observed frequencies given Table 2, is equal to a 12% difference between the groups.

The multiple logistic regression results (Approach 2) are given in Table 4. Importantly, after controlling for the 26 covariates, the UP treatment effect on the outcome (underserved residency choice) was significant, and estimated to be 17% (91% of UP graduates are predicted to select an underserved residency compared to 74% of graduates who did not complete UP. Predictors that were uniquely predictive of underserved residency choice (above and beyond UP participation) included interest in academic practice, other types of practice, and town practice.

The results from the PS 1:n nearest neighbor matching with replacement (PSM) and PS inverse probability weighting (IPW) approaches (Approach 3 and 4, respectively) are given in Table 5 and Figure 2, along with the prior two approaches. As can be seen, the PSM approach estimated the treatment effect to be significantly greater than zero, as a 23% difference between the groups favoring UP graduates in choosing an underserved residency. However, the IPW approach estimated the group difference to be slightly lower and non-significant at a 21% difference. As is readily seen in the table and graph, the standard error of treatment for IPW is double that of PSM, indicating that the approach may be suffering from multicollinearity or dimensionality. Had the standard error been smaller, the result for IPW would be significant.

3.3 COVARIATE BALANCE CHECKS FOR PSM AND IPW

Given that the purpose of PS approaches is to achieve reduced or eliminated selection bias, then the groups should be balanced on covariates (i.e., there should not be differences between groups on the covariates after matching or weighting). Table 6 shows the specific subject matches for PSM: of the 39 treatment subjects, 35 were matched to controls, and two of those controls (IDs 54 and 117) were responsible for 37% of the control matches.

Figure 3 displays the density of PS values after matching was complete for each of the two groups; the hope is that there is overlap between the two densities, or "common support" for making the ignorability assumption tenable. Although there are clearly differences in the two distributions, there does appear to be common overlap to warrant the idea that selection bias has been reduced in the matched sample.

To go a step further, covariate balance across three of the approaches (naive, PSM, and IPW) was evaluated (see Table 7). Standardized differences were computed for metrical variables (absolute value of mean difference between groups in standard deviations) and for categorical variables, frequency variance ratios between groups. Standardized differences of zero and variance ratios of 1 indicate perfect covariate balance (i.e., no difference between UP and non-UP groups). For PSM and IPW, the difference between each balance statistic and the naïve approach were computed to determine the percentage of bias reduction each contributed: an improvement of 3-5% reduction in bias is one suggested standard by Caliendo (2008) for evaluating improvement in covariate balance from the naïve approach (Caliendo & Kopeinig, 2008). For PSM match, 15 covariates had a reduction in bias, while for IPW there was a reduction in bias for 14 covariates, not all of which were overlapping reductions.

These balance indices were further evaluated using a 1-way within-subjects analysis of variance (ANOVA) to test whether the three approaches (naïve, PSM, and IPW) differed on standardized differences and variance ratios (see Table 7 for mean values). There were no significant main effects for standardized differences, *Greenhouse-Geisser Adjusted F*(1.12, 19.00) = 1.84, $p = 0.192$. Follow up paired *t*-tests among the three approaches (using Dunn-Sidak adjusted *p*-values to control Type I error to 0.05) revealed a significant mean balance difference between PSM and IPW across all measures 4.5%, adjusted $p < 0.05$. The comparison of differences in approaches on variance ratios, however, showed no main effect or pairwise differences (*p*s > 0.05).

3.4 Sensitivity Analysis

As a final analysis, a sensitivity analysis was performed using McNemar's test for the PSM approach to determine if there was evidence for any unmeasured (omitted) confounders (see Table 8). In the matched sample of 158 pairs, 43% of pairs were observed to both have selected the outcome (chose an underserved residency), and 12% of were observed to both not choose an underserved residency. Importantly, 34% of pairs were discordant on the outcome, with more treated subjects choosing the outcome compared to controls. Adjusting the *p* value to just below 0.05 yielded a gamma statistic of 1.94. This means that, for the PSM approach's matched sample, an unobserved covariate would need to produce a nearly twofold increase (1.94 to be specific) in the treatment group's odds of selecting an underserved residency. As previously described (Rosenbaum, 2005) the usual range for being concerned that a variable may have been omitted is when gamma is close to 1. For context, the gamma statistic for an unobserved covariate other than smoking to cause lung cause is around 5. From these results, the PSM's treatment estimate would appear to be fairly robust to an omitted confounder.

Chapter 4. DISCUSSION

4.1 Comparison of Different Approaches

The average treatment effect of UP completion on a graduate choosing an underserved residency depended on estimation approach. The naïve approach that would be confounded by selection bias estimated the smallest treatment impact (12%, not significant). The multiple logistic regression (MLR) and propensity score matching (PSM) approaches both yielded slightly higher (and significant) treatment effects: 17.2% and 23% higher probability of UP graduates selecting an underserved residency compared to non-UP graduates, respectively. Surprisingly, the inverse probability weighting (IPW) approach yielded a treatment effect estimate highly similar to the MLR and PSM approaches (21%), but the large standard error precluded the estimate from being significant. Covariate balance was not substantively different across the approaches; however, PSM was slightly better than IPW by approximately a 5% reduction in mean bias across all metrical covariates. Further, the PSM approach found that the estimate was robust to potential omitted variable bias

Given the somewhat divergent treatment effect findings, the question naturally centers on which approach is yielding the most accurate estimate. First, the naïve treatment effect estimate can be ruled out considering that 10 of the 26 pre-treatment covariates (38%) exhibited significant group differences when only 5% would be expected if the groups had been randomized (and therefore ignorability does not hold for the naïve approach). In simple terms, covariates should not be ignored.

Second, given the consistency of the point estimates (mean treatment effect), it seems likely that the true treatment effect is between 17% and 23% favoring UP over non-UP graduates. Further, both MLR and PSM approaches arrive at the same conclusion: the treatment effect is significantly different from zero. The fact that the IPW does not find the effect to be significant is due to the unusually large standard error that can easily occur in logistic regression when even moderate correlations among many covariates exist. It seems that, since other studies typically find consistent estimates between MLR and PS approaches (Shah, Laupacis, Hux, & Austin, 2005), it is likely the case that the IPW approach is inconsistent due to the current study's particular correlational characteristics across 26 covariates relative to a small sample size of $N = 158$ subjects.

Another consideration for any researcher is that each approach has benefits and risks. MLR is a straightforward approach that has been used for decades and is well understood by a wide audience. Although it does not discard any cases (i.e., it employs the entire sample), the sample-to-variable size ratio in this analysis is less than 6:1, much lower than the typically recommended 15:1 ratio. Hence, statistical power for this analysis is quite a bit lower than usually desired. While PSM may also achieve bias reduction through matching (and for the present study, slightly better mean covariate balance than IPW), it uses only a subset of the entire sample and the 1:n nearest neighbor matching with replacement algorithm may result in downwardly biased variance since a single control subject can act as a match for multiple treatment subjects. All this said, IPW has the benefit of including the entire sample (although weighted) in the analysis and reduces selection bias by using a composite (PS value) that captures the selection bias across all the covariates simultaneously). However, like MLR, creating the PS itself is susceptible to multicollinearity and dimensionality, and for the present study, this issue may have led to unusually high standard errors.

PS analysis approaches are still a relatively young field in methodology. There are many questions that are still unanswered regarding which method is best suited to answer specific research questions and how to clearly judge the success of covariate balance after using a individual approach. In the present study, the covariate balance achieved was not, on average, significantly different from the unaltered sample. It is possible that a better specified model for the PSM or IPW would have resulted in improved balance. It is also possible that the covariates included in these analyses were not as reliable as one would hope; there is much work that has shown that reliability of measures has a direct effect on treatment effect estimates. For the present study, the parsimonious way to proceed would be to stick with the simplest approach – the MLR – which reduces selection bias directly using observed covariates (as PS approaches do in a two-step manner) so long as dimensionality and multicollinearity are not an issue.

4.2 UNDERSERVED PATHWAY COMPLETION AND UNDERSERVED RESIDENCY MATCH

Completion of the UP resulted in 17% percentage point increase in graduates selecting to match to an underserved residency after controlling for a large set of potential confounders. Although treatment effect estimates from a quasi-experimental study typically only allows us to make associative links between treatment and the outcome, use of MLR and PS approaches allows for stronger, tentatively causal claims, about the link so long as we believe that there are no omitted variables in the analysis (unmeasured confounders). Although this study did not employ random assignment, there are several plausible reasons for why the UP effect on graduates' choice is likely to be causal. First, UP may support interest in caring for underserved populations throughout medical school, thus that when students prepare to match to a residency program, the have maintained this initial interest. UP may also protect against the well-documented decline in altruism that occurs in medical school; supporting altruism throughout school may help students select careers caring for underserved patients upon graduation. Second, students who have devoted significant extracurricular time to learning more about caring for underserved populations may be more attracted to residency programs that have a mission to prepare physicians for these careers.

This present study also suggests that the "linkage" approach to medical education outcomes could be a way to potentially identify future impacts of an intervention without the time delay resulting from waiting for students to complete the entire training process. For the desired results of increasing the number of physicians practicing in underserved areas, this study indicates that UP may be an important component of the pipeline that starts with interested students and likely delivers them to undeserved practice settings nearly a decade later. Studies of more mature programs, such as Rural Physician Associate Program, have found similar associations with the medical school experience and eventual practice outcome. This study suggests that claims of association or causality on the long-term outcome for a given medical school program could be generated earlier if later links in the chain are already pre-existing.

Notably, one of the strengths of this study also includes the sheer number of initial interest variables with which to control for in the treatment effect model. This makes an unobserved confounder less likely, as indicated by the gamma close to a value of two for the PSM approach. This said, like all studies, there are a number of limitations to our results, including that the study is still quasi-experimental: the results are not necessarily causal although the basis for causal inference is high considering the consistency in estimates and high number of covariates controlled for. Additional limitations include that the study took place within a single institution with a stated mission of producing primary care physicians. As such, these findings likely do not generalize to other schools and programs that do not have this kind of focus. Future research could include the addition of more variables known to impact residency choice that occur during medical school, such as United States Medical Licensing Exam scores. The sample size of the present study is also a strong limitation: future research might aggregate results across a number of universities and programs.

In summary, the UWSOM's UP program had a significant effect on graduates' choice to be matched to underserved residencies, and therefore it has a promising future that will require evaluation of longer-term outcomes. Importantly, the point estimate of the effect was consistent across three bias-reduction approaches, even if one had unusually large standard errors; researchers should take care to deal with selection bias whenever a quasi-experimental design is employed (i.e., the naïve approach is not recommended under any circumstance).Further, propensity score (PS) analysis methods are likely to be a useful tool for future medical education research, particularly when multiple logistic regression is not feasible due to multicollinearity or dimensionality. Although these methods cannot replace the use of a true experimental design, they do have a strong potential for addressing selection bias.

TABLES AND FIGURES

Table 1. Sample Description and Bivariate Correlations

Table 2. Descriptive Statistics and Naïve Comparison of UP and non-UP Students

Table 3. Naïve Logistic Regression Results Estimating UP Effect on Underserved Residency Choice (Selection Bias Not Accounted For)

Table 4. Multiple Logistic Regression Results Estimating UP Effect on Underserved Residency Choice (First Bias Reduction Approach)

Table 5. Comparison of Approaches for Predicting Graduates' Choice to Serve in an Underserved Residency

Table 6. Propensity Score Matched Pairs using 1:n Nearest Neighbor with Replacement

Table 7. Comparison of Covariate Balance between Naïve (Ignoring Covariates) and each PS Estimation Approach for the UP Effect

Table 8. Testing Accuracy of PS Approaches for Balancing Groups on Covariates

Table 9. Sensitivity Analysis for PS Matching Approach to UP Effect Estimation

Figure 1. Linkage Model of Medical Education Outcomes

Figure 2. Graphical Comparison of Predicted Probabilities by Statistical Approach

Figure 3. Propensity Score Density (Common Support) for UP (Right) and non-UP (Left)

Table 1.

Sample Descriptive Statistics and Correlations among Variables

Measure	*M*	*(SD)*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Outcomes																													
1. Underserved Residency	0.58	0.04	--																										
Covariates																													
2. UP Completed	0.25	0.03	.11	--																									
3. Female	0.70	0.04	.10	.09	--																								
4. Age	30.45	0.36	-.10	-.11	**-.16**	--																							
5. Person of color	0.22	0.03	.14	-.09	.01	-.04	--																						
6. English language learner	0.10	0.02	.03	-.10	.04	.07	**.59**	--																					
7. Low parent education	0.70	0.06	.03	-.01	.06	.05	**.31**	**.36**	--																				
8. High parent education	0.75	0.06	-.08	.09	-.05	-.06	**-.17**	**-.22**	**-.68**	--																			
9. Underserved background	0.12	0.03	.08	-.08	-.01	-.04	.14	**.20**	**.26**	**-.21**	--																		
10. Community <10,000	0.41	0.04	.06	.04	-.04	.05	-.15	-.06	.02	-.09	.05	--																	
11. Community >200 miles	0.23	0.03	.01	-.03	-.04	.05	-.10	-.03	.13	-.14	.03	**.35**	--																
12.Low accessibility community	0.15	0.03	.01	.00	-.03	.12	.04	.15	.09	-.07	.12	**.37**	**.49**	--															
13. Private practice	3.26	0.06	-.07	.07	**-.16**	-.14	**-.19**	**-.17**	**-.16**	.06	-.07	.01	.05	-.10	--														
14. Employed practice	3.50	0.06	.13	**-.20**	-.04	-.09	**.18**	.10	.03	.05	.10	.05	.02	.11	**-.21**	--													
15. Military practice	2.59	0.06	-.15	**-.22**	-.15	-.03	-.02	-.03	.04	-.02	.09	.10	.05	.04	.06	.08	--												
16. Public practice	3.57	0.05	**.16**	-.11	.14	.07	.09	.14	.09	-.02	**.16**	.10	.04	.01	**-.16**	**.37**	.06	--											
17. Academic practice	2.57	0.06	.07	**-.22**	-.11	.06	.08	.03	.04	-.07	.12	.14	-.08	-.04	.09	**.23**	**.28**	.06	--										
18. Industry practice	1.59	0.05	-.07	-.11	**-.28**	.06	.06	.00	.07	-.12	.14	.07	.04	-.04	**.16**	.07	**.30**	-.11	**.48**	--									
19. Other practice	2.28	0.10	**-.19**	-.02	**-.16**	-.13	.08	-.06	.04	-.10	.03	.11	.10	.00	.12	-.09	**.30**	-.06	**.17**	**.18**	--								
20. Metro practice	2.93	0.10	**.19**	-.09	.03	.03	**.34**	**.34**	.03	.06	.04	**-.37**	**-.29**	-.05	-.15	.10	.03	**.18**	.10	.05	-.03	--							
21. City practice	3.62	0.06	.13	**-.18**	-.09	.00	.03	-.01	-.04	.03	-.12	.05	.06	-.07	.10	.07	-.07	-.02	.07	**.19**	.02	.15	--						
22. Town practice	3.21	0.09	.04	**.19**	.02	.08	-.14	**-.18**	-.01	.01	.06	**.45**	.13	.07	.12	-.12	.10	.01	.07	.12	.03	**-.46**	.01	--					
23. Low access practice	2.92	0.07	-.05	**.22**	-.07	.09	**-.21**	**-.18**	**-.16**	**.17**	.10	**.26**	**.27**	.06	.12	-.15	.03	-.05	-.02	.06	.02	**-.39**	-.05	**.58**	--				
24. Rural practice	3.83	0.07	-.12	**.26**	.02	.04	**-.17**	**-.17**	-.04	.09	-.04	**.33**	**.20**	.08	**.16**	-.03	-.06	-.03	.02	-.02	.13	**-.56**	.00	**.67**	**.55**	--			
25. Underserved practice	3.81	0.07	.12	-.05	.11	.00	**.27**	**.19**	-.01	.08	.05	**-.21**	**-.19**	-.05	**-.19**	-.03	.06	**.22**	.03	-.03	.06	**.60**	.09	**-.18**	-.15	**-.24**	--		
26. Outpatient practice	4.11	0.05	.10	**.27**	**.24**	.00	.15	.07	-.01	.03	.09	.07	-.05	.05	-.01	.10	-.14	**.29**	.04	-.15	.07	-.01	.00	**.22**	.10	**.29**	.05	--	
27. Family medicine first	0.53	0.04	-.07	**.22**	.03	.00	.07	-.06	.04	-.08	-.04	.06	**-.18**	.01	.02	-.04	-.03	.05	.04	-.08	**.19**	-.13	.03	**.21**	.02	**.23**	-.01	**.39**	--
28. Tertiary/specialty practice	2.53	0.04	-.08	**-.24**	**-.31**	-.08	-.07	-.05	-.01	-.04	.00	.16	.07	.00	**.21**	.09	**.41**	**-.17**	**.37**	**.48**	.13	.07	.15	.06	-.04	-.13	.00	**-.33**	**-.22**

Note. $N = 158$ family medicine graduates. All categorical variables dummy coded (1 = yes, 0 = otherwise); most metrical variables were measured on a rating scale. Boldfaced correlations indicate significant relationships, $p < 0.05$.

Table 2.
Descriptive Statistics and Naïve Comparison of UP and non-UP Students

	Family Medicine Graduates					
	UP (treatment) n = 39		Non-UP (control) n = 119			
Variable	*M*	*(SD)*	*M*	*(SD)*	*Statistic*	*p*
Outcome						
Chose Underserved Residency	0.67	(0.48)	0.55	(0.50)	1.74	0.187
Covariates						
Female	0.77	(0.43)	0.67	(0.47)	1.31	0.253
Age	29.62	(3.67)	30.72	(4.78)	1.32	0.188
Person of color	0.15	(0.37)	0.24	(0.43)	1.15	0.283
English language learner	0.05	(0.22)	0.12	(0.32)	1.42	0.233
Low parental education	0.69	(0.80)	0.71	(0.82)	0.09	0.928
High parental education	0.87	(0.61)	0.71	(0.58)	-1.12	0.265
Underserved background	0.08	(0.27)	0.13	(0.34)	0.92	0.338
Community <10,000 pop.	0.44	(0.50)	0.39	(0.49)	0.20	0.651
Distance >200 miles	0.21	(0.07)	0.24	(0.43)	0.15	0.697
Low access community	0.15	(0.37)	0.15	(0.36)	0.00	0.969
Private practice	3.35	(0.75)	3.23	(0.78)	-0.86	0.390
Employed practice	**3.22**	**(0.82)**	**3.60**	**(0.79)**	**2.58**	**0.011**
Military practice	**2.28**	**(0.84)**	**2.70**	**(0.78)**	**2.84**	**0.005**
Public practice	3.44	(0.68)	3.61	(0.69)	1.36	0.177
Academic practice	**2.29**	**(0.78)**	**2.66**	**(0.68)**	**2.79**	**0.006**
Industry practice	1.47	(0.56)	1.63	(0.62)	1.43	0.156
Other practice	2.23	(1.33)	2.29	(1.27)	0.25	0.803
Metro practice	2.73	(1.42)	2.99	(1.17)	1.17	0.243
City practice	**3.38**	**(0.90)**	**3.70**	**(0.75)**	**2.22**	**0.028**
Town practice	**3.59**	**(1.00)**	**3.08**	**(1.15)**	**-2.47**	**0.014**
Low access practice	**3.25**	**(0.84)**	**2.81**	**(0.89)**	**-2.76**	**0.006**
Rural practice	**4.23**	**(0.61)**	**3.70**	**(0.92)**	**-3.35**	**0.001**
Underserved practice	3.74	(0.17)	3.84	(0.08)	0.57	0.569
Outpatient practice	**4.39**	**(0.46)**	**4.02**	**(0.62)**	**-3.44**	**0.001**
Family medicine first	**0.72**	**(0.46)**	**0.46**	**(0.50)**	**7.71**	**0.006**
Tertiary practice	**2.30**	**(0.64)**	**2.60**	**(0.50)**	**3.07**	**0.003**

Note. N = 158 family medicine graduates. All categorical variables dummy coded (1 = yes, 0 = otherwise); most metrical variables were measured on a rating scale. Statistical tests are as follows: metrical variable mean differences are based on 2-group t-test with df = 156; categorical variable frequency differences are based on 2-group χ^2 tests with df = 1. Boldfaced indicates a significant difference between groups, $p < 0.05$.

Table 3.
Naïve Logistic Regression Results Estimating UP Effect on Underserved Residency Choice (Selection Bias Not Accounted For)

	$\chi^2(1)$	*p*	*Pseudo R^2*	*b*	*(SE)*	*Wald*(1)	*p*
Model Fit	1.78	0.183	0.008				
Coefficients							
Intercept				0.19	(0.18)	1.01	0.314
Underserved Pathway				0.51	(0.39)	1.73	0.189

Note. *N* = 158 family medicine graduates. Outcome (choice of underserved residency) and treatment (underserved pathway, UP) dummy coded (1 = yes, 0 = no).

Table 4.
Multiple Logistic Regression Results Estimating UP Effect on Underserved Residency Choice (First Bias Reduction Approach)

	$\chi^2(27)$	*p*	*Pseudo* R^2	*b*	*(SE)*	*Wald*(1)	*p*
Model Fit	50.48	0.004	0.234				
Coefficients							
Intercept				1.03	(3.12)	0.11	0.741
Underserved Pathway				**1.28**	**(0.58)**	**4.79**	**0.029**
Gender				-0.30	(0.49)	0.35	0.538
Age				-0.09	(0.05)	2.98	0.084
Person of color				1.30	(0.75)	3.00	0.083
English language learner				-1.57	(0.98)	2.58	0.108
Low parental education				-0.32	(0.36)	0.79	0.373
High parental education				-0.67	(0.39)	3.00	0.083
Underserved background				0.38	(0.72)	0.28	0.595
Community <10,000				0.56	(0.56)	1.01	0.315
Distance >200 miles				0.37	(0.64)	0.35	0.556
Low access community				-0.04	(0.72)	0.00	0.960
Private practice				0.05	(0.29)	0.04	0.852
Employed practice				0.28	(0.31)	0.82	0.367
Military practice				-0.36	(0.30)	1.41	0.236
Public practice				0.36	(0.38)	0.92	0.337
Academic practice				**0.78**	**(0.35)**	**4.85**	**0.028**
Industry practice				0.03	(0.45)	0.00	0.125
Other practice				**-0.36**	**(0.18)**	**4.01**	**0.045**
Metro practice				0.35	(0.27)	1.73	0.189
City practice				0.56	(0.29)	3.62	0.057
Town practice				**0.64**	**(0.29)**	**4.86**	**0.027**
Low access practice				-0.11	(0.31)	0.12	0.727
Rural practice				-0.72	(0.38)	3.56	0.059
Underserved practice				0.18	(0.30)	0.37	0.542
Outpatient practice				0.01	(0.40)	0.00	0.973
Family medicine first				-0.92	(0.50)	3.41	0.065
Tertiary practice				-0.66	(0.51)	1.64	0.201

Note. N = 158 family medicine graduates. All categorical variables dummy coded (1 = yes, 0 = otherwise); most metrical variables were measured on a rating scale. Boldfaced indicates the statistic is significantly different from zero, $p < 0.05$.

Table 5.
Comparison of Approaches for Predicting Graduates' Choice to Serve in an Underserved Residency

	Predicted Probability			
Approach to Effect Estimation	*ATE*	*(SE)*	*p*	
Logistic Regression (Naïve)	12.04%	(2.59%)	0.189	
Multiple Logistic Regression (MLR)	17.21%	(3.00%)	0.029	*
Propensity Score Matching (PSM)	23.40%	(7.65%)	0.002	*
Inverse Probability Weighting (IPW)	20.90%	(15.88%)	0.188	

Note. N = 158 family medicine graduates.

Table 6.
Propensity Score Matched Pairs using 1:n Nearest Neighbor with Replacement

Subject ID	Matched Pairs	% of Pairs	Matched Subject IDs
1	2	1%	**39, 155**
5	2	1%	103, 114
49	6	4%	8, 31, 22, 43, 101, 138
54	40	25%	2, 6, 7, 9, 10, 11, 13, 14,17, 18, 19, 22, 24, 26, 27, 30, 37, 40, 42, 44, 47, 52, 53, 55, 58, 61, 68, 71, 72, 74, 75, 81, 86, 102, 107, 108, 116, 140, 147
57	1	1%	**117**
59	1	1%	**125**
60	6	4%	63, 88, 126, 127, 135, 136
63	3	2%	**60, 78, 137**
77	9	6%	**41, 66, 83, 87, 100, 128, 134, 148, 152**
80	1	1%	143
83	1	1%	77
84	1	1%	**153**
91	1	1%	105
92	1	1%	**131**
95	2	1%	**23, 133**
98	3	2%	**70, 119, 144**
101	1	1%	**49**
105	9	6%	**67**, **73**, **89**, **96**, **118**, **141**, **150, 157**
108	1	1%	**54**
110	4	3%	95, 115, 121, 156
113	1	1%	**149**
114	1	1%	**5**
115	1	1%	**110**
117	19	12%	3, 12, 28, 32, 38, 46, 48, 57, 63, 64, 76, 85, 109, 111, 124, 130, 132, 145, 154
119	1	1%	82
120	1	1%	**146**
125	8	5%	34, 36, 59, 79, 93, 112, 1**23**, 151
131	2	1%	21, 92
137	7	4%	15, 29, 35, 51, 56, 69, 99
143	1	1%	**80**
144	4	3%	4, 45, 97, 98
146	7	4%	16, 20, 50, 90, 94, 106, 120
149	2	1%	113, 158
153	6	4%	25, 84, 104, 129, 139, 142
155	2	1%	1, 122

Note: Bolded indicates treatment subject. Total of 158 matched pairs.

Table 7.
Comparison of Covariate Balance between Naïve (Ignoring Covariates) and each PS Estimation Approach for the UP Effect

Covariate	Standardized Differences (Metrical Variables)			Variance Ratios (Categorical Variables)		
	Naive	*PSM*	*IPW*	*Naive*	*PSM*	*IPW*
Female				0.82	1.05 [a]	1.09 [a]
Age	0.26	0.25	0.20 [a]			
Person of color				0.74	0.88 [a]	0.84 [a]
English language learner				0.48	0.36	0.41
Low parental education	0.02	0.47	0.54			
High parental education	0.20	0.26	0.40			
Underserved background				0.62	1.08 [a]	1.04 [a]
Community <10,000				1.05	1.05	1.03 [a]
Distance >200				0.92	1.18	1.37
Low access community				1.03	0.50	0.59
Private practice	0.16	0.02 [a]	0.06 [a]			
Employed practice	0.47	0.00 [a]	0.03 [a]			
Military practice	0.51	0.23 [a]	0.23 [a]			
Public practice	0.25	0.26	0.29			
Academic practice	0.50	0.37 [a]	0.43 [a]			
Industry practice	0.27	0.06 [a]	0.06 [a]			
Other practice	0.05	0.36	0.35			
Metro practice	0.21	0.12 [a]	0.24			
City practice	0.39	0.37 [a]	0.32 [a]			
Town practice	0.47	0.49	0.60			
Low access practice	0.52	0.38 [a]	0.38 [a]			
Rural practice	0.68	0.52 [a]	0.61 [a]			
Underserved practice	0.10	0.06 [a]	0.21			
Outpatient practice	0.68	0.19 [a]	0.13 [a]			
Family medicine first				0.83	0.98 [a]	1.00
Tertiary practice	0.53	0.19 [a]	0.34 [a]			
Means	*0.35*	*0.26*	*0.30*	*0.81*	*0.89*	*0.92*

Note. $N = 158$ family medicine graduates. Standardized mean differences for metrical variables are the mean difference between groups in standard deviations: closer to zero indicates better balance (less selection bias). Variance ratios are the ratio of treatment: control proportions in the focal category: closer to 1 indicates better balance (less selection bias). IPW model's overall balance fit was $\chi^2(27) = 6.14$, $p > .999$. [a] Indicates that the direction of the bias reduction achieved better balance compared to the naive approach.

Table 8.
Testing Accuracy of PS Approaches for Balancing Groups on Covariates

	Naïve		PSM		IPW		Naïve vs. PSM			Naïve vs. IPW			PSM vs. IWP		
Mean Balance Statistic	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*Diff*	*(SE)*	*p*	*Diff*	*(SE)*	*p*	*Diff*	*(SE)*	*p*
Standardized Difference	0.35	(0.21)	0.26	(0.16)	0.30	(0.18)	0.09	(0.06)	0.307	0.05	(0.06)	0.825	**-0.05**	**(0.02)**	**0.045**
Variance Ratio	0.81	(0.20)	0.89	(0.29)	0.92	(0.30)	-0.08	(0.11)	0.874	-0.11	(0.10)	0.681	-0.04	(0.03)	0.563

Note. N = 158 family medicine graduates. Standardized mean differences for metrical variables are the mean difference between groups in standard deviations: closer to zero indicates better balance (less selection bias). Variance ratios are the ratio of treatment: control proportions in the focal category: closer to 1 indicates better balance (less selection bias). Mean standardized difference based on 18 covariates, variance ratios based on 8 covariates. Boldfaced indicates the statistic was significantly different from zero using paired samples t-tests with Dunn-Sidak correction, $p < 0.05$. Lack of a difference between Naïve approaches and the two PS approaches likely due to high random variability in the standardized mean differences.

Table 9.
Sensitivity Analysis for PS Matching Approach to UP Effect Estimation

Matched Sample	Treated, outcome=yes		Treated, outcome=no		Total	Gamma	*p*
Control, outcome=yes	68	43%	17	11%			
Control, outcome=no	54	34%	19	12%			
Total	122		36		158	1.94	0.05

Note. N = 158 family medicine graduates. Gamma statistic is adjusted until the p-value is at the threshold for the predefined alpha level of significance, in this case <0.05. Unmeasured covariates would need to exert an impact in excess of the gamma to in order to change the treatment effect estimation.

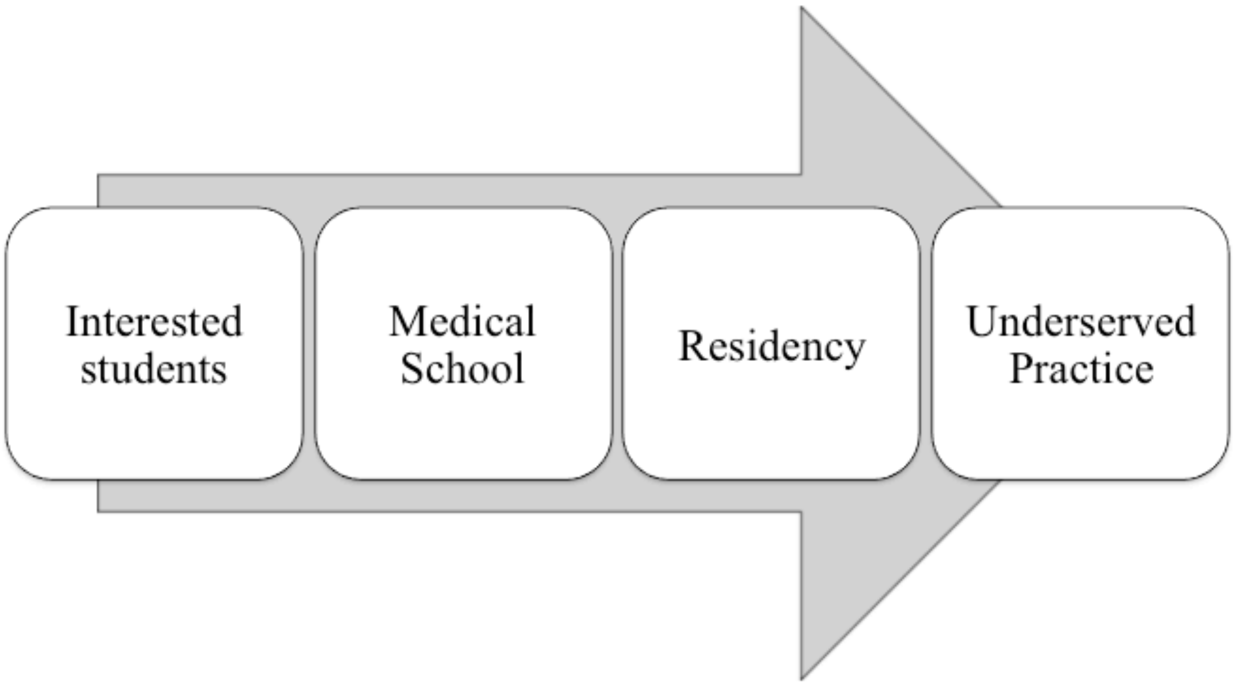

Figure 1. Linkage Model of Medical Education Outcomes

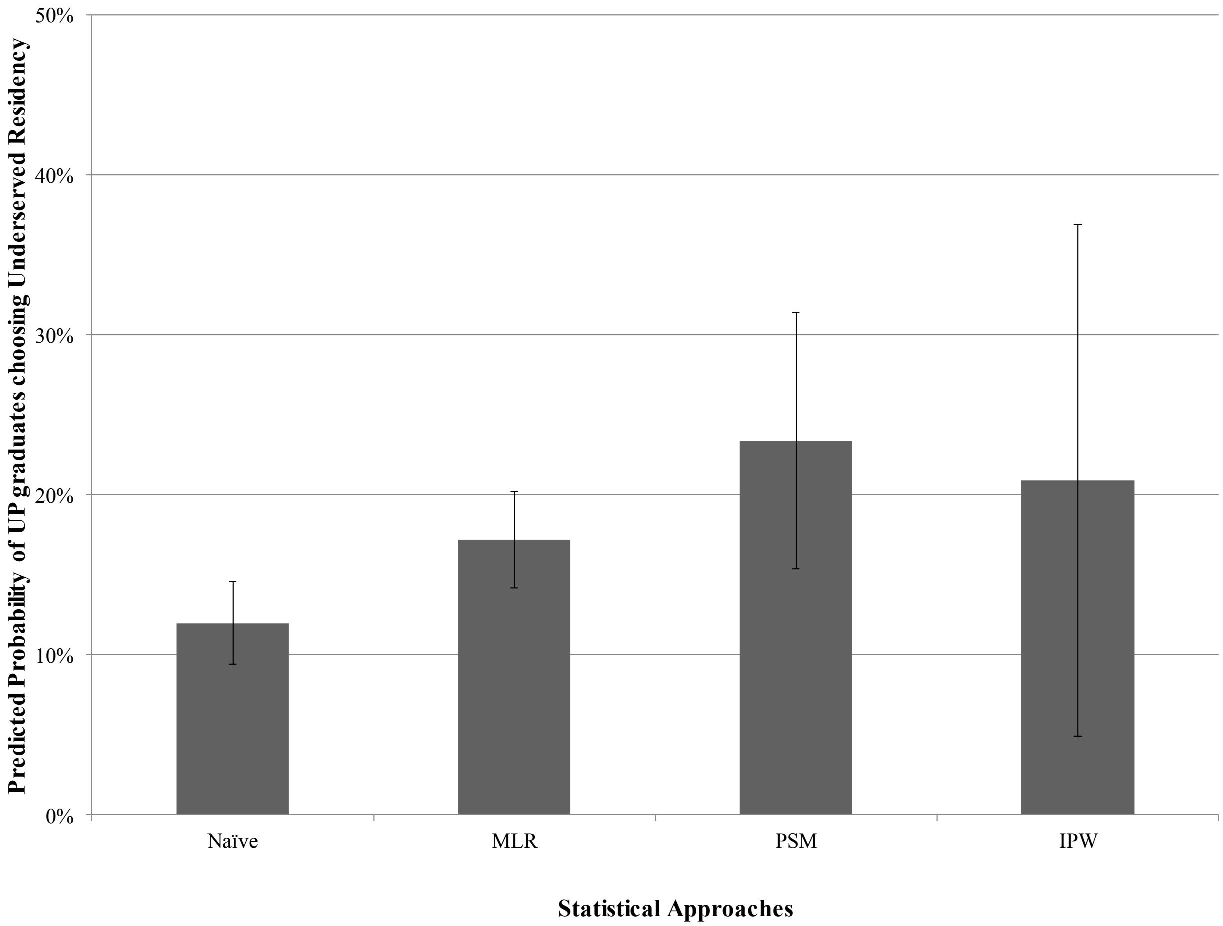

Figure 2. Graphical Comparison of Predicted Probabilities by Statistical Approach

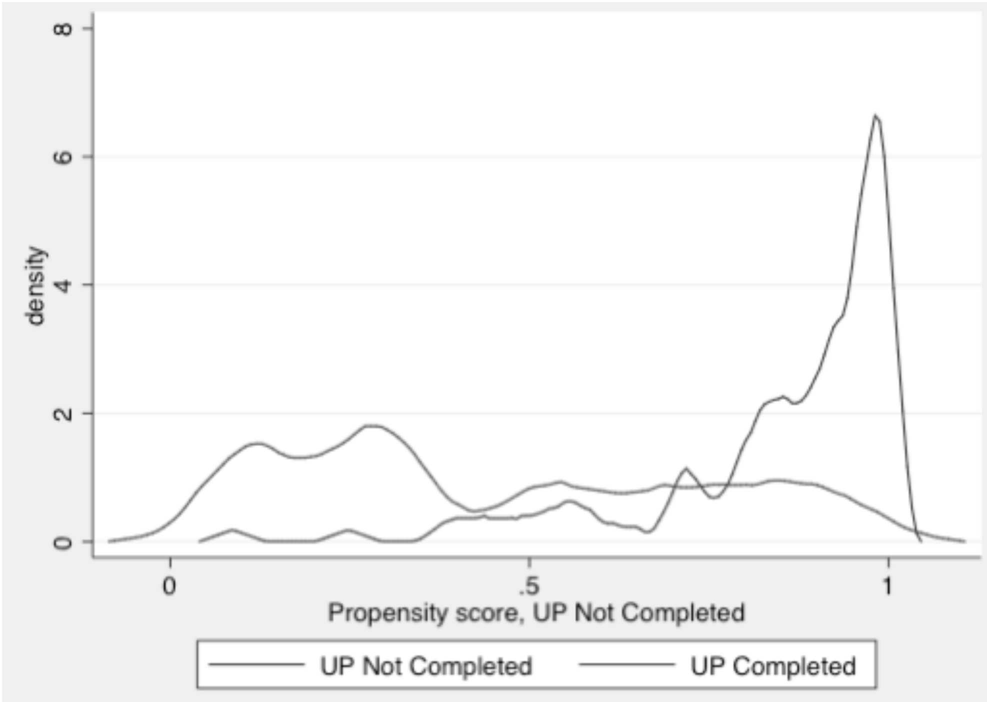

Figure 3. Propensity Score Density (Common Support) for UP (Right) and non-UP (Left)

BIBLIOGRAPHY

References

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine, 34*(28), 3661-3679.

Bazemore, A. W., Wingrove, P., Petterson, S. M., Peterson, L., Raffoul, M., & Phillips Jr, R. L. (2015). Graduates of teaching health centers are more likely to enter practice in the primary care safety net. *American Family Physician, 92*(10)

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31-72.

Chen, F. M., Bauchner, H., & Burstin, H. (2004). A call for outcomes research in medical education. *Academic Medicine, 79*(10), 955-960.

Cook, D. A., & West, C. P. (2013). Perspective: Reconsidering the focus on "outcomes research" in medical education: A cautionary note. *Academic Medicine : Journal of the Association of American Medical Colleges, 88*(2), 162-167. doi:10.1097/ACM.0b013e31827c3d78; 10.1097/ACM.0b013e31827c3d78

Council on Graduate Medical Education. (2010). *Advancing primary care.* (No. Twentieth).Health Resources and Services Adminstration.

Doescher, M. Fordyce, M. Skillman, S. Jackson, J. Rosenblatt, R. (2009). *Policy brief: Persistent primary care health professional shortage areas (HPSAs) and health care access in rural* *america.* (Policy brief). Seattle, WA: WWAMI Rural Health Research Center, University of Washington.

Flex Monitoring Team. (2016). Critical access hospital locations. Retrieved from http://www.flexmonitoring.org/data/critical-access-hospital-locations/

Guo, S., & Fraser, M. W. (2014). *Propensity score analysis: Statistical methods and applications: Statistical methods and applications* Sage Publications.

Ko, M., Heslin, K. C., Edelstein, R. A., & Grumbach, K. (2007). The role of medical education in reducing health care disparities: The first ten years of the UCLA/drew medical education program. *J Gen Intern Med, 22*(5), 625-31.

Luo, Z., Gardiner, J. C., & Bradley, C. J. (2010). Applying propensity score methods in medical research: Pitfalls and prospects. *Medical Care Research and Review : MCRR, 67*(5), 528-554. doi:10.1177/1077558710361486 [doi]

Morris, C. G., Johnson, B., Kim, S., & Chen, F. (2008). Training family physicians in community health centers: A health workforce solution. *Family Medicine, 40*(4), 271.

Phillips, R. L., Petterson, S., & Bazemore, A. (2013). Do residents who train in safety net settings return for practice? *Academic Medicine : Journal of the Association of American Medical Colleges, 88*(12), 1934-1940. doi:10.1097/ACM.0000000000000025; 10.1097/ACM.0000000000000025

Rabinowitz, H. K., Diamond, J. J., Markham, F. W., & Wortman, J. R. (2008). Medical school programs to increase the rural physician supply: A systematic review and projected impact of widespread replication. *Academic Medicine : Journal of the Association of American Medical Colleges, 83*(3), 235-243. doi:10.1097/ACM.0b013e318163789b

Rosenbaum, P. (2010). *Design of observational studies (springer series in statistics)*. New York: Springer.

Rosenbaum, P. (2005). Sensitivity analysis in observational studies. In B. S. Everitt, & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1809-1814). Chichester: John Wiley & Sons, Ltd,.

Rosenblatt, R. A., Andrilla, C. H., Curtin, T., & Hart, L. G. (2006). Shortages of medical personnel at community health centers: Implications for planned expansion. *JAMA : The Journal of the American Medical Association, 295*(9), 1042-1049. doi:10.1001/jama.295.9.1042

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.

Specialty and geographic distribution of the physician workforce: What influences medical student & resident choices? (2009). The Robert Graham Center. Retrieved from http://www.graham-center.org/online/graham/home/publications/monographs-books/2009/rgcmo-specialty-geographic.html

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science : A Review Journal of the Institute of Mathematical Statistics, 25*(1), 1-21. doi:10.1214/09-STS313 [doi]

APPENDIX A – SURVEY INSTRUMENT

E-2010 Students **NAME: ____________________________**

University of Washington School of Medicine
Biographical and Career Preference Inventory

I. Biographical Data

1. Please indicate the highest educational level attained to date by your father, mother, and (if married) spouse. Place the appropriate number in the space before each title.

_____ Father	1 = High school, grade 11 or below
	2 = High school graduate
	3 = Post-high school - No Certificate or Degree
_____ Mother	4 = Community/Junior College Program
	5 = Bachelor's Degree
	6 = Bachelor's Degree and Graduate School Hours
_____ Spouse	7 = Master's Degree
	8 = Master's Degree and Additional Hours
	9 = Doctoral Degree or Above

2. Indicate the present employment (or the previous employment, if no longer working or deceased) of your father, mother and spouse.

_____ Father	1 = Businessman/Businesswoman
	2 = Clerical, "white collar" worker
	3 = Craftsman, skilled trade
_____ Mother	4 = Housewife/Househusband
	5 = Professional (Lawyer, Nurse, Physician, etc.)
	6 = Semi-skilled/Unskilled Worker
_____ Spouse	7 = Technician (e.g. Lab, Repairmen, Technician)
	8 = Agriculture/Farmer/Rancher
	9 = Other

3. If either of your parents or your spouse has ever been or is presently engaged in health or health related occupations, indicate using the numbers listed below.

_____ Father	1 = Clinical Psychologist, Social Worker
	2 = Dentist
	3 = Dental Hygienist or Assistant
_____ Mother	4 = Hospital or Health Care Administrator
	5 = Lab, X-ray or Other Technician
	6 = Nurse
_____ Spouse	7 = Pharmacist
	8 = Physician
	9 = Physician Assistant, Medex
	0 = Other

4. Indicate the number of living sibling in your family (including yourself) _____

5. Indicate your position among the siblings (1=oldest, 2=second, etc.) _____

6. Indicate the size of the community in which you spent most of your pre-college years.

_____ Birth to 7 years of age
_____ 8 to 13 years of age
_____ 14 to 18 years of age

1 = Metropolitan, Over 1 million
2 = Metropolitan, 500,000 - 1 million
3 = Metropolitan, 100,000 - 500,000
4 = Large City 50,000 - 100,000
5 = Medium City 25,000 - 50,000
6 = Small City 10,000 - 25,000
7 = Large Town 5,000 - 10,000
8 = Medium Town 1,000 - 5,000
9 = Small Rural Town (Under 1,000)

7. Indicate the state or country in which you spent most of your pre-college years.

_____ Birth to 7 years of age
_____ 8 to 13 years of age
_____ 14 to 18 years of age

1 = Washington
2 = Alaska
3 = Montana
4 = Idaho
5 = Other U.S. State
6 = Foreign Country
7 = Wyoming

8. Indicate the distance you were from the nearest metropolitan are during most of your pre-college years.

_____ Birth to 7 years of age
_____ 8 to 13 years of age
_____ 14 to 18 years of age

1 = Over 200 Miles
2 = 100 - 200 Miles
3 = 50 - 100 Miles
4 = Under 50 Miles

9. Indicate your accessibility to the nearest metropolitan area during most of your pre-college years.

_____ Birth to 7 years of age
_____ 8 to 13 years of age
_____ 14 to 18 years of age

1 = Very accessible
2 = Moderately accessible (longer drive or inconvenient access)
3 = Low accessibility (difficult drive and/or difficult access)

10. Please provide the name of the High School and City from which you graduated.

____

II. CAREER PREFERENCES

1. Career Specialty: Please rate all of the following areas of medicine in terms of your current inclination to practice in them. (Indicate your likelihood by placing the number from the following scale before each career title).

1 = Strongly inclined to **AVOID**
2 = Moderately inclined to **AVOID**
3 = Neutral
4 = Moderately inclined to **SELECT**
5 = Strongly inclined to **SELECT**

_____ Family Medicine
_____ Allergy
_____ Anesthesiology
_____ Basic Medical Sciences
_____ Cardiology
_____ Dermatology
_____ Emergency Medicine
_____ Gastroenterology
_____ General Surgery
_____ Internal Medicine
_____ Neurology
_____ Neurosurgery
_____ Obstetrics/Gynecology
_____ Oncology
_____ Ophthalmology
_____ Orthopedics
_____ Otolaryngology
_____ Pathology
_____ Pediatrics
_____ Physical Medicine & Rehabilitation
_____ Plastic Surgery
_____ Psychiatry
_____ Pulmonary Medicine
_____ Public Health/Policy Studies
_____ Radiology
_____ Thoracic Surgery
_____ Urology
_____ Other Specialty

2. Type of Practice: Please indicate your current inclination or disinclination for each of the following characteristics of medical practice. Please use the following scale:

1 = Strongly inclined to **AVOID**
2 = Moderately inclined to **AVOID**
3 = Neutral
4 = Moderately inclined to **SELECT**
5 = Strongly inclined to **SELECT**

A. Depth and breadth of specialization
_____ Primary Care
_____ Tertiary Care

If **TERTIARY** Care, of what variety?
_____ Highly specialized practice
_____ Administration
_____ Research
_____ Academic

B. Please rate each of the following career/practice settings, or affiliations, in terms of your current inclination or disinclination to work in them. Use the same scale previously provided.

Private Practice Facilities

_____ Solo private practice
_____ Shared private practice with one other physician
_____ Private group practice of one specialty
_____ Small group practice of mixed specialties
_____ Large group practice, clinic based
_____ Large group practice, hospital based

Public Health Care Agencies

_____ Veteran's Affairs Hospital
_____ Military Hospital
_____ City, County or State Hospital
_____ City, County or State Health Department
_____ Public Health Service

Medical Schools

_____ As a basic science professor
_____ As a clinical professor
_____ As a research physician
_____ As an administrator

Other Employment Settings

_____ Research organization, institute or foundation
_____ Industry
_____ Pharmaceutical corporation
_____ Insurance company
_____ Church or religious medical position
_____ Other: Please specify ____

C. Please indicate your inclination or disinclination to practice under each of the following styles or managing payments for medical services

_____ Fee for service
_____ Pre-paid service (e.g. Group Health, HMO's, PPO's)

D. Please rate your inclination or disinclination to direct your work to each of the following categories of medical care:

_____ Crisis or acute care problems
_____ Chronic care, on-going management concerns
_____ Health maintenance or preventive care programs.

E. Location of Practice: Please indicate your current inclination or disinclination to practice in a community with each of the following characteristics. Please use the following scale:

1 = Strongly inclined to **AVOID**
2 = Moderately inclined to **AVOID**
3 = Neutral
4 = Moderately inclined to **SELECT**
5 = Strongly inclined to **SELECT**

Region

_____ Washington
_____ Wyoming
_____ Alaska
_____ Montana
_____ Idaho
_____ Other U.S. State ________________________________
_____ Other Foreign Country ________________________________

Size of Community

_____ Large Metropolitan - over 1 million population
_____ Medium Metropolitan - 500,000 to 1 million population
_____ Small Metropolitan - 100,000 to 500,000 population
_____ Large City - 50,000 to 100,000 population
_____ Medium City - 25,000 to 50,000 population
_____ Small City - 10,000 to 25,000 population
_____ Large Town - 5,000 to 10,000 population
_____ Medium Town - 1,000 to 5,000 population
_____ Small Town - under 1,000 population

Distance from nearest Metropolitan Area

_____ Over 200 miles
_____ 100 - 200 miles
_____ 50 - 100 miles
_____ Under 50 miles

Accessibility to Metropolitan Area

_____ Very accessible - easy car access or public transportation
_____ Moderately accessible - longer drive or somewhat costly and/or inconvenient public transportation
_____ Low accessibility - difficult or impossible to drive and/or costly/infrequent public transportation

F. Please indicate your inclination or disinclination to work within the following areas.

_____ URBAN socio-economically deprived/depressed area
_____ RURAL socio-economically deprived/depressed area
_____ FOREIGN socio-economically deprived/depressed area

III. Language Skills and Background

1. Was English your first spoken language? _____ Yes _____ No

2. If English was **NOT** your first spoken language, at what age did you begin to speak English?
   _____ Age 5 or younger
   _____ Age 6 to 10
   _____ Age 11 to 15
   _____ Age 16 or older

3. What was the dominant language in your household prior to your completion of high school?
   _____ English
   _____ Another language: ________________________

4. What language do you know the best?
   _____ English
   _____ English and another language about equally
   _____ Another language: ________________________

IV. Current Career Plans

1. What is your current preferred medical specialty? ____________________

2. How certain are you that you will actually practice within this medical specialty?

1	2	3	4	5
Not at all certain		Moderately Certain		Absolutely Certain

3. What other specialties, if any, are you also considering?
   Second choice: __________________________
   Third choice: __________________________

4. How important are the following factors in your currently preferred specialty.
   (1 = Very Unimportant, 3 = Neutral, 5 = Very Important)

Emphasis on patient education and prevention	1	2	3	4	5
Emphasis on procedural skills	1	2	3	4	5
Emphasis on primary care	1	2	3	4	5
Specialty is compatible with my personality	1	2	3	4	5
Emphasis on continuity of care	1	2	3	4	5
Opportunity to work with acute problems	1	2	3	4	5
Preference for working in an office/clinic setting	1	2	3	4	5
Preference for working in a rural environment	1	2	3	4	5
Preference for working with underserved populations	1	2	3	4	5
Preference for working with diverse patient populations	1	2	3	4	5
Having close personal relationships with patients	1	2	3	4	5
Diagnostic challenge/intellectual content of the specialty	1	2	3	4	5
Geographic flexibility in seeking employment	1	2	3	4	5

APPENDIX B – STATA CODE

Logistic Regression (Naïve Approach)

```
. logit yanysite txupcompleted

Iteration 0:   log likelihood = -107.68739
Iteration 1:   log likelihood =  -106.8005
Iteration 2:   log likelihood = -106.79944
Iteration 3:   log likelihood = -106.79944

Logistic regression                       Number of obs     =        158
                                          LR chi2(1)        =       1.78
                                          Prob > chi2       =     0.1827
Log likelihood = -106.79944               Pseudo R2         =     0.0082

------------------------------------------------------------------------------
    yanysite |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
txupcompleted |    .507744   .3863777     1.31   0.189    -.2495425     1.26503
       _cons |   .1854032    .184128     1.01   0.314    -.1754811    .5462875
------------------------------------------------------------------------------
```

Multiple Logistic Regression (MLR)

```
. logit yanysite txupcompleted gender age poc ell pedu1 pedu3 undback comanyl10  distanym20
> 0 aclow privprac empprac vamprac pubprac acadprac industprac otherprac metro city town lo
> wacc  ruralenv otherund ouptrel fm1st tertspec
Iteration 0:   log likelihood = -107.68739
Iteration 1:   log likelihood = -83.038946
Iteration 2:   log likelihood = -82.448652
Iteration 3:   log likelihood = -82.445991
Iteration 4:   log likelihood = -82.445991

Logistic regression                         Number of obs    =        158
                                            LR chi2(27)      =      50.48
                                            Prob > chi2      =     0.0040
Log likelihood = -82.445991                 Pseudo R2        =     0.2344
```

yanysite	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
txupcompleted	1.276203	.5829324	2.19	0.029	.1336763	2.418729
gender	-.3049079	.4949599	-0.62	0.538	-1.275011	.6651956
age	-.087616	.0507703	-1.73	0.084	-.1871239	.011892
poc	1.295989	.7479216	1.73	0.083	-.1699105	2.761888
ell	-1.568206	.9766235	-1.61	0.108	-3.482353	.3459411
pedu1	-.3239083	.3634003	-0.89	0.373	-1.03616	.3883432
pedu3	-.6709056	.387387	-1.73	0.083	-1.43017	.0883589
undback	.3840982	.7233984	0.53	0.595	-1.033737	1.801933
comanyl10	.5593642	.5568127	1.00	0.315	-.5319687	1.650697
distanym200	.3742684	.635668	0.59	0.556	-.871618	1.620155
aclow	-.036598	.7237478	-0.05	0.960	-1.455118	1.381922
privprac	.0542789	.2918004	0.19	0.852	-.5176393	.6261971
empprac	.284204	.3148431	0.90	0.367	-.3328772	.9012852
vamprac	-.3605592	.304204	-1.19	0.236	-.9567882	.2356697
pubprac	.364414	.3799323	0.96	0.337	-.3802395	1.109068
acadprac	.779674	.354126	2.20	0.028	.0855998	1.473748
industprac	-.6943397	.4526331	-1.53	0.125	-1.581484	.1928049
otherprac	-.3646183	.1821873	-2.00	0.045	-.7216989	-.0075377
metro	.3488362	.265404	1.31	0.189	-.1713461	.8690185
city	.5590482	.293751	1.90	0.057	-.0166931	1.13479
town	.6439939	.29207	2.20	0.027	.0715472	1.216441
lowacc	-.1077964	.3082512	-0.35	0.727	-.7119577	.4963648
ruralenv	-.7165574	.3800545	-1.89	0.059	-1.461451	.0283356
otherund	.1807179	.2966813	0.61	0.542	-.4007669	.7622026
ouptrel	.0135319	.3957943	0.03	0.973	-.7622107	.7892745
fm1st	-.9182077	.4975663	-1.85	0.065	-1.89342	.0570044
tertspec	-.6576604	.513876	-1.28	0.201	-1.664839	.3495182
_cons	1.033508	3.123572	0.33	0.741	-5.08858	7.155596

PS Matching (PSM)

```
. teffects psmatch (yanysite) (txupcompleted gender age poc ell pedu1 pedu3 undback comanyl
> 10  distanym200 aclow privprac empprac vamprac pubprac acadprac industprac otherprac metr
> o city town lowacc  ruralenv otherund ouptrel fm1st tertspec)

Treatment-effects estimation                Number of obs      =        158
Estimator      : propensity-score matching     Matches: requested =          1
Outcome model  : matching                              min =          1
Treatment model: logit                                 max =          1
------------------------------------------------------------------------------
             |              AI Robust
    yanysite |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ATE          |
txupcompleted |
   (1 vs 0)  |   .2341772   .0765258     3.06   0.002     .0841895     .384165
------------------------------------------------------------------------------
```

IPW

```
. teffects ipw (yanysite) (txupcompleted gender age poc ell pedu1 pedu3 undback comanyl10
> distanym200 aclow privprac empprac vamprac pubprac acadprac industprac otherprac metro ci
> ty town lowacc  ruralenv otherund ouptrel fm1st tertspec)

Iteration 0:   EE criterion =  8.565e-27
Iteration 1:   EE criterion =  4.284e-32

Treatment-effects estimation                    Number of obs     =        158
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
------------------------------------------------------------------------------
             |               Robust
    yanysite |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ATE          |
txupcompleted |
   (1 vs 0)  |   .2088311   .1587831     1.32   0.188    -.1023781    .5200402
-------------+----------------------------------------------------------------
POmean       |
txupcompleted |
          0  |   .5524305   .0612761     9.02   0.000     .4323315    .6725296
------------------------------------------------------------------------------
```