# The Educational Pipeline for Health Care Professionals: Understanding the Source of Racial Differences

Ivora Hinton, Jessica Howell, Elizabeth Merwin,
Steven Stern, Sarah Turner, Ishan Williams, and Melvin Wilson*
University of Virginia, Department of Economics and School of Nursing

August 2007

**Abstract**

While the representation of black students in medical schools rose dramatically from approximately 2 percent in 1965 to more than 7 percent in 1975, gains in the representation of blacks among health professionals have slowed, with blacks currently representing slightly more than 8 percent of first year medical students in the U.S. The underrepresentation of black Americans in the healthcare professions may have direct implications for the health outcomes of minority patients, underscoring the importance of understanding how individual characteristics, student achievement, and undergraduate experience differentially affect the completion of graduate training in the health sciences by race. We specify a model of individuals' postsecondary decisions including college enrollment, college type, and baccalaureate degree completion, jointly with the decision to enter a health care occupation that requires an advanced degree. We estimate the parameters of the model with maximum likelihood using data from the National Longitudinal Study class of 1972. Our estimates highlight the importance of factors operating prior to post-baccalaureate study, such as pre-collegiate achievement, in explaining the differential representation by race in the health professions. Without accounting for the impact of college type on the likelihood of completing a bachelor's degree, blacks appear to be somewhat more likely to pursue a career in the health professions. In contract, blacks are less likely to pursue a career in the health professions once we account for the impact of college type on the race-specific likelihood of baccalaureate degree completion. Our results emphasize the importance of jointly examining the full chain of educational decisions in understanding racial differences in representation within professional healthcare occupations.

*We would like to thank Michelle Bucci and Elizabeth Katz for excellent research assistance.

# 1 Introduction

Over the last half century, the representation of blacks in the pool of health professionals with graduate education (*e.g.*, physicians, dentists, psychologists, etc.), as well as other careers requiring post-baccalaureate training, has grown episodically rather than continuously.[1,2] Immediately after the passage of the Civil Rights Act, barely 2 percent of all medical students were black. Just a decade later in 1975, more than 7 percent of first year medical students were black. In the subsequent quarter century, however, gains in black representation among health professionals have slowed with blacks currently comprising slightly more than 8 percent of first year medical students (Association of American Medical Colleges, 2005). While recent cohorts entering medical school are unquestionably more racially diverse than those entering the profession three decades ago, blacks still receive advanced training in the health sciences at rates far below their population share of about 15.4 percent (U.S. Census Bureau, 2007).

The underrepresentation of blacks in the health professions is a concern not just for reasons of social equality, but also because members of the black community may have unique health care needs that may be better addressed and more successfully treated by black health care professionals who are knowledgeable about cultural aspects of health and care. Blacks have significantly more health problems than other groups, including high rates of diabetes, heart disease, prostate cancer, HIV/AIDS, breast cancer, and infant mortality (U.S. Department of Health and Human Services, 2001). Many of these health disparities can be explained partially by demographic factors, lack of health insurance, and decreased access to care or inferior care. If black health care professionals possess some comparative advantage in treating black patients, the underrepresentation of black health care professionals would have a direct effect on health outcomes in the black community and the attendant racial gaps in health.

The medical literature on this topic examines four hypotheses on the effect of minority health care professionals on the health outcomes of minority patients: service patterns, concordance, trust in health care, and professional advocacy (U.S. Department of Health and Human Services, 2006).[3,4] The service patterns hypothesis states that minority health professionals improve access to care for members of minority communities because they are more likely than their non-minority health professional peers to serve minority populations. The empirical evidence

[1] Health professionals, along with their representation in the data analyzed, include physicians (33.2%), therapists (17.4%), dentists (14.3%), registered nurses (6.1%), pharmacists (5.1%), psychologists (4.6%), optometrists (4.1%), veterinarians (3.6%), biological scientists (3.6%), dieticians (2.0%), health technicians (1.5%), podiatrists (1.0%), and chiropractors (0.5%). Additionally, because we define health professional status as being in a health occupation and possessing a post-baccalaureate degree, there are a very small number of other health services technicians that are categorized as health professionals.

[2] "Black" is used in the dataset as the category for racial identification and will be used in this paper as the more inclusive term representing African Americans and other black individuals.

[3] Most studies focus on the effect of minority physicians on minority patient outcomes. The empirical evidence for non-physician health professionals (*e.g.,* therapists, registered nurses, dentists, etc.) is more limited.

[4] In the medical literature, the minority groups examined are traditionally underrepresented groups including blacks, Latinos, and Native Americans. Primarily due to data limitations, we focus exclusively on blacks in the present study.

generally supports this hypothesis. Saha (2000) and Bach et. al. (2004) find that black physicians are more likely to locate in areas with larger proportions of blacks. Komaromy (1996) and Penn (1986) find that this result is also present among minority physicians who are graduates of some of the most prestigious medical schools in the U.S., who presumably would have substantial choice over their location. Additionally, data indicate that black doctors serve a disproportionate share of black patients, relative to their representation among all physicians.[5] While this latter result is partially due to black patients' higher propensity to choose a black physician, the analyses also clearly demonstrate that minority physicians' choice of practice location plays a significant role (Bach et. al., 2004 and Saha, 2000). Even in light of this evidence on service patterns among black health professionals, one still needs to believe that increased access to care by the black community leads to improved health outcomes for those individuals.[6,7]

The concordance hypothesis states that race-concordant care (*e.g.,* a black patient visiting a black physician) may be associated with greater trust by the patient and better communication between individuals and health care professionals regarding seriousness of illness and proper implementation of prescribed treatment (Rosenheck, 1995 and Cooper, et. al., 2003). Race concordance may improve minority health outcomes via several channels, namely, by increasing access to care, utilization of care, and/or the quality of the care received.[8] Research by Kaplan, Greenfield, and Ware (1989), Giron, et. al. (1998), Stewart (1995), and Ware and Davies (1983) suggests that better communication between race-concordant patient-physician pairs is associated with greater patient involvement in decision-making and higher overall patient satisfaction, which is associated with improved continuity of care, timely and accurate diagnoses, adherence to effective programs, and health outcomes. Finally, Saha (1999) finds that race concordance among blacks is also associated with greater likelihood of preventive care and a smaller probability of foregoing needed health care.

The trust in health care hypothesis states that greater health care workforce diversity increases minority community members' propensity to use health services by increasing their trust in the system. The professional advocacy hypothesis states that minority health care professionals are more likely to advocate for policies and programs designed to improve health care access, usage, and quality, and thereby, health outcomes, among minority populations. While both the trust and professional advocacy hypotheses seem plausible, there exists insufficient empirical evidence to make statements about the merit of either. Finally, the motivation for increasing minority representation among health professionals may simply be a production function story. As the Association

[5] Bach et. al. (2004) find that, in a national sample of Medicare beneficiaries, 22 percent of black patient visits nationwide were to black physicians, which is substantially greater than the average proportion of black physicians within the areas where black patients sought care (12.5 percent) and the proportion of black physicians nationally (5.5 percent).

[6] To date, there have been no randomized trials that explicitly test for this link (U.S. Department of Health and Human Services, 2006).

[7] Stern et. al. (2007) find that interactions of availability of black health professionals with individual measures of race have significant effects on various measures of mental and physical health.

[8] A variety of additional studies focus on language-concordant care (see Flaskerud, 1990), which is not entirely relevant given our focus on blacks.

of American Medical Colleges argued in an amicus brief (2002) in the Supreme Court case regarding the use of affirmative action in University of Michigan graduate admissions, racial diversity among students in medical education is a direct input to the training of all physicians, producing physicians who are "culturally competent" and "who are better prepared to serve a varied patient population."[9]

A variety of national programs have attempted to increase minority representation among health professionals. Some of these programs, like the Minority Medical Education Program sponsored by the Robert Wood Johnson Foundation, seek to improve academic preparation for medical school, thereby increasing the number of minorities who complete advanced degrees in the health sciences. Other programs combine increased minority recruiting and academic preparation with financial subsidies for graduate school or for serving underserved populations following advanced degree completion (*e.g.,* the Health Careers Opportunity Program, Centers of Excellence, Minority Faculty Fellowship Program, and National Health Service Corps). Although there is empirical evidence to suggest that some of these programs successfully increase racial and ethnic diversity in health professional schools (Grumbach et. al., 2003), it is not clear that these successes translate into similar gains in diversity among the health care workforce or into improved access for underserved populations (U.S. Department of Health and Human Services, 2006).

The question addressed in this paper concerns how individual characteristics and achievement observed at the precollegiate level affects the chain of decisions leading to training as a health care professional by race. We evaluate how the representation of health care professionals, by race, would differ if observed between-group differences, such as gaps in parental education, were eliminated. We trace individuals' decisions about college enrollment, college degree completion, advanced degree completion, and choice of a health occupation that requires an advanced degree in the context of a unified economic model that allows for the correlation of unobservable determinants of each of these outcomes. For example, if individuals who are more likely to complete baccalaureate degrees are also more likely to complete advanced degrees in the health sciences for unobservable reasons, then simple estimates of the determinants of the decision to become a health professional would be biased. By jointly modeling these decisions, we are able to examine the extent to which the overall "leakage" from the pipeline into a professional health care occupation stems from precollegiate factors, differences in collegiate attainment, or gaps at the transition from undergraduate to graduate study in the health sciences. Finally, our parameter estimates enable us to focus on how changes in the precollegiate characteristics of students over time might be expected to narrow the racial gaps in professional degree attainment in the health sciences.

The paper begins with the consideration of the historical context of the underrepresentation of black Americans in the health professions and college completion more generally. Section 3 presents a theoretical model of the individual decisions described above and then generalizes that basic model to allow for variation in the type of postsecondary institutions individuals choose to attend for their baccalaureate training. The data is described

[9] The Court ruled against the undergraduate admissions policy at the University of Michigan in *Gratz v. Bollinger, et al.* and supported the "narrowly tailored" use of race by the University of Michigan law school in *Grutter v. Bollinger, et al.*.

in Section 4, results are discussed in Section 5, and model fit is examined in Section 6. An analysis of how racial gaps in the health professions might evolve over time based on the estimates of the model parameters are presented in Section 7. Section 8 concludes.

# 2 Historical Context

There is no question that segregated universities and labor market discrimination limited the incentives and opportunities for black Americans to pursue advanced training in the health care fields in the first part of the twentieth century. Yet, there is clear evidence that these institutions changed dramatically in the late 1960s and early 1970s, providing new incentives for blacks to enter the health professions (Freeman, 1976). The legacy of segregation and discrimination in education affects the contemporary representation of blacks in the health care professions through two channels. First, the stock of health care professionals today includes those educated in the 1960s when there were greater institutional limits on the participation of blacks in medical education and, as such, it will take several decades before the overall representation of blacks in the health care professions reflects the current representation among education in health sciences programs. Secondly, segregation and discrimination in prior decades may contribute to racial differences in parental wealth, parental education, and pre-collegiate educational opportunities which mediate collegiate attainment that is a prerequisite for attainment of advanced degrees in the health professions. Because changes in the representation of blacks among health care professionals may be largely a function of differences in educational outcomes well before graduate school, we begin by tracing out changes in the overall gaps in educational attainment by race.

## 2.1 Overall Educational Attainment

The historical underrepresentation of blacks in the health sciences is embedded in the history of separate and unequal schooling opportunities in the United States, in both the elementary and secondary grades and at the collegiate level. For those born at the start of the twentieth century, the educational attainment of whites was nearly double that of blacks with a gap of about 3.6 years of schooling (DeLong, Goldin, and Katz, 2003). Well before the modern Civil Rights movement, the gap began to close, starting with those born in about 1910 and continuing through the cohorts born in the 1960s. Despite the narrowing of this gap, there remains a persistent gap between blacks and whites of about 0.6 years of total schooling even for those born in the 1970s (DeLong, Goldin, and Katz, 2003). Figure 1 indicates that the racial gap in high school completion among 30-34 year olds has narrowed dramatically over the last 35 years, shrinking from 23 percentage points to about 2.5 percentage points.

Segregated institutions affected black Americans at all levels of education in the first part of the twentieth century. Particularly in the South, opportunities for collegiate study often were limited to a set of underfunded colleges as Southern states maintained explicitly segregated systems of education in the 1940s, with slightly more than 100 institutions of higher education for blacks. Before the Supreme Court dismantled elementary and secondary segregation through the historic *Brown v. Board of Education* case in 1954, cases in Missouri and Texas opened opportunities for minorities in graduate professional schools. Yet, the initial judicial desegregation proved insufficient in generating change at a number of universities in the South, with the University of Mississippi and the University of Alabama remaining segregated into the early 1960s.

The 1960s brought dramatic changes affecting both educational choices and labor market opportunities for black Americans. In 1964, Johnson signed into law the Civil Rights Act followed by an executive action providing enforcement to anti-discrimination measures. The Johnson administration went further, calling for "affirmative efforts to provide opportunities for black Americans" (Bowen and Bok, 1998). Spurred by campus activism, many leading colleges and universities went beyond compliance in employment practices and made active efforts to recruit black students. Indeed, there were dramatic changes in the representation of black students at leading colleges and universities, with black representation in Ivy League institutions rising from 2.3 percent in 1967 to 6.3 percent in 1976 (Karen, 1991). Selective colleges and universities initiated proactive programs to increase the representation of minority students in both graduate and undergraduate programs.

Growth in the educational attainment of black students in the late 1960s and early 1970s went beyond the boundaries of elite colleges to include both increased high school graduation and greater collegiate attendance at a range of postsecondary institutions. Figure 2 shows the proportion of the population ages 30-34 with some college attainment from 1968 to 2003, while Figure 3 presents the data for college graduation by race. The gap in collegiate attainment between black and white adults narrows appreciably over the interval for both measures, although there remains a non-trivial difference in outcomes. In 2003, white adults were about 15 percent (7.3 percentage-points) more likely than black adults to have some college experience and 57 percent (8.3 percentage-points) more likely to have completed college (Current Population Survey, 2003). The causes of these differences in collegiate attainment are of significant policy concern in their own right, likely reflecting a combination of differences in secondary school quality, family background, and the capacity to finance college. At the same time, the persistence of these differences suggests that some of the racial gap in the health professions is caused by differences generated earlier in the educational pipeline.

While some of the narrowing of the gap between black and white students in college enrollment in the 1970s represents structural changes in opportunities brought about by the Civil Rights movement, it is also the case that some of the growth in the representation of black Americans in college can be traced to broader changes in socioeconomic conditions, including increased odds of parental high school attainment. Kane (1994) finds that rising relative educational attainment of black parents was an important determinant of high school graduation and college enrollment for blacks in the 1980s, though these effects were offset somewhat by rising tuition during this era. Emphasizing the magnitude of this change in recent decades, Kane (1994) notes that, "In 1970, approximately

32 percent of black 18-19 year old youths had mothers who were high school graduates. By 1988, that proportion had doubled to roughly 63 percent." In the econometric model, we investigate how a closing of the racial gap in achievement and parental characteristics would likely affect differences in educational outcomes. While the achievement gap between black and white students at the time of expected college entry has narrowed somewhat over the last three decades, progress has been slow and uneven. Krueger, Rothstein and Turner (2006) note that the black-white gap in the performance of 17-year old students on the National Assessment of Education Progress narrowed from over one standard deviation in 1970 to about three quarters of a standard deviation in reading (and a larger gap in math), though nearly all of the convergence occurred before 1990. One implication, which follows in our empirical analysis, is that differences in the representation of blacks and whites at the post-baccalaureate stage can be traced to gaps generated much earlier in the educational pipeline.

## 2.2 Graduate Training in Health Professions

Through the first part of the 20th century, two institutions - Howard University in Washington and Meharry Medical College in Nashville - trained the overwhelming majority of black physicians. Blackwell (1981) estimates that, in 1967, approximately 83 percent of the 6,000 practicing black physicians received training at one of these two schools. With the Civil Rights movement, many medical schools explicitly endorsed the objective of increasing minority representation in the health professions, and the Association of American Medical Colleges (AAMC) endorsed this position in 1968.

The combination of expected returns in the health professions and the new recruiting efforts of medical schools brought a dramatic increase in the representation of blacks in medical schools from the late 1960s through the early 1970s. The number of black students enrolled as first year medical students jumped from 266 in 1968 to 1106 in 1974, rising from 2.7 percent of the entering class to 7.5 percent (Figure 4) (Association of American Medical Colleges, 2005). The latter half of the 1970s and the 1980s brought some stagnation in the representation of black students in medical schools before the share rose again in the late 1980s.

The mid-1970s brought judicial scrutiny to efforts to increase the representation of minority students in medical schools through preferential admissions. While the 1974 *DeFunis v. Odegaard* case involving the differential treatment of minority candidates at the University of Washington Law School was dismissed, a case involving the application of Allan Bakke to medical school at the University of California at Davis entered the legal system in 1974 and lead to a landmark Supreme Court ruling in 1978. In a quite narrow ruling, the court held that admissions policies could not use a quota system or "set aside" places for minority students, but that student race could be considered among other factors in circumstances where racial diversity could be thought to yield educational benefits (Bowen and Bok, 1998).

Changes in the representation of black students in medical education translate to changes in the stock of minority health professionals with long lags. Not only is the process of medical education lengthy, but it also takes many years for cohorts with relatively low minority representation to retire and be replaced by cohorts that are somewhat more racially diverse.

## 3 The Basic Model

The model of individual behavior combines decisions about college enrollment, college degree completion, and choice of a health care occupation that requires an advanced degree. Assume that each individual, indexed by $i$, has some unobserved propensity to choose each of these outcomes, where their propensities are functions of individual and family characteristics denoted by $X_i$. In practice, $X_i$ contains information on gender, race, academic ability, parental education, and urbanicity of the location in which the individual attended high school. For each individual $i$, let $y_{1i}^*$ be the latent value of enrolling in college, $y_{2i}^*$ be the latent value of completing a four-year college degree conditional on enrolling, and $y_{3i}^*$ be the latent value of becoming a health professional with an advanced degree conditional on completing a four-year college degree. Each of these choices can be expressed as functions of observable individual-specific characteristics in $X_i$ as well as an unobservable component denoted by $u_{ji}$, $j = 1, 2, 3$;[10]

$$y_{1i}^* = X_{1i}\beta_1 + u_{1i} \tag{1}$$

$$y_{2i}^* = X_{2i}\beta_2 + u_{2i} \tag{2}$$

$$y_{3i}^* = X_{3i}\beta_3 + u_{3i}. \tag{3}$$

Define the vector of unobservables for individual $i$ as $u_i = (u_{1i}, u_{2i}, u_{3i})'$ and allow these unobservable factors to be correlated across individual $i$'s three choices by assuming $u_i \sim iidN(0, \Omega)$.

An individual's propensities to enter college, complete a baccalaureate degree, and select a health occupation that requires an advanced degree are all unobserved in the data. Instead, we observe binary outcomes indicating whether or not individual $i$ actually made these choices. Thus, let $y_{1i}$, $y_{2i}$, and $y_{3i}$ represent entry into college, completion of college, and entry into a health profession requiring an advanced degree, respectively. Mathematically, $y_{ki} = \prod_{l=1}^{k} 1(y_{li}^* \geq 0)$ for $k = 1$, 2, and 3, respectively. The definitions of these binary outcome variables are used to specify individual $i$'s probabilities of making various choices that are possible in the data, where the four possible outcomes and their associated probabilities, conditional on observable individual characteristics, are:

1. Do not enroll in college ($P_1 = \Pr[y_{1i} = 0 \mid X_{1i}]$);

[10] Note that $X_{ji} \subseteq X_i$ and $\bigcup_{j=1}^{3} X_{ji} = X_i$; however, we do not have to assume that $X_{ji} \cap X_{ki} = \emptyset$ for $j \neq k$ (*i.e.*, the explanatory variables for each set can have common elements). Also, because we have assumed that there are no endogenous variables in $X_i$, we do not need the typical identification conditions that are usually satisfied by having, for each equation, at least one variable belonging to $X_i$ having a zero restriction on the associated coefficient and not having zero restrictions in the other two equations for that variable.

2. Enroll in college, but do not graduate with a baccalaureate degree ($P_2 = \Pr[y_{1i} = 1, y_{2i} = 0 \mid X_{1i}, X_{2i}]$);

3. Enroll in college, graduate with a baccalaureate degree, but do not choose a health profession that requires an advanced degree ($P_3 = \Pr[y_{1i} = 1, y_{2i} = 1, y_{3i} = 0 \mid X_{1i}, X_{2i}, X_{3i}]$);

4. Enroll in college, graduate with a baccalaureate degree, and choose a health profession that requires an advanced degree ($P_4 = \Pr[y_{1i} = 1, y_{2i} = 1, y_{3i} = 1 \mid X_{1i}, X_{2i}, X_{3i}]$).

Each of these four probabilities are functions of the model parameters to be estimated, $\theta$, which include $\beta_1, \beta_2, \beta_3$, and $\Omega$, and are conditional on observed individual characteristics in $X_i$. The assumed joint normality of the unobservables ($u_{ki}$) in equations (1), (2), and (3) enable each of the four probabilities listed above to be expressed in terms of univariate, bivariate, and trivariate normal distribution and density functions. The detailed expressions for these four choice probabilities are presented in Appendix A.1.

The model parameters are estimated by maximum likelihood estimation (MLE), which involves specifying the log-likelihood function, to which each individual in the sample makes a contribution. An individual's log-likelihood contribution is the log probability of observing the choices made by that individual in the data, and it can be written as

$$\begin{aligned} L_i(\theta) &= (1 - y_{1i}) \log P_1(X_i; \theta) + y_{1i}(1 - y_{2i}) \log P_2(X_i; \theta) \\ &\quad + y_{1i} y_{2i} (1 - y_{3i}) \log P_3(X_i; \theta) + y_{1i} y_{2i} y_{3i} \log P_4(X_i; \theta). \end{aligned}$$

Summing over all individuals' log-likelihood contributions, the value of the parameters in $\theta$ that maximizes $L(\theta) = \sum_i L_i(\theta)$ is the maximum likelihood estimator of $\theta$.[11]

## 3.1 Decomposing the Effect of a Change in Individual Characteristics

We use the model described above to measure the predicted effect of increasing one of the explanatory variables, like black parental education, for example, on the probability that an individual becomes a health professional with an advanced degree. Furthermore, the predicted effect of increased black parental educational attainment can be decomposed into the component effects in each stage of the educational pipeline that we specify in the model. This is particularly useful for determining where in the pipeline black representation is predicted to be affected by such a change (*i.e.,* college entrance, college completion, or transition to health professional). For example, if $j$ indexes the different individual characteristics in which we are interested, the partial derivative of $P_4$ with respect to $X_{1ij}$ tells us the effect of increasing characteristic $j$ on the probability of becoming a health professional due to its effect on the propensity to enroll in college. The partial derivative of $P_4$ with respect to $X_{2ij}$ tells us the effect of increasing characteristic $j$ on the probability of becoming a health professional

[11] The asymptotic covariance matrix of the MLE $\widehat{\theta}$ can be estimated in the usual way as $\widehat{C}\left(\widehat{\theta}\right) = \left[\frac{1}{n}\sum_i \frac{\partial L_i(\widehat{\theta})}{\partial \theta}\frac{\partial L_i(\widehat{\theta})}{\partial \theta'}\right]^{-1}$.

conditional on enrolling in college due to its effect on the college completion. Finally, the partial derivative of $P_4$ with respect to $X_{3ij}$ provides the effect of increasing characteristic $j$ on the probability of becoming a health professional conditional on college completion. Thus, if characteristic $j$ is parental educational attainment, the three derivatives described here indicate how increased parental attainment would change an individual's probability of becoming a health professional at three important stages of the process; college enrollment, college completion, and post-baccalaureate career and degree decisions.

## 3.2 Altering the Model to Permit Variation in College Type and Quality

One issue that we abstract from in the basic theoretical model presented above is that college-bound individuals select and attend institutions of varying characteristics. The attributes of the college attended may be important in this model if variation in those attributes influences individuals' college completion rates, propensity to obtain an advanced degree, or propensity to choose a health care occupation.[12] In this section, we generalize the model so that colleges chosen at the baccalaureate level are permitted to differ along two dimensions: institutional quality (proxied by institutional selectivity) and whether the institution is a historically black college or university (HBCU).[13] We cannot simply add college quality and an HBCU indicator to the explanatory variables in equations (1), (2), and (3) because individuals *choose* these attributes through their application and enrollment decisions, making both variables endogenous. Instead, enrollment at colleges of varying quality or at an HBCU are modeled as additional latent choice variables.

Assume that $y_{1i}^*$ is a latent variable measuring the quality of the non-HBCU undergraduate college individual $i$ can attend or, more simply, $y_{1i}^*$ measures whether individual $i$ has the qualifications to be admitted to a non-HBCU college of a particular quality,

$$y_{1i}^* = X_{1i}\beta_1 + u_{1i}.$$

Next assume that $y_{2i}^*$ is a latent variable measuring the value of attending an HBCU.[14] It may be that an individual's propensity to choose an HBCU is a function of the quality of the non-HBCU colleges to which they could obtain admission. Thus, we allow $y_{2i}^*$ to be a function of $y_{1i}^*$ as well as observable individual characteristics and an unobservable component,

[12] Bowen and Bok (1999) demonstrate that graduate degree completion in general and completion of an MD in particular, is much higher among graduates of selective colleges and universities than among the overall pool of college graduates. Among graduates of the selective *College and Beyond* institutions, 56% of both blacks and whites went on to receive MA, professional, or PhD degrees; nationally, the share of college graduates completing advaned study are much lower with 34 percent of blacks and 38 percent of whites receiving advanced degrees (Figure 4.2, Bowen and Bok).

[13] An institution's status as historically black may be especially important for our research question regarding black representation in the health professions. According to the American Association of Medical Colleges, the top three undergraduate institutions that send black students to medical school (in percentage terms) are Xavier, Howard, and Spelman, which are all HBCUs. (http://www.aamc.org/data/facts/2005/mblack.htm)

[14] While there is some variation in institutional selectivity (our measure of quality) among HBCUs, we observe very few individuals enrolling in the highest quality HBCUs and it is not econometrically feasible to model quality variation in HBCUs.

$$y_{2i}^* = \alpha_1 y_{1i}^* + \alpha_2 (y_{1i}^*)^2 + X_{2i}\beta_2 + u_{2i}.$$

Note that $\alpha_1$ is not identified if $X_{1i} \subseteq X_{2i}$, which is the case given that we have a somewhat limited set of individual attributes in our dataset. Thus, we set $\alpha_1 = 1$ and think of $\beta_2$ as the degree to which $X_{2i}$ affects $y_{2i}^*$ in excess of $y_{1i}^*$, the value of attending college. Assume that individual $i$ attends an HBCU if and only if the value from doing so is positive, or $y_{2i} = 1(y_{2i}^* > 0)$.

Finally, we need to specify the quality of non-HBCU colleges attended by individuals in the sample and how this additional variation changes the basic model. Define college quality threshold values $\tau_k$, $k = 0, 1, .., K$, that divide up the support of $y_{1i}^*$ into regions consistent with the data. Individual $i$ attends a non-HBCU of quality level $k$ if and only if he/she does not attend an HBCU and if the quality of the non-HBCU attended falls between thresholds $\tau_k$ and $\tau_{k+1}$. Mathematically, we observe the set of $K$ possible non-HBCU college choices given by

$$y_{1i} = k \text{ iff } \tau_k < y_{1i}^* \leq \tau_{k+1} \cap y_{2i}^* < 0.$$

We can define $k = 0$ as the case of not attending college and allow college quality to be increasing in $y_{1i}^*$ and $k$. Without loss of generality, we can also define $\tau_0 = -\infty$, $\tau_1 = 0$, and $\tau_{K+1} = \infty$.

Equations (2) and (3) in the basic model specified the latent values of completing college and becoming a health professional with an advanced degree, respectively. We respecify these equations as

$$y_{3i}^* = X_{3i}\beta_3 + u_{3i};$$

$$y_{4i}^* = X_{4i}\beta_4 + u_{4i}.$$

Note that, in the equations for $y_{3i}^*$ and/or $y_{4i}^*$, we could have added interaction terms between $X_{ji}$ and attributes of the college chosen ( $y_{1i}^*$, and $y_{2i}$). In particular, this would have allowed graduation rates and/or entry into the health profession to depend upon interactions between race and attendance at an HBCU, college preparation measures like parental education and SAT score, and quality of the baccalaureate institution attended. Because our sample contains relatively small numbers of blacks in the college completion and health professional stages, including these race interactions is not feasible. In order to determine whether these interactions, in fact, matter, we construct Lagrange Multiplier tests of whether the actual quality of the school attended, $w_{1i} = y_{1i}^*(1 - y_{2i})$, and/or attendance at a HBCU, $y_{2i}$, affects the propensity to become a health professional, $y_{4i}^*$, for blacks. This involves respecifying $y_{4i}^* = X_{4i}\beta_4 + \gamma_1 w_{1i} + \gamma_2 w_{2i} + u_{4i}$, where $w_{2i}$ is $y_{2i}$ interacted with being black. Under the null hypothesis that $(\gamma_1, \gamma_2) = 0$, this corresponds to the model that we estimate below. The $\chi^2$ statistic for this test is 0.001, which supports the null hypothesis that these interaction terms do not matter.

As in the basic model, we assume that the vector of unobservable components of the latent variable equations are $u_i = (u_{1i}, u_{2i}, u_{3i}, u_{4i})'$ and $u_i \sim iidN(0, \Omega)$ with diagonal elements of $\Omega$ equal to one for identification purposes. Also, as in the basic model, all four $y_{ki}^*$ variables are latent and we actually observe the binary outcomes $y_{ki} = \prod_{l=1}^{k} 1\left(y_{li}^* \geq 0\right)$ for $k = 1$, 2, 3, and 4, respectively.

There are now seven possible outcomes we might observe in the data for each individual. These possible outcomes, along with their associated conditional probabilities of occurring in the data, are:

1. Do not enroll in college ($P_1 = \Pr\left[y_{1i} = 0, y_{21} = 0 \mid X_{1i}, X_{2i}\right]$);

2. Enroll in a non-HBCU of type $k$, but do not graduate with a baccalaureate degree ($P_{2k} = \Pr\left[y_{1i} = k, y_{2i} = 0, y_{3i} = 0 \mid X_{1i}, X_{2i}, X_{3i}\right]$);

3. Enroll in an HBCU, but do not graduate with a baccalaureate degree ($P_3 = \Pr\left[y_{2i} = 1, y_{3i} = 0 \mid X_{1i}, X_{2i}, X_{3i}\right]$);

4. Enroll in a non-HBCU of type $k$, graduate with a baccalaureate degree, but do not choose a health profession that requires an advanced degree ($P_{4k} = \Pr\left[y_{1i} = k, y_{2i} = 0, y_{3i} = 1, y_{4i} = 0 \mid X_{1i}, X_{2i}, X_{3i}, X_{4i}\right]$);

5. Enroll in an HBCU, graduate with a baccalaureate degree, but do not choose a health profession that requires an advanced degree ($P_5 = \Pr\left[y_{2i} = 1, y_{3i} = 1, y_{4i} = 0 \mid X_{1i}, X_{2i}, X_{3i}, X_{4i}\right]$);

6. Enroll in a non-HBCU of type $k$, graduate with a baccalaureate degree, and choose a health profession that requires an advanced degree ($P_{6k} = \Pr\left[y_{1i} = k, y_{2i} = 0, y_{3i} = 1, y_{4i} = 1 \mid X_{1i}, X_{2i}, X_{3i}, X_{4i}\right]$);

7. Enroll in an HBCU, graduate with a baccalaureate degree, and choose a health profession that requires an advanced degree ($P_7 = \Pr\left[y_{2i} = 1, y_{3i} = 1, y_{4i} = 1 \mid X_{1i}, X_{2i}, X_{3i}, X_{4i}\right]$).

The explicit forms of these probabilities are provided in Appendix A.2.[15]

Again, the probabilities discussed above form individual $i$'s log-likelihood contribution:

$$\begin{aligned}
L_i(\theta) \quad = \quad & 1\left(y_{1i} = 0\right)\left(1 - y_{2i}\right) \log P_1\left(X_{1i}, X_{2i}; \theta\right) \\
& + \left(1 - y_{2i}\right)\left(1 - y_{3i}\right) \sum_{k=1}^{K} 1\left(y_{1i} = k\right) \log P_{2k}\left(X_{1i}, X_{2i}, X_{3i}; \theta\right) \\
& + y_{2i}\left(1 - y_{3i}\right) \log P_3\left(X_{1i}, X_{2i}, X_{3i}; \theta\right) \\
& + \left(1 - y_{2i}\right) y_{3i}\left(1 - y_{4i}\right) \sum_{k=1}^{K} 1\left(y_{1i} = k\right) \log P_{4k}\left(X_{1i}, X_{2i}, X_{3i}, X_{4i}; \theta\right) \\
& + y_{2i} y_{3i}\left(1 - y_{4i}\right) \log P_5\left(X_{1i}, X_{2i}, X_{3i}, X_{4i}; \theta\right) \\
& + \left(1 - y_{2i}\right) y_{3i} y_{4i} \sum_{k=1}^{K} 1\left(y_{1i} = k\right) \log P_{6k}\left(X_{1i}, X_{2i}, X_{3i}, X_{4i}; \theta\right) \\
& + y_{2i} y_{3i} y_{4i} \log P_7\left(X_{1i}, X_{2i}, X_{3i}, X_{4i}; \theta\right),
\end{aligned}$$

[15] For some individuals in our data, we observe that they enroll in a non-HBCU four-year college, but the identity of the institution is unknown. The Technical Appendix also includes the way in which choice probabilities $P_2$, $P_4$, and $P_6$ are affected by this missing information.

and we maximize $\sum_i L_i(\theta)$ over the parameters in $\theta$ to get consistent, asymptotically normal estimates of $\theta$. These parameter estimates will also be used to decompose the effect of changing individual characteristics on choices made at various stages in the pipeline.

# 4 Data

The primary data we employ is from the National Longitudinal Study of the High School Class of 1972 (NLS-72). The National Center for Education Statistics (NCES) of the U.S. Department of Education designed and conducted this study and refers to it as "probably the richest archive ever assembled on a single generation of Americans" (NCES, 1994). Participants in the study were high school seniors in the spring of 1972, and follow-up surveys of these respondents were conducted in 1973, 1974, 1976, 1979, and 1986. The database contains information from high school records as well as postsecondary transcripts (collected in 1984). Because the original 18-year-old respondents were last interviewed when they were approximately 32-years-old, we believe this panel dataset is sufficiently long to allow individuals to acquire post-baccalaureate training and choose an occupation in a health profession. This dataset is supplemented with information on college and university selectivity rankings from *Barron's Profiles of American Colleges* (1994). We collapse the scale of ten selectivity rankings in *Barron's* into five categories such that higher level institutions are associated with higher quality and better reputation. Postsecondary institutions that are Historically Black Colleges and Universities (HBCUs) are coded as a separate category and not assigned a selectivity ranking. Additionally, attendance at a two-year, non-vocational postsecondary institution is considered college entry if the individual eventually completed a four-year bachelors degree.

Summary statistics for the sample of high school graduates, college entrants, college graduates, and health professionals with advanced degrees are provided in Table 1. In the sample of approximately 13,000 high school graduates, 72 and 73 percent of respondents' fathers and mothers, respectively, have at least a high school education, while 19 and 11 percent of respondents' fathers and mothers, respectively, have a baccalaureate or advanced degree. Consistent with early-1970s data from the October Current Population Survey analyzed in Kane (1994), 51 percent of our sample of high school graduates enroll in some type of non-vocational postsecondary institution. Table 1 also indicates the types of postsecondary institutions chosen. For example, 9.3 percent of high school graduates begin their college career at a two-year college, while 4.6 percent start at highly-selective (level 5) four-year institutions. Most college-bound high school graduates enter college at an institution of moderate selectivity, or level 3. Reading Table 1 from left-to-right, our sample changes in predictable ways as we follow these respondents through the educational pipeline from high school graduation through college entrance and completion and, finally, to becoming a health professional with an advanced degree. The sample becomes more male, less racially diverse, and socioeconomic status (proxied by parental educational attainment) increases.[16] The students who successfully complete each additional stage are also of higher academic ability, as proxied by student SAT score, and less likely to be from rural and farming communities.[17] Table 1 also indicates that nearly 60 percent of college entrants graduate with a baccalaureate degree and 5 percent of those degree recipients go on to obtain advanced degrees and select a health occupation.

Because our primary interest in this paper is in racial differences, Table 2 identifies between-group differences in the samples of whites and non-whites at various stages of the educational pipeline. The data are consistent with known differences in demographics and the socioeconomic status between whites and minorities. At the first observable point in the pipeline in the NLS-72, we see that white high school graduates are much more likely to have better-educated parents than non-white high school graduates; 72.2 percent of white fathers have at least a high school diploma compared to only 45.2 percent of non-white fathers. Differences in precollegiate academic ability, proxied by SAT score, are also substantial. White high school graduates are fairly evenly distributed across the four SAT quartiles, while 60 percent of non-white graduates fall in the lowest 20 percent of SAT scores in the sample.[18] Likewise, non-white high school graduates are much less likely (6.0 percent) to score in the top SAT quartile than white high school graduates (28.9 percent). These observed differences in academic preparation are consistent with well-documented test scores gaps between whites and minorities.[19]

Table 2 also indicates that the racial gaps that exist upon high school graduation are still present and, in some cases, exacerbated further in the educational pipeline. The between-group differences in parental educational attainment actually grow more pronounced when we look at college entrants compared to high school graduates, as do differences in the representation in the highest SAT quartile. As we look at these individuals further out into the pipeline, minority representation in the top SAT quartile does not increase as quickly as it does for whites, so the between-group difference continues to grow until the health professional stage.

# 5 Results

## 5.1 Basic Model

The parameter estimates from the basic model are presented in Table 3. Conditional on parental education, SAT score, and urbanicity, blacks are more likely than whites to enroll in college and complete a baccalaureate degree, and these effects are more pronounced for black females than black males (-0.217). It is also the case

[16] Due to substantial missing parental income data in the NLS-72, we use only parental educational attainment.

[17] Not all high school students take the SAT test; some opt for the ACT test or no college entrance exam at all. In addition, the NLS-72 survey respondents took a standardized test with sections on vocabulary, picture numbers (associative memory), reading, letter groups, mathematics, and mosaic comparisons. Using the scaled math scores and scaled reading scores, we employed regression analysis to generate a prediction of the SAT score.

[18] It is possible that these differences are somewhat exaggerated by the fact that blacks are less likely to have observed SAT scores in the data and that the extrapolation procedure may not predict as well at extremes in the distribution.

[19] See Johnson and Neal (1998).

that Hispanic and Asian individuals, conditional on observables, are more likely to enroll than whites, but they are less likely to complete a baccalaureate degree. The result that shifts the unconditional deficit in black college enrollment to greater enrollment probability for blacks conditional on parental background and a student's high school achievement is well-established in the prior empirical work. The seminal work by Manski and Wise (1983) shows that, conditional on observable characteristics, blacks from both the North and the South are substantially more likely to enroll in college than their white counterparts, while blacks from the South are also appreciably more likely to persist in college. Kane (1999) finds a similar advantage in enrollment using data from the NELS for students expected to graduate from high school in 1992.

An individual's SAT score is positively associated with enrolling and completing college, as is having a parent with a college degree. Because our sample respondents were born in approximately 1954, their parents' generation had high school completion rates that were approximately half of what they are today (Goldin, 2003). Thus, it is not particularly surprising that even high school completion by parents increases respondents' college enrollment rates by nearly the same magnitude as college completion by parents. College degree attainment is only positively affected by maternal high school completion for non-blacks; the effect of paternal high school completion for non-blacks is statistically insignificant. Individuals who attended high school in a rural or farming area are less likely to enroll in and more likely to graduate from college than those in non-rural areas. While the role of urbanicity does not appear to have a differential effect on college entrance for black high school graduates, blacks who attended high school in rural areas are much more likely to graduate from college than non-blacks in non-rural areas, conditional on entering college and other observable characteristics.

The effect of covariates on the likelihood of choosing occupations requiring advanced degrees in the health professions is shown in the third panel of Table 3. The probability of following this path increases with an individual's SAT score, while parental education has a mixed effect on a college graduate's decision to become a health professional. The effect of race is strikingly positive, and this effect is more pronounced for black males than black females (0.066), indicating that, conditional on college completion and other attributes, the probability of pursuing a health profession is higher for blacks than for whites. This result, as the discussion in the next section demonstrates, does not persist when the type of college in which an individual enrolls is incorporated as a determinant of degree completion. Growing up in a rural area decreases the likelihood of choosing to become a health professional, and this negative effect is much stronger for black individuals from rural areas (-2.126).

The lower panel of Table 3 displays the estimated covariances between unobservable factors in each of these three stages. Surprisingly, those individuals who are more likely to enter college for unobservable reasons are *less* likely to complete a four-year degree for unobservable reasons.[20] The correlation in unobservables works in

[20] This result also appears when estimating the basic model with data from the National Educational Longitudinal Survey (NELS), which tracks the postsecondary choices of the high school class of 1992. Note that NELS is not suitable for estimating the complete model that includes the decision to enter a health profession with an advanced degree because NELS respondents are not followed through their career and graduate educational choices.

the anticipated direction for the other choices. Unobservables that make it more likely that a person completes college are positively related to those unobservables that encourage a person to become a health professional with an advanced degree.

## 5.2 Model with Variation in College Type and Quality

The parameter estimates from the structural model that includes college quality and historically black institutions are presented in Table 4. Recall that our impetus for adding variation in college attributes is that variation in the types of colleges individuals attend may influence college completion rates, propensity to obtain an advanced degree, or propensity to choose a health care occupation. Many of the qualitative conclusions regarding the determinants of college entry are the same as in the basic model discussed above, but there are some noticeable differences in other stages of decision making.

Conditional on college entry *and* the attributes of the college chosen, as well as parental education, academic ability, and urbanicity, blacks are now even more likely than whites to complete a four-year degree. In the basic model, Hispanics and Asians were conditionally less likely than whites to complete a college degree. In the quality-adjusted model, both groups are conditionally *more* likely to complete. The effect of coming from a rural area on college degree completion changes sign between the basic and quality-adjusted models, indicating that growing up in a rural area and college quality are negatively correlated. In the quality-adjusted model, individuals from rural areas who enter college are less likely to graduate from college, conditional on other factors, although this effect is mitigated for blacks from rural areas.

The parameter estimates in Table 4 also enable us to examine the determinants of choosing an HBCU institution. Black high school graduates are, not surprisingly, more likely than whites to choose (and be chosen by) a historically black college or university, and this effect is somewhat stronger for black females than black males (-0.234). Individual SAT score is negatively associated with choosing an HBCU. We also included a measure of the quality of non-HBCU institution individual $i$ could attend in the HBCU equation in both linear and quadratic form. The parameter estimates on these non-HBCU college quality terms are labeled $\alpha_1$ and $\alpha_2$ in Table 4. Only $\alpha_2$ is identified; thus we set $\alpha_1$ equal to 1 and estimate $\alpha_2$. The negative estimated value of $\alpha_2$ indicates that an individual's propensity to choose an HBCU initially increases, but eventually decreases in the quality of non-HBCU alternatives available. This result indicates that high school graduates with the ability to garner admissions offers from top-tier non-HBCU institutions are less likely to select a historically black institution.

The final column of parameter estimates in Table 4 refers to individuals' propensities to become a health care professional with an advanced degree conditional on all previous choices and outcomes in earlier stages of the educational pipeline. Several interesting results emerge that are distinct from the basic model that did not allow for variation in colleges by quality or HBCU status. First, the parameter estimate on black is now negative, indicating that black college graduates are less likely than observationally equivalent whites to go into the health professions with advanced degrees. This negative effect is stronger for black males than black females (-0.070). Having college-educated parents continues to have mixed implications for the decision to become a health care professional and the effect of coming from a rural area continues to be negative.[21]

The bottom panel of Table 4 presents the covariances between unobservable factors that influence college entry, choice of an HBCU, degree completion, and becoming a health professional. The unobservables that make it more likely that an individual starts college are positively correlated with unobservable determinants of choosing an HBCU, but still negatively associated with the unobservables factors that determine degree completion.[22] There is a negative covariance between unobservables in the decision to enroll at an HBCU and degree completion, but a positive covariance between enrolling at an HBCU and becoming a health professional. Once we control for attributes of the undergraduate institution selected, the covariance between unobservable determinants of degree completion and unobservable determinants of becoming a health professional is negative, though not statistically significant.

### 5.2.1 Marginal Effects of Individual Characteristics on Choice Probabilities

To understand how the parameter estimates from the quality-adjusted model in Table 4 affect the probabilities of entering college, enrolling in a college with certain characteristics, completing college, and becoming a health care professional with an advanced degree, we calculate marginal effects of each of the covariates. The marginal effects presented in Table 5 are conditional on successfully completing all previous stages in the educational pipeline as well as on other observable characteristics. For each coefficient, we calculate the marginal effect using the means of the data at the relevant decision points.

A primary question in this analysis is how race affects the probability of different outcomes in the collegiate pipeline. We present the estimated effects in Table 5 relative to outcomes predicted for white females. For example, the second row of marginal effects, labeled "Black", indicates how the probability of each outcome would be expected to differ for a black female relative to a white female, evaluated at the means of the other covariates. Relative to a white woman with the same characteristics, a black woman is appreciably more likely to enroll in college (27.97 percentage points), to attend a selective four-year college or an HBCU, and to complete an undergraduate degree (25.17 percentage points). Yet, conditional on college enrollment, college type, and degree completion, there is a decline in progress to the health professions of 2.55 percentage points for black females relative to white females. Given that the overall share of college graduates who become health professionals is

[21] The omission of various interaction terms from the health care professional equation is due to small sample sizes among black health care professionals.

[22] *A priori*, we believed that this covariance would be positive once we controlled for variation in the attributes of the college chosen. In fact, when we estimate the quality-adjusted model with NELS data, the covariance between unobservables in college entry and completion is positive.

about 5 percent, this is a sizeable effect. Similar statements can be made about going from a white male to a black male by combining the marginal effects in the second and third rows of Table 5. For the Hispanic and Asian group membership, there is a positive marginal effect on college entry and undergraduate degree receipt, while membership in these groups is not linked to the health professional outcome in a statistically significant way.

The marginal effect of a 100-point increase on individual SAT score is, not surprisingly, associated with a higher probability of college entry, college selectivity, and degree completion. The probability of college enrollment increases by 9.68 percentage points when SAT increases by 100 points. Conditional on college entry and observables, the probability of attending a more selective four-year institution increases by approximately 3 percentage points at moderately- to most-selective colleges with a 100-point SAT score increase. Higher SAT scores are also associated with an increased probability (3.13 percentage points) of becoming a health professional with an advanced degree.

Table 5 also indicates the effect of parental education on children's educational and career outcomes. Among non-blacks, having a father with a high school education is associated with a 7.63 percentage point increase in college entry, higher probabilities of attending a more selective institution (conditional on college entry), a 3.68 percentage point increase in the probability of degree receipt (conditional on college entry and college type), and a 2.25 percentage point increase in the probability of becoming a health professional (conditional on college entry and type and degree receipt). Mothers of non-blacks with high school diplomas have a similar impact on their children's probability of progressing through the educational pipeline. Non-black parents who also complete a baccalaureate degree also increase their children's probability of progressing through the pipeline. Non-black college-educated fathers (relative to high school educated fathers) increase their child's probability of college entry by 7.61 percentage points, of going to a more selective college by 2.3 to 2.8 percentage points, of degree completion by 6.61 percentage points, and of becoming a health professional by 1.46 percentage points. Non-black college-educated mothers have similar marginal effects.

The marginal effects of parental educational attainment differ substantially by race. For blacks, the marginal effects of their father's high school degree completion on college entry, completion, and health professional are 10.98, 7.35, and 2.27 percentage points. Maternal high school completion has mixed effects for blacks. While a black individual with a father who also completes a college degree has a 6.95 higher probability of entering college, paternal college completion actually substantially reduces the child's probability of completing a college degree (7.36 percentage point decline). There is no difference between blacks and non-blacks in the marginal effect of paternal college completion on the probability of becoming a health care professional with an advanced degree.

Finally, moving from a rural to a non-rural location has no statistically significant effect on the college entry and degree completion probabilities of either blacks or non-blacks, but does have a small positive and statistically significant effect on the probability of becoming a health professional for both groups. Originating from a rural area also has no discernible effect on the probability of going to a more selective four-year institution for either non-blacks or blacks. It is interesting to note that the parameter estimate on the effect of being from a rural area in the health professional stage is negative in Table 4 (-0.309) and positive (0.0049) in Table 5. This result stems from selection and correlation in the unobservable determinants of the decisions to enter college, complete a degree, and become a health professional, thereby demonstrating the importance of jointly modeling these decisions in the way that we do. Although high school graduates from rural areas are *less* likely to become health professionals with advanced degrees than those from non-rural areas, conditioning on college entry, college selectivity, graduation, and unobservables indicates that college graduates from rural areas are slightly *more* likely to become health professionals than their non-rural peers.

## 6 Specification Tests

The quality-adjusted model presented at the end of Section 3 specifies the probabilities of observing a variety of different educational and career outcomes. Because we model the decision to enter college, the type of college chosen (HBCU or non-HBCU in one of five selectivity categories or selectivity unknown or two-year college), degree completion, and choosing a health profession that requires an advanced degree, there are 22 different educational/career paths available to each individual.[23] We use the parameter estimates in Table 4 to compute predicted probabilities that individuals choose each educational/career path and compare the predicted behavior with actual outcomes. Table 6 presents predicted and actual proportions of individuals choosing each educational/career path. Although predicted behavior appears to be very similar to actual behavior in many cases, we also divide the sample into quintiles based on predicted probabilities to facilitate the construction of more formal specification test statistics.

We perform $\chi^2$ goodness-of-fit tests to more rigorously examine how well the model fits the choices and outcomes that we actually observe in the NLS-72 data. The null hypothesis for this statistical test is that the proportions predicted by the model equal the actual proportions in the data, thus, test statistics that fall below the critical value indicate that the model fits the data well. $\chi^2$ goodness-of-fit statistics for each outcome, by quintile and overall, are presented in Table 7.[24]

[23] The seven probabilities listed in Section 3 have nested within them the choice of college type, which expands the total number of choices from seven to 22. For example, the educational/career paths available to individuals include: (1) Do not enter college, (2) Enter an HBCU, but do not complete a degree, (3) Enter a non-HBCU of level 5 selectivity, but do not complete a degree, (4) Enter a non-HBCU of level 5 selectivity, complete a degree, but do not become a health professional with an advanced degree, and so on.

[24] Test statistics are reported for 16 out of the 22 educational/career paths due to insufficient variation in choice probabilities for six of the possible outcomes. The six paths omitted for this reason include paths that involve becoming a health professional with an advanced degree if the undergraduate college was an HBCU, a two-year institution, or a four-year non-HBCU of selectivity level 2 (the lowest selectivity for non-special four-year institutions), and any path that involves choosing a “special” four-year instituions

Overall, the model fails this specification test. However, a closer examination of the disaggregation by outcome and quintile reveals that the model does a poor job primarily in those outcomes that involve college entrance with no degree completion, particularly at lower quality institutions. This result stems from the fact that the degree completion equation in the model, $y^*_{3i}$, does not take into account how well students are matched with the postsecondary institution in which they are enrolled. For example, assume that an individual enrolls in a level 2 college (lowest selectivity category) for unobservable reasons (*i.e.,* a small $u_{1i}$ in the $y^*_{1i}$ equation). If this individual compares favorably to his college peers, this has a positive effect on his propensity to complete a baccalaureate degree ($y^*_{3i}$). In our model, $y^*_{3i}$ is not a function of the difference between $y^*_{1i}$ and the average $y^*_1$ among student $i$'s peers, which causes us to understate degree completion for some college entrants. Despite this shortcoming of the model, the lower half of Table 7 indicates that the model does a good job of predicting those educational/career outcomes that involve moderately and more selective undergraduate institutions.

# 7 Projecting the Distribution of Black Health Care Professionals

## 7.1 Predicting Representation Using NELS Data

The estimates presented in the Section 5 are for the single cohort of students graduating from high school in 1972. This leaves open the important question of the extent to which behavioral parameters or the distribution of explanatory variables have changed in the most recent decades in ways that should affect the entry of blacks into the health care professions. Students from the high school class of 1972 would have begun to graduate from college in 1976 with entry into the health professions with an advanced degree beginning in the early 1980s. The National Educational Longitudinal Survey (NELS) is a similarly structured survey of the high school class of 1992. We use these data to examine the extent to which there have been large changes in behavior (affecting the estimated coefficients) or changes in the distribution of population characteristics over this interval predicting changes in the representation of blacks in the health professions.[25]

Given the well-known increases in parental education for students from the 1992 cohort relative to the 1972 cohort, particularly for black students, it is no surprise that college enrollment and college completion are expected to increase over this interval. The first two sets of bars in Figures 5 illustrate this prediction with college enrollment rates rising and completion rates conditional on enrollment also increasing. To illustrate, evaluating the NELS data using the estimates from NLS-72 leads to a predicted rise in enrollment for whites from 47 percent to 53 percent and a rise in enrollment for blacks from 43 percent to 55 percent. In turn, college completion conditional on enrollment is predicted to rise somewhat for whites from 54 percent to 58.7 percent while increasing more markedly for blacks. Moving to the final set of bars which show the actual and predicted representation among

(level 1 non-HBCU).

[25] See Appendix A.3 for a discussion of the NELS data.

advanced degree health professionals conditional on college graduation, we see that there are few changes in predicted health professional shares for either whites or blacks.

### 7.2 Understanding the Stagnation of the Representation of Black Health Care Professionals

Changing values of explanatory variables and outcomes at intermediate stages in the educational attainment process have not been sufficiently large to generate sustained changes in the predicted representation of blacks in the health care professions. In addition to the projections that we have made using micro data, we turn to data from large national data sets, including Census files, to document the stagnation of the representation of blacks in the health care professions. Focusing only on the ratio of black to white health care professionals, changes over the last two decades in the representation of blacks in the health professions have been modest and consistent with our projections using the NLS-72 and NELS micro data. Using data from the 2000 decennial Census, we present the number of black and white health care professionals with advanced degrees by age in Figure 6. What is clear from this presentation is that there is only the most modest upward trend in the black/white ratio among younger workers which is driven by an erosion in the number of whites choosing the health professions rather than a sustained increase in blacks choosing health care professions.

While we emphasize that much of the overall gap in the representation of blacks can be traced to outcomes at the precollegiate and collegiate levels, the question of why we have not observed greater increases in the representation of blacks in health care professions remains primary. The value to entering the health care professions is necessarily relative to other outside options. One hypothesis for the failure to achieve greater gains in post-baccalaureate health care programs is that outside options for black college graduates improved far more rapidly than opportunities in the health professions. As such, demand from professions like law and business, where the gap in wages between black and white professions narrowed rapidly in the 1970s and 1980s, drew many high achieving blacks to MBA programs and law schools. To illustrate, the number of blacks enrolled in law school increased from 3,744 in 1971-72 to 9,529 in 2006-2007, representing an increase of more than 250 percent (American Bar Association, 2007). That demand for advanced study in the health professions has not increased markedly among blacks is borne out in data showing major undergraduate fields of study in 1977 and 1997 by race (see Appendix A.4). If life sciences study at the undergraduate level is an indication of future advanced study in the health professions, black participation in these fields has fallen off over the last two decades at a rate somewhat greater than that observed for whites.

A final point to emphasize in understanding the difficulty associated with increasing the representation of blacks in the health professions is that admission opportunities have not changed appreciably over the last three decades in medical schools, though there has been some increased capacity in nursing programs. As shown in Figure 7, while the total number of advanced degrees awarded in nursing fields has increased, MD degrees have remained relatively flat at a bit more than 15,000 per year over the course of the last quarter century. Thus, if the supply of opportunities in medical school is fixed, increasing the probability of the admission of a black student to medical school requires an improvement in the achievement of black students relative to students from other groups. We note that our model does not incorporate this important supply-side dimension in medical education, though we speculate that increased competition for opportunities in medical school will work against gains for representation among black students.

# 8 Discussion and Conclusion

The dramatic underrepresentation of blacks in the health professions is a cause for policy concern because it may capture group differences in educational achievement and opportunities as well as potentially affecting the quality of health provision in the United States. For the cohort that we follow that graduated from high school in 1972, the representation of blacks declined from 11 percent at the point of high school graduation, to 9 percent at college entry, to 7.2 percent at college graduation, and to 4.1 percent at the stage of entry to the health professions. Much of this erosion in the representation of blacks through the pipeline is accounted for by circumstances and choices prior to the entry into post-baccalaureate study in a health field. Still, the estimation of a model in which we control for type of college, measured by both institutional selectivity and status as an HBCU, generates substantial underrepresentation of blacks in the transition from baccalaureate degree receipt to participation in a health profession requiring an advanced degree.

There is little evidence to suggest that changes over the last three decades in student achievement or parental circumstances have been sufficiently large in absolute terms and relative to other groups to predict substantial changes in the representation of blacks among those with advanced degrees in the health professions. The underrepresentation of blacks in the health professions is part of the more general social and economic problems generating substantial group differences, entrenched before the college years.

Our evidence suggests that further efforts to understand the pathway from undergraduate degree receipt to entry in advanced degree health programs by race and type of undergraduate experience may be a constructive direction for future research. Still, we caution that, even with a compelling public policy interest to increase the representation of blacks in the health professions, efforts to target students at the margin between college completion and entry to a graduate program in the health professions may well generate substantial distortions in the educational marketplace in the absence of a full understanding of the causes of race-specific differences in the collegiate pipeline.

## References


[1] Association of American of Medical Colleges. 2005. *AAMC Data Book*. Washington, D.C.

[2] America Bar Association. 2007. Downloaded from: http://www.abanet.org/legaled/statistics/charts/stats%20-%2013.pdf.

[3] Bach, P.B., H.H. Pham, D. Schrag, R.C. Tate, and J.L. Hargraves. 2004. "Primary Care Physicians Who Treat Blacks and Whites." *New England Journal of Medicine.* 351(6): 575-84.

[4] Barron's Educational Series. 1994. *Barron's Profiles of American Colleges*. 20th ed. Hauppage, NY: Barron's Educational Series.

[5] Bowen, William and Derek Bok. 1998. *The Shape of the River: Long Term Consequences of Considering Race in College and University Admissions*. Princeton: Princeton University Press.

[6] Cooper, L.A., D.L. Roter, R.L. Johnson, D.E. Ford, D.M. Steinwachs, and N.R. Powe. 2003. "Patient-Centered Communication, Ratings of Care, and Concordance of Patient and Physician Race." *American College of Physicians*, 139: 907-915.

[7] DeLong, J. Bradford, Claudia Goldin and Lawrence Katz. 2003. "Sustaining U.S. Economic Growth." in H. Aaron, J. Lindsay, and P. Nivola, eds., *Agenda for the Nation*. Washington, D.C.: Brookings Institution Press.

[8] Flaskerud, J.H. and P.Y. Liu. 1990. "Influence of Therapist Ethnicity and Language on Therapy Outcomes of Southeast Asian Clients." *International Journal of Social Psychiatry.* 36(1): 18-29.

[9] Freeman, Richard. 1976. *The Black Elite: The New Market for Highly Educated Black Americans*. San Francisco: Carnegie Commission on Higher Education.

[10] Giron, M., P. Manjon-Arce, J. Puerto-Barber, E. Sanchez-Garcia, and M. Gomez-Beneyto. 1998. "Clinical Interview Skills and Identification of Emotional Disorders in Primary Care." *American Journal of Psychiatry.* 155: 530-5.

[11] Goldin, Claudia. 2003. "The Human Capital Century." *Education Next*. Winter: 73-78.

[12] Jenkins, M. 1947. "The Availability of Higher Education for Negroes in Southern States." *Journal of Negro Education.* 16(3): 459-473.

[13] Johnson, William R. and Derek Neal. 1998. "Black-White Test Score Differences and Black-White Earnings Differences." in Christopher Jencks and Meredith Phillips, eds. *The Black-White Test Score Gap*. Brookings Institution Press.

[14] Kane, Thomas. 1994. "College Entry by Blacks since 1970: The role of College Costs, Family Background and Returns to Education." *Journal of Political Economy*. 102: 878-911.

[15] Kane, Thomas. 1999. *The Price of Admission: Rethinking How Americans Pay for College*. Washington, D.C.: Brookings Institution Press.

[16] Kaplan S.H., S. Greenfield, and J.E. Ware, Jr. 1989. "Assessing the Effects of Physician-Patient Interactions on the Outcomes of Chronic Disease." *Medical Care*. 27: S110-27.

[17] Karen, David. 1991. "The Politics of Class, Race and Gender: Access to Higher Education in the United States, 1960-1986." *American Journal of Education.* 99(2): 208-37.

[18] Komaromy, M., K. Grumbach, M. Drake, et. al. 1996. "The Role of Black and Hispanic Physicians in Providing Health Care for Underserved Populations." *New England Journal of Medicine.* 334(20): 1305-10.

[19] Krueger, Alan, Jesse Rothstein, and Sarah E. Turner. forthcoming. "Race, Income and College in 25 Years: Evaluating Justice O'Connor's Conjecture." *American Law and Economics Review*.

[20] Lakhan, Shaheen Emmanuel. 2003. "Diversification of U.S. Medical Schools Via Affirmative Action Implementation." *BMC Medical Education*. 3:6 doi:10.1186/1472-6920-3-6.

[21] Lavy, Victor, Michael Palumbo, and Steven N. Stern. 1998. "Simulation of Multinomial Probit Probabilities and Imputation." in Thomas Fomby and R. Carter Hill, eds. *Advances in Econometrics*. JAI Press.

[22] Manski, Charles F., and David A. Wise. 1983. *College Choice in America*. Cambridge, Mass.: Harvard University Press.

[23] National Center for Educational Statistics, U.S. Department of Education. 1994. *National Longitudinal Study of the High School Class of 1972*. CD-ROM.

[24] Penn N.E., P.J. Russell, H.J. Simon, et. al. 1986. "Affirmative Action at Work: A Survey of Graduates of the University of California, San Diego, Medical School." *American Journal of Public Health.* 76(9): 1144-6.

[25] Rosenheck R., A. Fontana, and C. Cottrol. 1995. "Effect of Clinician-Veteran Racial Pairing in the Treatment of Posttraumatic Stress Disorder." *American Journal of Psychiatry.* 152(4): 555-63.

[26] Saha, S., M. Komaromy, T.D. Koepsell, and A.B. Bindman. 1999. "Patient-Physician Racial Concordance and the Perceived Quality and Use of Health Care." *Archives of Internal Medicine.* 159(9): 997-1004.

[27] Saha, S., S.H. Taggart, M. Komaromy, and A.B. Bindman. 2000. "Do Patients Choose Physicians of Their Own Race?" *Health Affairs.* 19(4): 76-83.

[28] Stern, Steven N., Elizabeth Merwin, Emily Hauenstein, Ivora Hinton, Virginia Rovnyak, Melvin Wilson, Ishan Williams, Irma Mahone. 2007. "The Effect of Rurality on Mental and Physical Health." Unpublished manuscript.

[29] M.A. Stewart. 1995. "Effective Physician-Patient Communication and Health Outcomes: A Review." *Canadian Medical Association Journal*. 152: 1423-33.

[30] U.S. Department of Health and Human Services. 2001. "Mental Health: Culture, Race, and Ethnicity - A Supplement to Mental Health: A Report of the Surgeon General." Substance Abuse and Mental Health Services Administration, Center for Mental Health Services.

[31] U.S. Department of Health and Human Services. 2006. "The Rationale for Diversity in the Health Professions: A Review of the Evidence." Health Resources and Services Administration, Bureau of Health Professions, ftp://ftp.hrsa.gov/bhpr/workforce/diversity.pdf.

[32] Ware, J.E. Jr. and A.R. Davies. 1983. "Behavioral Consequences of Consumer Dissatisfaction with Medical Care." *Evaluation and Program Planning.* 6: 291-7.

# A Appendix

Define $\Phi(\cdot)$ as the standard normal distribution function, $\phi(\cdot)$ as the standard normal density function, $B(\cdot,\cdot;\rho)$ as the standard bivariate normal distribution function (with correlation $\rho$), $b(\cdot,\cdot;\rho)$ as the standard bivariate normal density function (with correlation $\rho$), $t(\cdot,\cdot,\cdot;\Omega)$ as the standard trivariate normal density with covariance matrix $\Omega$,[26] and

$$\begin{aligned}\rho_{jk|m} &= Corr\left(u_{ji}, u_{ki} \mid u_{mi}\right) \\ &= \frac{\Omega_{jk} - \Omega_{jm}\Omega_{km}}{\sqrt{\left(1-\Omega_{jm}^2\right)\left(1-\Omega_{km}^2\right)}}.\end{aligned}$$

## A.1 Choice Probabilities in the Basic Model

Recall from Section 3 that there are four choice probabilities in the basic model. The conditional probability of not going to college is[27]

$$P_1\left(X_i;\theta\right) = \Pr\left[y_{1i}=0 \mid X_{1i}\right] = \Phi\left(-X_{1i}\beta_1\right); \tag{4}$$

the conditional probability of going to college but not finishing is

$$\begin{aligned}P_2\left(X_i;\theta\right) &= \Pr\left[y_{1i}=1, y_{2i}=0 \mid X_{1i}, X_{2i}\right] \\ &= \int_{-X_{1i}\beta_1}^{\infty}\int_{-\infty}^{-X_{2i}\beta_2} b\left(u_{1i}, u_{2i}; \Omega_{12}\right) du_{2i} du_{1i} \\ &= B\left(X_{1i}\beta_1, -X_{2i}\beta_2; -\Omega_{12}\right);\end{aligned} \tag{5}$$

the conditional probability of finishing college but not becoming a health professional with an advanced degree is

$$\begin{aligned}&P_3\left(X_i;\theta\right) = \Pr\left[y_{1i}=1, y_{2i}=1, y_{3i}=0 \mid X_{1i}, X_{2i}, X_{3i}\right] \\ = &\int_{-X_{1i}\beta_1}^{\infty}\int_{-X_{2i}\beta_2}^{\infty}\int_{-\infty}^{-X_{3i}\beta_3} t\left(u_{1i}, u_{2i}, u_{3i}; \Omega\right) du_{3i} du_{2i} du_{1i} \\ = &\int_{-X_{1i}\beta_1}^{\infty} B\left(\frac{X_{2i}\beta_2 + \Omega_{12}u_{1i}}{\sqrt{1-\Omega_{12}^2}}, \frac{-X_{3i}\beta_3 - \Omega_{13}u_{1i}}{\sqrt{1-\Omega_{13}^2}}; -\rho_{23|1}\right)\phi\left(u_{1i}\right) du_{1i};\end{aligned} \tag{6}$$

and the conditional probability of becoming a health professional with an advanced degree is

$$\begin{aligned}&P_4\left(X_i;\theta\right) = \Pr\left[y_{1i}=1, y_{2i}=1, y_{3i}=1 \mid X_{1i}, X_{2i}, X_{3i}\right] \\ = &\int_{-X_{1i}\beta_1}^{\infty}\int_{-X_{2i}\beta_2}^{\infty}\int_{-X_{3i}\beta_3}^{\infty} t\left(u_{1i}, u_{2i}, u_{3i}; \Omega\right) du_{3i} du_{2i} du_{1i} \\ = &\int_{-X_{1i}\beta_1}^{\infty} B\left(\frac{X_{2i}\beta_2 + \Omega_{12}u_{1i}}{\sqrt{1-\Omega_{12}^2}}, \frac{X_{3i}\beta_3 + \Omega_{13}u_{1i}}{\sqrt{1-\Omega_{13}^2}}; \rho_{23|1}\right)\phi\left(u_{1i}\right) du_{1i}.\end{aligned} \tag{7}$$

[26] Note that the standard trivariate normal density function has a covariance matrix with diagonal elements of $\Omega$ equal to 1.
[27] Note that an implication of equation (1) is that $\Pr\left[y_{1i}=0 \mid X_i\right] = \Pr\left[y_{1i}=0 \mid X_{1i}\right]$. Similar statements can be made about equations (5) through (7) using equations (1) through (3).

## A.2 Choice Probabilities in Model with College Quality and HBCUs

Recall from Section 3 that there are seven choice probabilities in the model that allows for variation in college characteristics. The conditional probability of not going to college is

$$\begin{aligned} P_1(X_{1i}, X_{2i}; \theta) &= \Pr[y_{1i} = 0, y_{21} = 0 \mid X_{1i}, X_{2i}] \\ &= \Pr[y_{1i}^* < 0, y_{21}^* = 0 \mid X_{1i}, X_{2i}] \\ &= \Pr\left[u_{1i} < -X_{1i}\beta_1, u_{2i} < -\alpha_1 y_{1i}^* - \alpha_2 (y_{1i}^*)^2 - X_{2i}\beta_2\right] \\ &= \int_{-\infty}^{-X_{1i}\beta_1} \int_{-\infty}^{-\widetilde{y}_{2i}^*(u_{1i})} f(u_{1i}, u_{2i})\, du_{2i} du_{1i} \end{aligned} \tag{8}$$

where

$$\widetilde{y}_{2i}^*(u_{1i}) = \alpha_1 y_{1i}^*(u_{1i}) + \alpha_2 (y_{1i}^*(u_{1i}))^2 + X_{2i}\beta_2;$$

the conditional probability of going to a non-HBCU college of type $k$ but not finishing is

$$\begin{aligned} & P_{2k}(X_{1i}, X_{2i}, X_{3i}; \theta) \\ =\; & \Pr[y_{1i} = k, y_{2i} = 0, y_{3i} = 0 \mid X_{1i}, X_{2i}, X_{3i}] \\ =\; & \Pr[\tau_k < y_{1i}^* \leq \tau_{k+1}, y_{2i}^* < 0, y_{3i} = 0 \mid X_{1i}, X_{2i}, X_{3i}] \\ =\; & \int_{\tau_k - X_{1i}\beta_1}^{\tau_{k+1} - X_{1i}\beta_1} \int_{-\infty}^{-\widetilde{y}_{2i}^*(u_{1i})} \int_{-\infty}^{-X_{3i}\beta_3} f(u_{1i}, u_{2i}, u_{3i})\, du_{3i} du_{2i} du_{1i} \\ =\; & \int_{\tau_k - X_{1i}\beta_1}^{\tau_{k+1} - X_{1i}\beta_1} B\left(-\frac{\widetilde{y}_{2i}^*(u_{1i}) + \Omega_{12}u_{1i}}{\sqrt{1-\Omega_{12}^2}}, -\frac{X_{3i}\beta_3 + \Omega_{13}u_{1i}}{\sqrt{1-\Omega_{13}^2}}; \rho_{23|1}\right) \phi(u_{1i})\, du_{1i}; \end{aligned} \tag{9}$$

the conditional probability of going to an HBCU institution but not finishing is

$$\begin{aligned} & P_3(X_{1i}, X_{2i}, X_{3i}; \theta) \\ =\; & \Pr[y_{2i} = 1, y_{3i} = 0 \mid X_{1i}, X_{2i}, X_{3i}] \\ =\; & \int_{-\infty}^{\infty} \int_{-\widetilde{y}_{2i}^*(u_{1i})}^{\infty} \int_{-\infty}^{-X_{3i}\beta_3} f(u_{1i}, u_{2i}, u_{3i}) \prod_{j=1}^{3} du_{ji} \\ =\; & \int_{-\infty}^{\infty} B\left(\frac{\widetilde{y}_{2i}^*(u_{1i}) + \Omega_{12}u_{1i}}{\sqrt{1-\Omega_{12}^2}}, -\frac{X_{3i}\beta_3 + \Omega_{13}u_{1i}}{\sqrt{1-\Omega_{13}^2}}; -\rho_{23|1}\right) \phi(u_{1i})\, du_{1i}; \end{aligned} \tag{10}$$

the conditional probability of going to a non-HBCU college of type $k$, finishing, but not becoming a health professional with an advanced degree is

$$\begin{aligned} & P_{4k}(X_{1i}, X_{2i}, X_{3i}, X_{4i}; \theta) \\ =\; & \Pr[y_{1i} = k, y_{2i} = 0, y_{3i} = 1, y_{4i} = 0 \mid X_{1i}, X_{2i}, X_{3i}, X_{4i}] \\ =\; & \int_{\tau_k - X_{1i}\beta_1}^{\tau_{k+1} - X_{1i}\beta_1} \int_{-\infty}^{-\widetilde{y}_{2i}^*(u_{1i})} \int_{-X_{3i}\beta_3}^{\infty} \int_{-\infty}^{-X_{4i}\beta_4} f(u_i) \prod_{j=1}^{4} du_{ji} \\ =\; & \int_{\tau_k - X_{1i}\beta_1}^{\tau_{k+1} - X_{1i}\beta_1} \int_{-\infty}^{-\widetilde{y}_{2i}^*(u_{1i})} B\left(x_{3i}, -x_{4i}; -\rho_{34|12}\right) b(u_{1i}, u_{2i}) \prod_{j=1}^{2} du_{ji}; \end{aligned} \tag{11}$$

the conditional probability of going to an HBCU institution, finishing, but not becoming a health professional with an advanced degree is

$$\begin{aligned} & P_5\left(X_{1i}, X_{2i}, X_{3i}, X_{4i}; \theta\right) \quad (12) \\ = & \Pr\left[y_{2i}=1, y_{3i}=1, y_{4i}=0 \mid X_{1i}, X_{2i}, X_{3i}, X_{4i}\right] \\ = & \int_{-\infty}^{\infty} \int_{-\widetilde{y}_{2i}^{*}(u_{1i})}^{\infty} \int_{-X_{3i}\beta_3}^{\infty} \int_{-\infty}^{-X_{4i}\beta_4} f\left(u_i\right) \prod_{j=1}^{4} du_{ji} \\ = & \int_{-\infty}^{\infty} \int_{-\widetilde{y}_{2i}^{*}(u_{1i})}^{0} B\left(x_{3i}, -x_{4i}; -\rho_{34|12}\right) b\left(u_{1i}, u_{2i}\right) \prod_{j=1}^{2} du_{ji}; \end{aligned}$$

the conditional probability of going to a non-HBCU college of type $k$, finishing, and becoming a health professional with an advanced degree is

$$\begin{aligned} & P_{6k}\left(X_{1i}, X_{2i}, X_{3i}, X_{4i}; \theta\right) \quad (13) \\ = & \Pr\left[y_{1i}=k, y_{2i}=0, y_{3i}=1, y_{4i}=1 \mid X_{1i}, X_{2i}, X_{3i}, X_{4i}\right] \\ = & \int_{\tau_k - X_{1i}\beta_1}^{\tau_{k+1} - X_{1i}\beta_1} \int_{-\infty}^{-\widetilde{y}_{2i}^{*}(u_{1i})} \int_{-X_{3i}\beta_3}^{\infty} \int_{-X_{4i}\beta_4}^{\infty} f\left(u_i\right) \prod_{j=1}^{4} du_{ji} \\ = & \int_{\tau_k - X_{1i}\beta_1}^{\tau_{k+1} - X_{1i}\beta_1} \int_{-\infty}^{-\widetilde{y}_{2i}^{*}(u_{1i})} B\left(-x_{3i}, x_{4i}; -\rho_{34|12}\right) b\left(u_{1i}, u_{2i}\right) \prod_{j=1}^{2} du_{ji}; \end{aligned}$$

the conditional probability of going to an HBCU institution, finishing, and becoming a health professional with an advanced degree is

$$\begin{aligned} & P_7\left(X_{1i}, X_{2i}, X_{3i}, X_{4i}; \theta\right) \quad (14) \\ = & \Pr\left[y_{2i}=1, y_{3i}=1, y_{4i}=1 \mid X_{1i}, X_{2i}, X_{3i}, X_{4i}\right] \\ = & \int_{-\infty}^{\infty} \int_{-\widetilde{y}_{2i}^{*}(u_{1i})}^{\infty} \int_{-X_{3i}\beta_3}^{\infty} \int_{-X_{4i}\beta_4}^{\infty} f\left(u_i\right) \prod_{j=1}^{4} du_{ji} \\ = & \int_{-\infty}^{\infty} \int_{-\infty}^{-\widetilde{y}_{2i}^{*}(u_{1i})} B\left(x_{3i}, x_{4i}; \rho_{34|12}\right) b\left(u_{1i}, u_{2i}\right) \prod_{j=1}^{2} du_{ji} \end{aligned}$$

where

$$\begin{aligned} x_{3i} &= \frac{X_{3i}\beta_3 + \sum_{j=1}^{2} \Omega_{j3} u_{ji}}{\sqrt{\Omega_{33|12}^2}}; \\ x_{4i} &= \frac{X_{4i}\beta_4 + \sum_{j=1}^{2} \Omega_{j4} u_{ji}}{\sqrt{\Omega_{44|12}^2}}. \end{aligned}$$

There are some observations where we observe the individual enrolling in a four year non-HBCU, but are not able to observe the quality of the institution. The relevant likelihood contributions change from equation (9) to

$$
\begin{aligned}
& P^*_{2k}\left(X_{1i}, X_{2i}, X_{3i}; \theta\right) && (15)\\
= \;& \Pr\left[y_{1i} \geq 2, y_{2i} = 0, y_{3i} = 0 \mid X_{1i}, X_{2i}, X_{3i}\right] \\
= \;& \Pr\left[\tau_2 < y^*_{1i}, y^*_{2i} < 0, y_{3i} = 0 \mid X_{1i}, X_{2i}, X_{3i}\right] \\
= \;& \int_{\tau_2 - X_{1i}\beta_1}^{\infty} \int_{-\infty}^{-\widetilde{y}^*_{2i}(u_{1i})} \int_{-\infty}^{-X_{3i}\beta_3} f\left(u_{1i}, u_{2i}, u_{3i}\right) du_{3i} du_{2i} du_{1i} \\
= \;& \int_{\tau_2 - X_{1i}\beta_1}^{\infty} B\left(-\frac{\widetilde{y}^*_{2i}\left(u_{1i}\right) + \Omega_{12} u_{1i}}{\sqrt{1-\Omega_{12}^2}}, -\frac{X_{3i}\beta_3 + \Omega_{13} u_{1i}}{\sqrt{1-\Omega_{13}^2}}; \rho_{23|1}\right) \phi\left(u_{1i}\right) du_{1i};
\end{aligned}
$$

from equation (11) to

$$
\begin{aligned}
& P^*_{4k}\left(X_{1i}, X_{2i}, X_{3i}, X_{4i}; \theta\right) && (16)\\
= \;& \Pr\left[y_{1i} \geq 2, y_{2i} = 0, y_{3i} = 1, y_{4i} = 0 \mid X_{1i}, X_{2i}, X_{3i}, X_{4i}\right] \\
= \;& \int_{\tau_2 - X_{1i}\beta_1}^{\infty} \int_{-\infty}^{-\widetilde{y}^*_{2i}(u_{1i})} \int_{-X_{3i}\beta_3}^{\infty} \int_{-\infty}^{-X_{4i}\beta_4} f\left(u_i\right) \prod_{j=1}^{4} du_{ji} \\
= \;& \int_{\tau_2 - X_{1i}\beta_1}^{\infty} \int_{-\infty}^{-\widetilde{y}^*_{2i}(u_{1i})} B\left(x_{3i}, -x_{4i}; -\rho_{34|12}\right) b\left(u_{1i}, u_{2i}\right) \prod_{j=1}^{2} du_{ji};
\end{aligned}
$$

and from equation (13) to

$$
\begin{aligned}
& P^*_{6k}\left(X_{1i}, X_{2i}, X_{3i}, X_{4i}; \theta\right) && (17)\\
= \;& \Pr\left[y_{1i} \geq 2, y_{2i} = 0, y_{3i} = 1, y_{4i} = 1 \mid X_{1i}, X_{2i}, X_{3i}, X_{4i}\right] \\
= \;& \int_{\tau_2 - X_{1i}\beta_1}^{\infty} \int_{-\infty}^{-\widetilde{y}^*_{2i}(u_{1i})} \int_{-X_{3i}\beta_3}^{\infty} \int_{-X_{4i}\beta_4}^{\infty} f\left(u_i\right) \prod_{j=1}^{4} du_{ji} \\
= \;& \int_{\tau_2 - X_{1i}\beta_1}^{\infty} \int_{-\infty}^{-\widetilde{y}^*_{2i}(u_{1i})} B\left(-x_{3i}, x_{4i}; -\rho_{34|12}\right) b\left(u_{1i}, u_{2i}\right) \prod_{j=1}^{2} du_{ji}.
\end{aligned}
$$

Note that $k = 1$ corresponds to enrolling in a two year college and so is not consistent with such an observation.

Equations (8) through (17) are the probabilities for the ten possible events that can occur in the data. The log likelihood contribution for $i$ when the quality of non-HBCU institutions is observed is

$$
\begin{aligned}
L_i(\theta) = {} & 1(y_{1i}=0)(1-y_{2i})\log P_1(X_{1i},X_{2i};\theta) \\
& +(1-y_{2i})(1-y_{3i})\sum_{k=1}^{K} 1(y_{1i}=k)\log P_2(X_{1i},X_{2i},X_{3i};\theta) \\
& +y_{2i}(1-y_{3i})\log P_3(X_{1i},X_{2i},X_{3i};\theta) \\
& +(1-y_{2i})y_{3i}(1-y_{4i})\sum_{k=1}^{K} 1(y_{1i}=k)\log P_4(X_{1i},X_{2i},X_{3i},X_{4i};\theta) \\
& +y_{2i}y_{3i}(1-y_{4i})\log P_5(X_{1i},X_{2i},X_{3i},X_{4i};\theta) \\
& +(1-y_{2i})y_{3i}y_{4i}\sum_{k=1}^{K} 1(y_{1i}=k)\log P_6(X_{1i},X_{2i},X_{3i},X_{4i};\theta) \\
& +y_{2i}y_{3i}y_{4i}\log P_7(X_{1i},X_{2i},X_{3i},X_{4i};\theta),
\end{aligned}
$$

and the adjustments required when the quality of non-HBCU institutions are not observed involves changing the appropriate term to its replacement. As in the basic model, we maximize $\sum_i L_i(\theta)$ over $\theta$ to get consistent, asymptotically normal estimates of $\theta$.

## A.3 The National Educational Longitudinal Study (NELS)

The National Education Longitudinal Study of 1988 (NELS) consists of a cohort of eighth graders in 1988 who were surveyed through 2000 by the National Center for Education Statistics (NCES). The NCES randomly sampled schools across the country and then randomly sampled students within those schools. We use all students who were seniors during the 1991-1992 academic year, which yields a nationally representative sample of 7,920 observations. We utilize those variables that correspond to information available in the NLS-72; namely, sex, race/ethnicity, parental education, academic ability, college enrollment, college choice, and degree completion. The means and standard deviations of these variables are provided below.

**NELS Summary Statistics**

| | Mean | Std. Dev. |
|---|---|---|
| *Proportion* | | |
| Male | .47 | .50 |
| Black Male | .08 | .28 |
| Black | .03 | .18 |
| Hispanic | .12 | .33 |
| Asian | .07 | .25 |
| Dad: HS Grad | .74 | .44 |
| Dad: College Grad | .29 | .46 |
| Mom: HS Grad | .77 | .42 |
| Mom: College Grad | .24 | .43 |
| Enroll in college | | |
| Two-year | .31 | .46 |
| HBCU | .02 | .13 |
| Four-year, level unknown | .07 | .25 |
| Four-year, level 5 (highest) | .08 | .28 |
| Four-year, level 4 | .08 | .28 |
| Four-year, level 3 | .20 | .40 |
| Four-year, level 2 (lowest) | .10 | .30 |
| Four-year, level 1 (special) | .00 | .05 |
| Complete college degree | .41 | .49 |
| SAT Score (Predicted) | 850 | 200 |
| *N* | 7,920 | |

Source: Authors' calculations using NELS88.

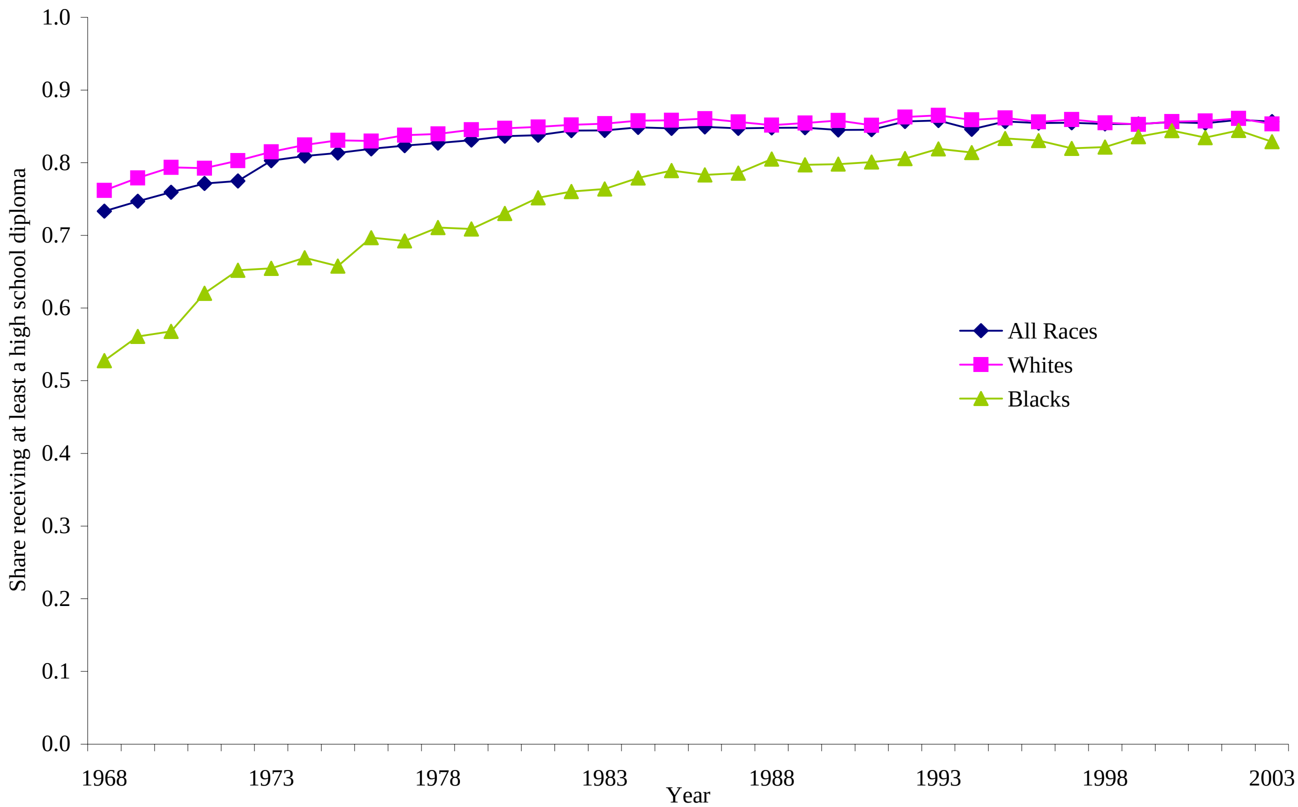
### A.4 Distribution of Undergraduate Majors by Field and Race

| | 1977 | | | 1997 | | |
|---|---|---|---|---|---|---|
| | Black | White | Total | Black | White | Total |
| *Selected Fields* | | | | | | |
| Engineering | 1.3% | 3.5% | 3.7% | 1.8% | 3.1% | 3.8% |
| Physical Sciences | 0.7% | 1.7% | 1.7% | 0.6% | 0.9% | 1.0% |
| Math & Computer Sciences | 1.0% | 1.5% | 1.5% | 3.2% | 2.5% | 2.9% |
| Life Sciences | 11.6% | 15.2% | 14.8% | 9.0% | 12.9% | 12.2% |
| Psychology | 3.1% | 3.2% | 3.2% | 3.7% | 3.8% | 3.7% |
| Social Sciences | 9.1% | 7.5% | 7.6% | 6.6% | 5.2% | 5.6% |
| Education | 20.8% | 15.0% | 15.1% | 9.2% | 11.0% | 10.0% |
| Business & Management | 18.2% | 16.8% | 17.0% | 20.3% | 18.4% | 19.1% |
| Total (N=) | 126,864 | 1,588962, | 1,860,510 | 217,545 | 1,853,808 | 2,577065 |

Source: Authors' tabulations from the Earned Degrees Conferred Survey.

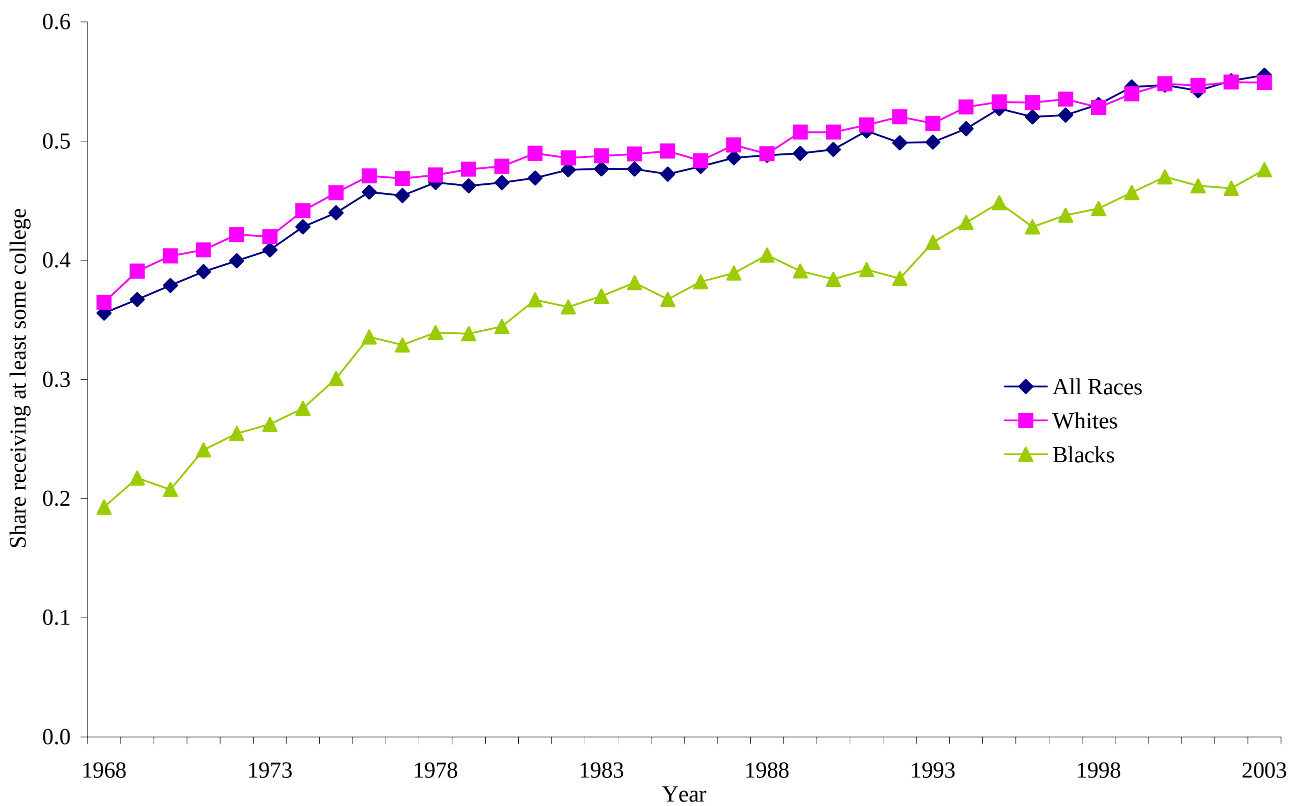
## B Tables & Figures

Figure 1: High School Attainment of 30-34 Year Olds by Year and Race



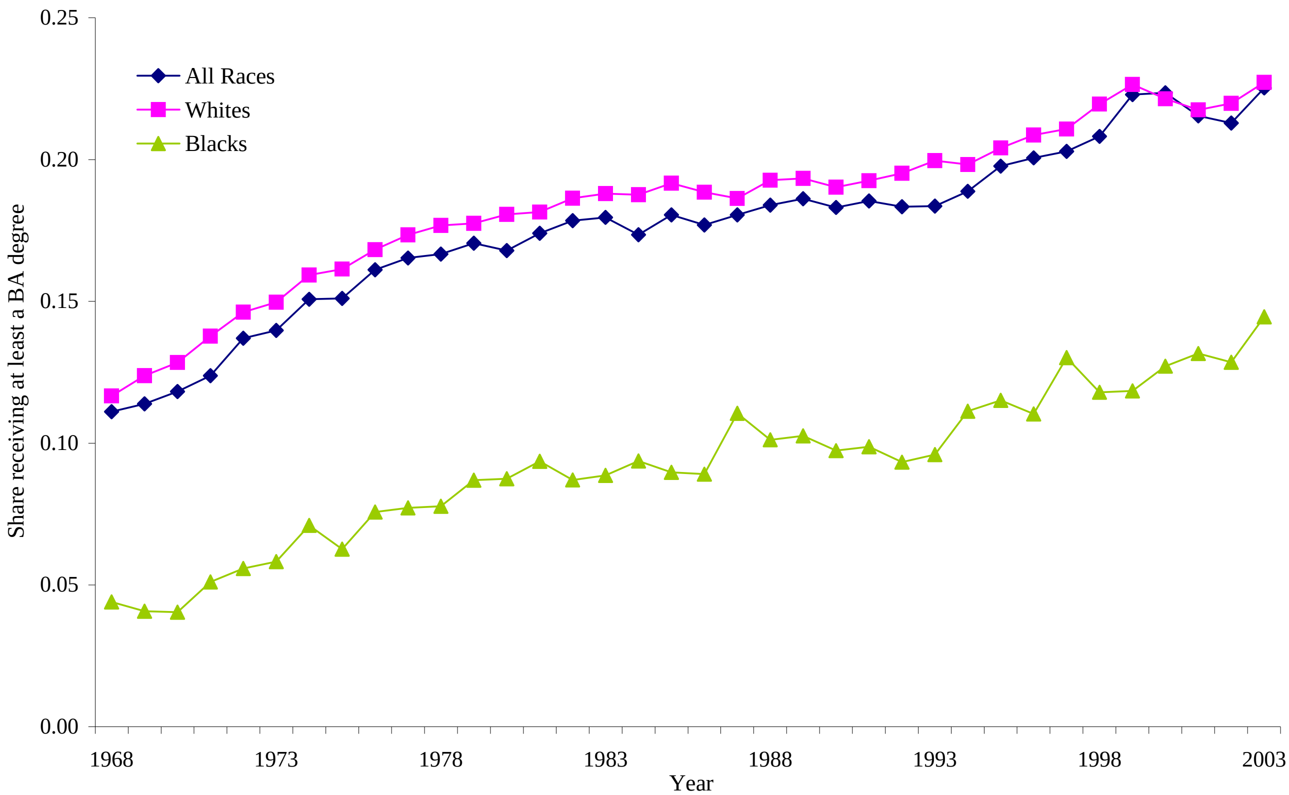

Source: Authors' tabulations from the Current Population Survey (Education and School Enrollment, October).

Figure 2: College Participation of 30-35 Year Olds by Year and Race



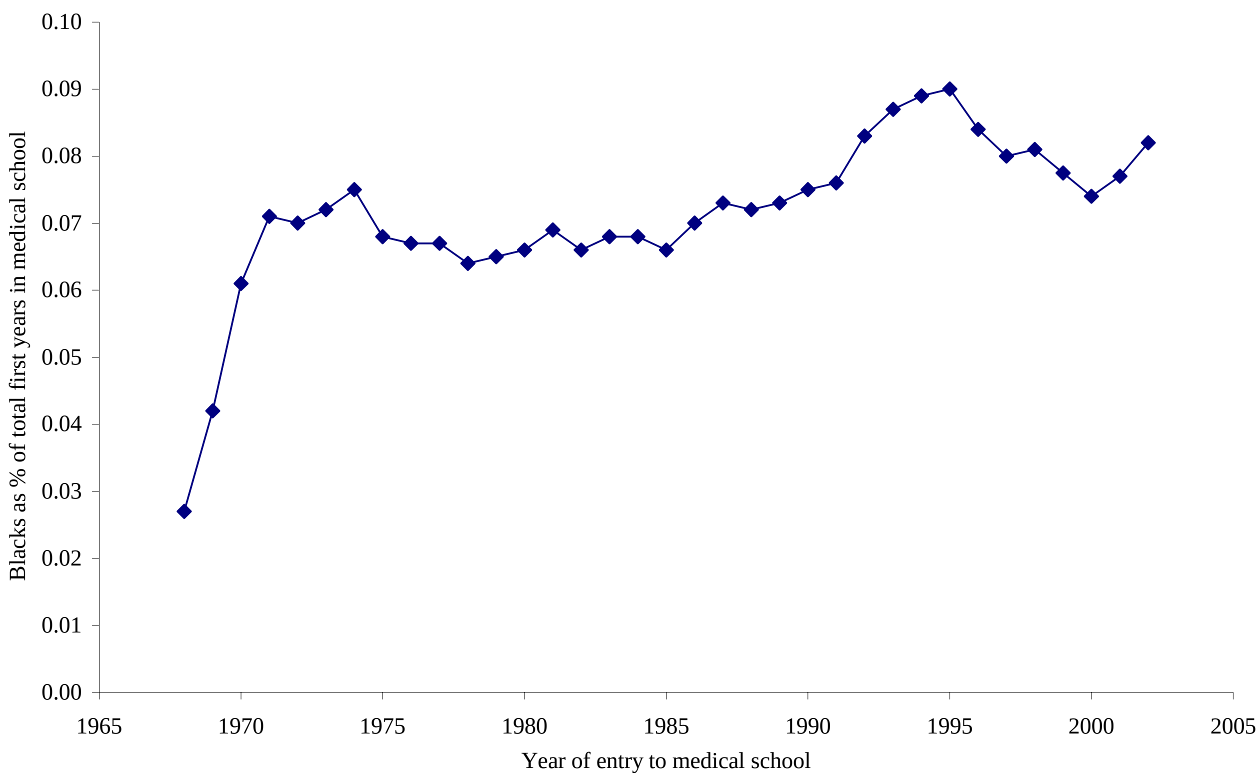

Source: Authors' tabulations from the Current Population Survey (Education and School Enrollment, October).

Figure 3: Baccalaureate Degree Receipt of 30-35 Year Olds by Year and Race



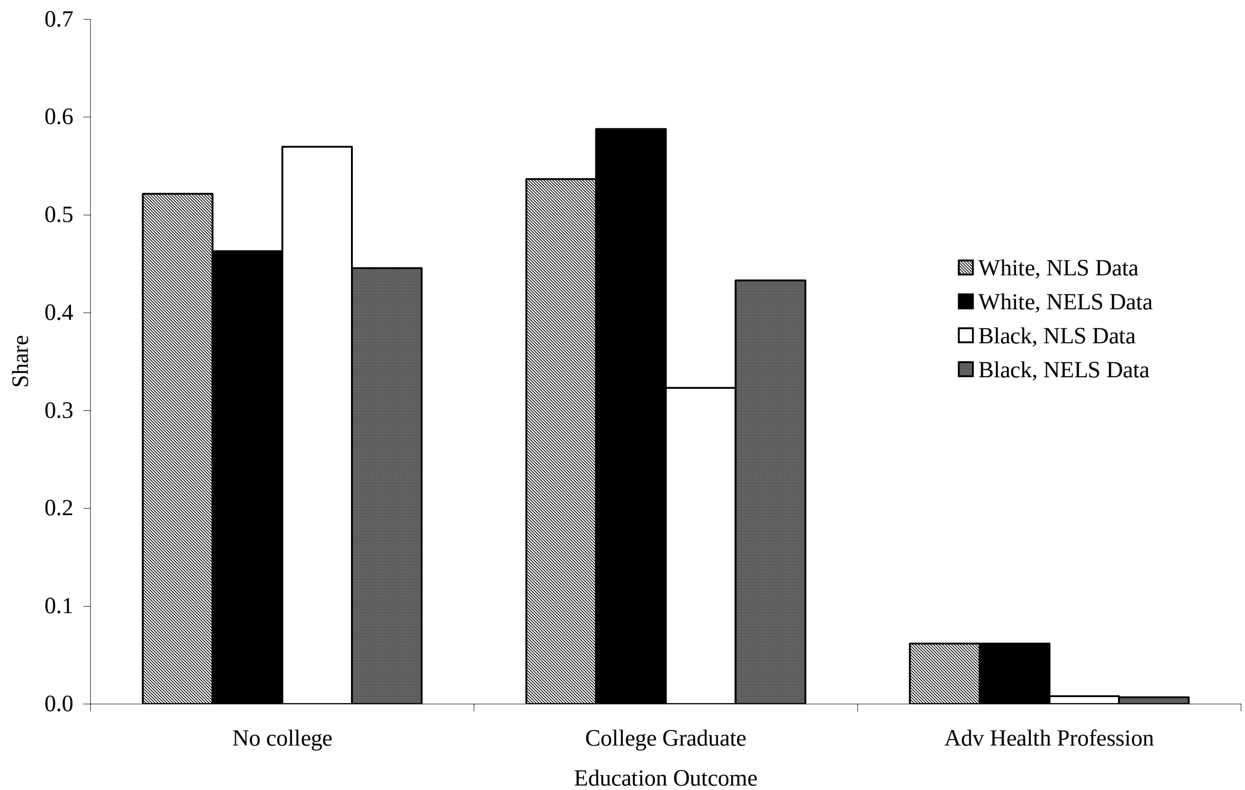

Source: Authors' tabulations from the Current Population Survey (Education and School Enrollment, October).

Figure 4: First Year Enrollment in Medical School, Proportion Black Over Time



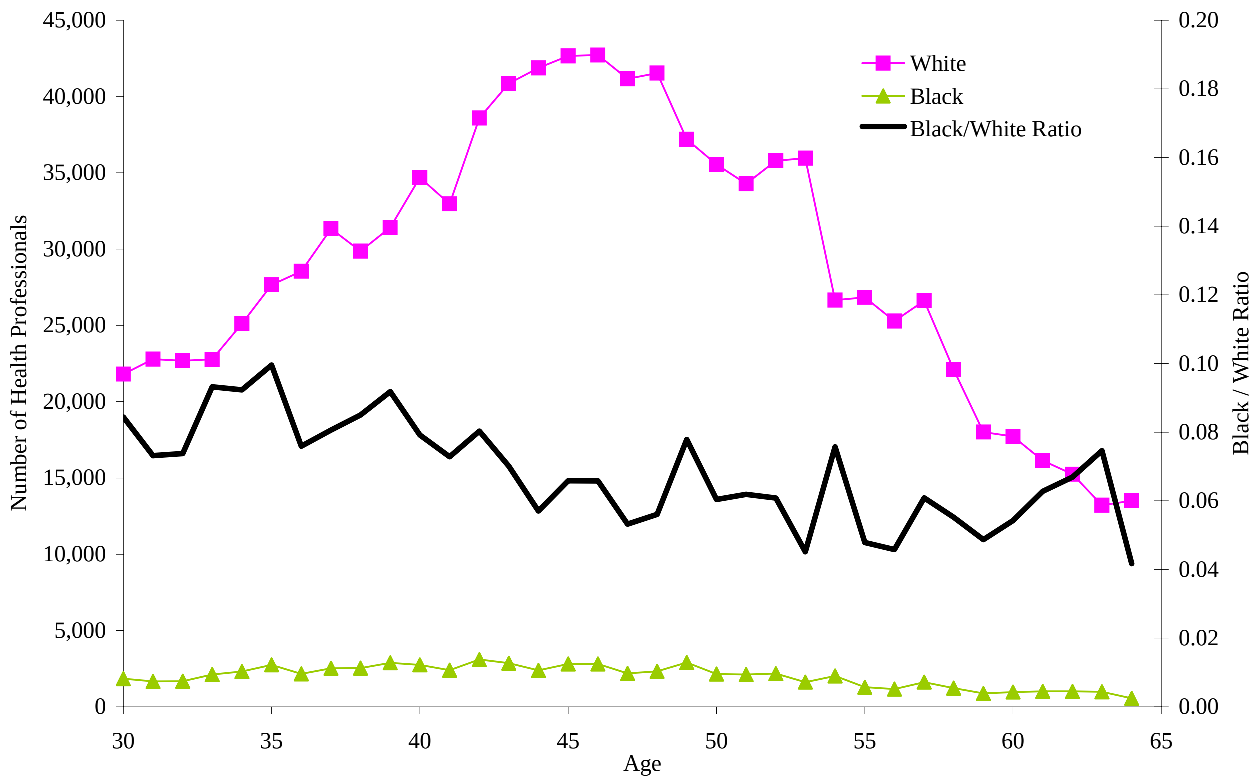

Source: Association of American Medical Colleges (2005).

Figure 5: Differences in Mean Probabilities Across Datasets Utilizing NLS-72 Parameter Estimates



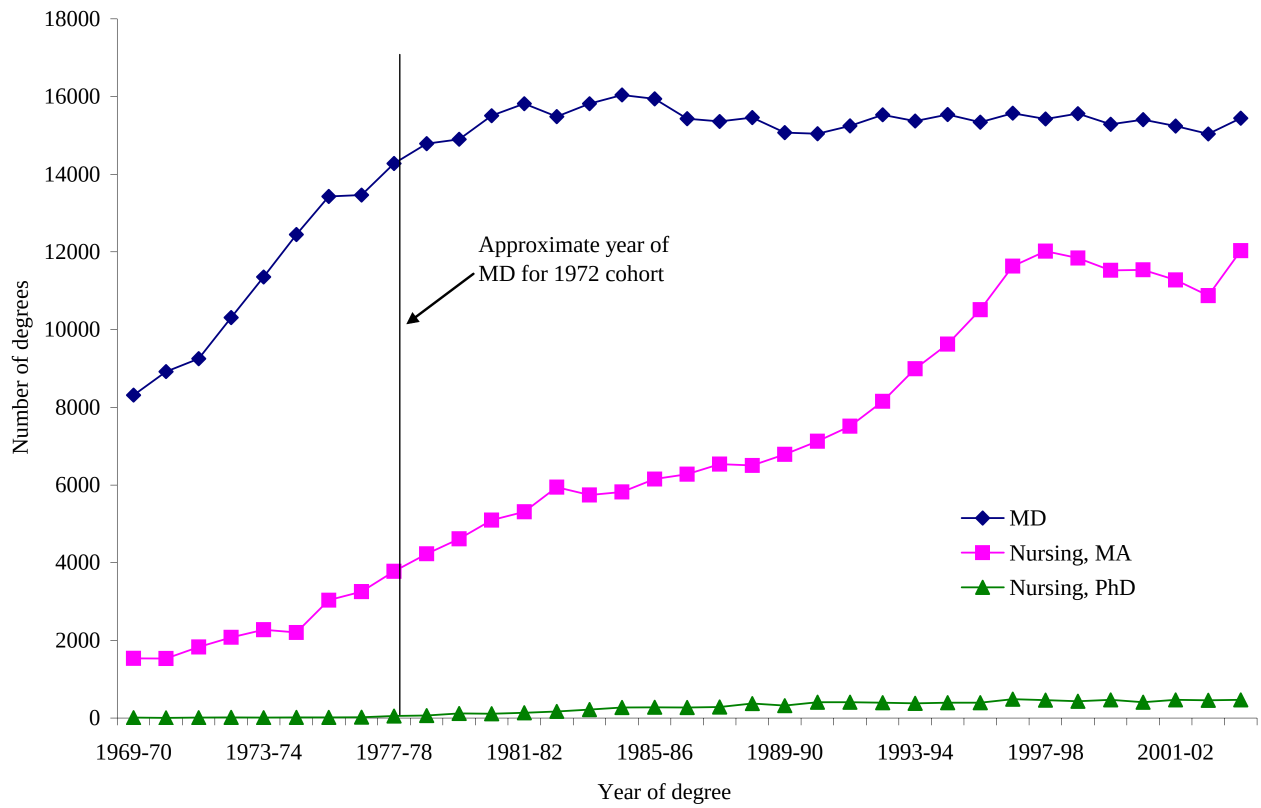

Source: Authors' calculations based on NLS-72 and NELS.

Figure 6: Black Health Professionals by Age, 2000




Source: IPUMS Census data, 2000.

Figure 7: Advanced Degrees Awarded in Major Health Professions




Source: Authors' tabulations from Earned Degrees Surveys (Department of Education).

Table 1: Summary Statistics of Individual Characteristics, NLS-72

| | **High School Graduates** | **College Entrants*** | **College Graduates†** | **Health Professionals‡** |
|---|---|---|---|---|
| *Proportion* | | | | |
| Female | .518 | .504 | .491 | .404 |
| Male | .482 | .496 | .509 | .596 |
| White | .824 | .850 | .884 | .917 |
| Black | .111 | .090 | .072 | .041 |
| Hispanic | .041 | .036 | .021 | .016 |
| Asian | .012 | .019 | .020 | .026 |
| Other | .012 | .005 | .003 | .000 |
| Dad: HS Grad | .717 | .785 | .826 | .850 |
| Dad: College Grad | .188 | .288 | .351 | .451 |
| Mom: HS Grad | .729 | .823 | .866 | .902 |
| Mom: College Grad | .107 | .168 | .211 | .295 |
| Enroll in college | .509 | 1.000 | 1.000 | 1.000 |
| Two-year | .093 | .182 | .080 | .047 |
| HBCU | .012 | .024 | .027 | .016 |
| Four-year, level unknown | .074 | .145 | .051 | .000 |
| Four-year, level 5 (highest) | .046 | .090 | .127 | .218 |
| Four-year, level 4 | .070 | .137 | .183 | .254 |
| Four-year, level 3 | .148 | .282 | .369 | .316 |
| Four-year, level 2 (lowest) | .073 | .130 | .152 | .135 |
| Four-year, level 1 (special) | .006 | .011 | .011 | .016 |
| Complete college degree | .299 | .586 | 1.000 | 1.000 |
| Become health professional | .015 | .029 | .050 | 1.000 |
| From rural community | .206 | .155 | .151 | .077 |
| Average SAT Score | 824 | 917 | 976 | 1094 |
| (standard deviation) | (217) | (206) | (194) | (177) |
| *N* | 13,014 | 6,629 | 3,885 | 193 |

* Refers to individuals enrolling in non-vocational two-year or any four-year colleges or universities.
† Refers to individuals receiving a baccalaureate degree from any four-year colleges or universities.
‡ Refers to individuals with post-baccalaureate degree who choose an occupation in the health professions.
Note: Proportion refers to the group indicated by the column heading. Categories may not sum to one due to rounding or non-exhaustive category choice.

Source: Authors' calculations using NLS-72.

Table 2: Between-Group Differences in Characteristics, NLS-72

| | High School Graduates | | | College Entrants* | | | College Graduates† | | | Health Professionals‡ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | White | Non-White | Diff | White | Non-White | Diff | White | Non-White | Diff | White | Non-White | Diff |
| *Proportion* | | | | | | | | | | | | |
| Female | .510 | .553 | -.043 | .494 | .557 | -.064 | .481 | .566 | -.084 | .407 | .375 | .032 |
| Male | .490 | .447 | .043 | .506 | .443 | .064 | .519 | .434 | .084 | .593 | .625 | -.032 |
| Dad: HS Grad | .722 | .452 | .270 | .826 | .551 | .275 | .858 | .586 | .272 | .870 | .625 | .245 |
| Dad: Coll Grad | .215 | .072 | .143 | .319 | .112 | .207 | .379 | .138 | .241 | .463 | .313 | .151 |
| Mom: HS Grad | .776 | .505 | .271 | .867 | .573 | .294 | .899 | .613 | .287 | .927 | .625 | .302 |
| Mom: Coll Grad | .118 | .058 | .060 | .182 | .088 | .095 | .222 | .118 | .104 | .299 | .250 | .049 |
| SAT Quartiles | | | | | | | | | | | | |
| Q1 (lowest) | .178 | .595 | -.418 | .063 | .418 | -.355 | .028 | .290 | -.262 | .006 | .125 | -.119 |
| Q2 | .254 | .228 | .026 | .188 | .271 | -.083 | .130 | .276 | -.146 | .034 | .188 | -.154 |
| Q3 | .279 | .117 | .162 | .308 | .192 | .116 | .294 | .232 | .062 | .175 | .063 | .113 |
| Q4 (highest) | .289 | .060 | .229 | .442 | .120 | .322 | .548 | .203 | .345 | .785 | .625 | .160 |

* Refers to individuals enrolling in non-vocational two-year or any four-year colleges or universities.
† Refers to individuals receiving a baccalaureate degree from any four-year colleges or universities.
‡ Refers to individuals with post-baccalaureate degree who choose an occupation in the health professions.
Note: Proportion refers to the group indicated by the column heading. Categories may not sum to one due to rounding or non-exhaustive category choice.

Source: Authors' calculations using NLS-72.

Table 3: Parameter Estimates - Basic Model

| | **College Entry*** | | | **College Degree**† | | | **Health Professional**‡ | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate | | SE | Estimate | | SE | Estimate | | SE |
| *Variables* | | | | | | | | | |
| Constant | -2.822 | ** | 0.047 | -1.257 | ** | 0.181 | -2.092 | ** | 0.286 |
| Male | -0.023 | * | 0.013 | -0.006 | | 0.013 | 0.098 | ** | 0.016 |
| Black | 0.841 | ** | 0.019 | 0.361 | ** | 0.022 | 0.290 | ** | 0.020 |
| Black*Male | -0.217 | ** | 0.012 | -0.082 | ** | 0.012 | 0.066 | ** | 0.017 |
| Hispanic | 0.598 | ** | 0.010 | -0.198 | ** | 0.010 | 0.134 | ** | 0.010 |
| Asian | 0.721 | ** | 0.009 | -0.030 | ** | 0.010 | 0.075 | ** | 0.009 |
| SAT Score / 1000 | 2.942 | ** | 0.044 | 1.656 | ** | 0.122 | 0.913 | ** | 0.173 |
| Dad: HS Grad | 0.242 | ** | 0.020 | 0.015 | | 0.025 | -0.132 | ** | 0.037 |
| Dad: Coll Grad | 0.289 | ** | 0.011 | 0.137 | ** | 0.013 | -0.009 | | 0.015 |
| Black*Dad: HS Grad | 0.054 | ** | 0.015 | 0.060 | ** | 0.017 | 0.089 | ** | 0.030 |
| Black*Dad: Coll Grad | -0.054 | ** | 0.011 | -0.461 | ** | 0.011 | -0.295 | ** | 0.016 |
| Mom: HS Grad | 0.212 | ** | 0.021 | 0.099 | ** | 0.028 | 0.020 | | 0.035 |
| Mom: Coll Grad | 0.340 | ** | 0.011 | 0.122 | ** | 0.012 | 0.051 | ** | 0.013 |
| Black*Mom: HS Grad | -0.353 | ** | 0.016 | -0.280 | ** | 0.018 | -0.443 | ** | 0.027 |
| Black*Mom: Coll Grad | -0.189 | ** | 0.011 | 0.190 | ** | 0.011 | 0.320 | ** | 0.021 |
| Rural | -0.229 | ** | 0.011 | 0.043 | ** | 0.011 | -0.228 | ** | 0.010 |
| Black*Rural | 0.007 | | 0.010 | 0.399 | ** | 0.010 | -2.126 | ** | 0.009 |
| *Covariance Matrix* | College Entry | | | College Degree | | | Health Professional | | |
| College Entry | 1.000 | | | | | | | | |
| College Degree | -0.311 | ** | 0.024 | 1.000 | | | | | |
| Health Professional | -0.446 | ** | 0.021 | 0.500 | ** | 0.042 | 1.000 | | |

* Refers to individuals enrolling in non-vocational two-year or any four-year colleges or universities.

† Refers to individuals receiving a baccalaureate degree from any four-year colleges or universities.

‡ Refers to individuals with post-baccalaureate degree who choose an occupation in the health professions.

Source: Authors' calculations using NLS-72.

Table 4: Parameter Estimates - Model with Variation in Colleges

| | **College Entry*** | | | **HBCU**⋆ | | | **College Degree**† | | | **Health Professional**‡ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | | SE | Estimate | | SE | Estimate | | SE | Estimate | | SE |
| *Variables* | | | | | | | | | | | | |
| Constant | -2.774 | ** | 0.048 | -2.654 | ** | 0.380 | -2.569 | ** | 0.173 | -2.806 | ** | 0.239 |
| Male | -0.006 | | 0.014 | 0.275 | ** | 0.064 | -0.020 | | 0.014 | 0.110 | ** | 0.014 |
| Black | 0.723 | ** | 0.022 | 2.040 | ** | 0.165 | 0.739 | ** | 0.028 | -0.509 | ** | 0.020 |
| Black*Male | -0.213 | ** | 0.013 | -0.234 | ** | 0.055 | -0.212 | ** | 0.012 | -0.070 | ** | 0.016 |
| Hispanic | 0.443 | ** | 0.010 | -2.324 | ** | 0.011 | 0.050 | ** | 0.011 | 0.156 | ** | 0.009 |
| Asian | 0.536 | ** | 0.009 | 0.652 | ** | 0.070 | 0.220 | ** | 0.011 | 0.025 | ** | 0.010 |
| SAT Score / 1000 | 2.930 | ** | 0.045 | -0.589 | ** | 0.192 | 2.439 | ** | 0.153 | 1.351 | ** | 0.171 |
| Dad: HS Grad | 0.232 | ** | 0.022 | 0.002 | | 0.435 | 0.105 | ** | 0.027 | -0.127 | ** | 0.030 |
| Dad: Coll Grad | 0.249 | ** | 0.013 | -0.538 | ** | 0.135 | 0.186 | ** | 0.015 | 0.010 | | 0.015 |
| Black*Dad: HS Grad | 0.110 | ** | 0.017 | -0.187 | | 0.360 | 0.105 | ** | 0.018 | | | |
| Black*Dad: Coll Grad | -0.025 | ** | 0.011 | 0.152 | | 0.116 | -0.431 | ** | 0.013 | | | |
| Mom: HS Grad | 0.176 | ** | 0.024 | -0.318 | | 0.454 | 0.201 | ** | 0.032 | -0.026 | | 0.032 |
| Mom: Coll Grad | 0.249 | ** | 0.011 | 0.474 | ** | 0.040 | 0.195 | ** | 0.013 | 0.072 | ** | 0.013 |
| Black*Mom: HS Grad | -0.303 | ** | 0.018 | 0.340 | | 0.384 | -0.426 | ** | 0.020 | | | |
| Black*Mom: Coll Grad | -0.119 | ** | 0.012 | 0.027 | | 0.032 | 0.157 | ** | 0.013 | | | |
| Rural | -0.190 | ** | 0.011 | 0.196 | ** | 0.035 | -0.048 | ** | 0.010 | -0.309 | ** | 0.010 |
| Black*Rural | -0.052 | ** | 0.011 | 0.122 | ** | 0.030 | 0.406 | ** | 0.011 | | | |
| $\boldsymbol{\alpha_1}$ | | | | 1.000 | | | | | | | | |
| $\boldsymbol{\alpha_2}$ | | | | -0.403 | ** | 0.044 | | | | | | |
| *Quality Thresholds* | | | | | | | | | | | | |
| 2-year / 4-year Level 1 | -1.198 | ** | 0.010 | | | | | | | | | |
| 4-year Level 1 / Level 2 | -3.633 | ** | 0.009 | | | | | | | | | |
| 4-year Level 2 / Level 3 | -1.157 | ** | 0.010 | | | | | | | | | |
| 4-year Level 3 / Level 4 | -0.220 | ** | 0.010 | | | | | | | | | |
| 4-year Level 4 / Level 5 | -0.423 | ** | 0.009 | | | | | | | | | |
| *Covariance Matrix* | College Entry | | | HBCU | | | College Degree | | | Health Professional | | |
| College Entry | 1.000 | | | | | | | | | | | |
| HBCU | 0.463 | ** | 0.213 | 1.000 | | | | | | | | |
| College Degree | -1.027 | ** | 0.190 | -1.921 | ** | 0.034 | 1.000 | | | | | |
| Health Professional | 0.002 | | 0.056 | 0.426 | ** | 0.123 | -0.181 | | 0.134 | 1.000 | | |

* Refers to individuals enrolling in non-vocational two-year or any four-year colleges or universities.

⋆ Refers to individuals enrolling in four-year historically black colleges or universities.

† Refers to individuals receiving a baccalaureate degree from any four-year colleges or universities.

‡ Refers to individuals with post-baccalaureate degree who choose an occupation in the health professions.

Note: Proportion refers to the group indicated by the column heading. Categories may not sum to one due to rounding or non-exhaustive category choice.

Source: Authors' calculations using NLS-72.

Table 5: Marginal Effects from the Model with Variation in Colleges:
Predicted Change in the Probability of Completing Each Stage in the Educational Pipeline Conditional on Successfully Completing Previous Stage(s)

| | College | **Attributes of the College Attended***⋆* | | | | | | | College | Health |
|---|---|---|---|---|---|---|---|---|---|---|
| | Entry* | 2-yr | 4-yr1 | 4-yr2 | 4-yr3 | 4-yr4 | 4-yr5 | HBCU | Degree† | Prof.‡ |
| *Variables* | | | | | | | | | | |
| Male | .0009 | **-.0006** | **-.0001** | **-.0007** | **-.0016** | **-.0009** | **-.0007** | .0054 | -.0069 | **.0235** |
| Black | **.2797** | -.0196 | **-.0011** | -.0063 | .0358 | **.0534** | **.0819** | .1358 | **.2517** | **-.0255** |
| Black * Male | **-.0708** | **-.0025** | **-.0004** | **-.0063** | **-.0233** | **-.0164** | **-.0150** | -.0063 | **-.0718** | -.0048 |
| Hispanic | **.1296** | **.0001** | **.0004** | **.0087** | **.0459** | **.0404** | **.0471** | -.0129 | **.0175** | .0136 |
| Asian | **.1832** | **-.0064** | **-.0001** | .0045 | **.0468** | **.0474** | **.0589** | .0321 | **.0778** | .0027 |
| SAT (100-pt. inc.) | **.0968** | **-.0019** | **.0001** | **.0046** | **.0311** | **.0285** | **.0296** | **.0017** | **.0885** | **.0313** |
| Dad's Education | | | | | | | | | | |
| HS Grad | **.0763** | **.0018** | **.0004** | **.0065** | **.0266** | **.0197** | **.0184** | .0027 | **.0368** | **.0225** |
| College Grad | **.0761** | **.0006** | **.0003** | **.0056** | **.0280** | **.0234** | **.0242** | -.0068 | **.0661** | **.0146** |
| Black * Dad's Educ. | | | | | | | | | | |
| HS Grad | **.0335** | **.0010** | **.0002** | **.0031** | **.0126** | **.0095** | **.0097** | -.0025 | **.0367** | **.0002** |
| College Grad | -.0066 | **-.0005** | **-.0001** | **-.0009** | -.0033 | **-.0022** | **-.0021** | **.0024** | **-.1398** | **.0001** |
| Mom's Education | | | | | | | | | | |
| HS Grad | **.0539** | **.0021** | **.0003** | **.0054** | **.0209** | **.0152** | **.0142** | -.0039 | **.0703** | **.0278** |
| College Grad | **.0883** | **-.0015** | **.0001** | **.0037** | **.0238** | **.0214** | **.0235** | .0162 | .0691 | **.0134** |
| Black * Mom's Educ. | | | | | | | | | | |
| HS Grad | **-.0935** | **-.0054** | **-.0007** | **-.0106** | **-.0354** | **-.0236** | **-.0208** | .0028 | **-.1394** | **.0002** |
| College Grad | **-.0381** | **-.0014** | **-.0002** | **-.0036** | **-.0133** | **-.0095** | **-.0088** | **-.0014** | **.0553** | **.0001** |
| Rural | -.0004 | **.0001** | **.0000** | .0001 | .0001 | .0000 | .0000 | **-.0006** | .0000 | **.0049** |
| Rural * Black | -.0003 | **.0001** | **.0000** | **.0001** | **.0001** | **.0000** | **.0000** | **-.0006** | .0000 | **.0004** |

* Refers to individuals enrolling in non-vocational two-year or any four-year colleges or universities.

⋆ The college attended falls into one of seven mutually exclusive categories: (1) 2-yr: non-vocational community colleges; (2) 4-yr1: special four-year institutions; (3) 4-yr2: non/less selective four-year colleges; (4) 4-yr3: somewhat selective four-year colleges; (5) 4yr-4: very selective four-year colleges; (6) 4-yr5: most selective four-year colleges; (7) HBCU: historically black colleges or universities regardless of institutional selectivity.

† Refers to individuals receiving a baccalaureate degree from any four-year colleges or universities.

‡ Refers to individuals with post-baccalaureate degree who choose an occupation in the health professions.

Note: Marginal effects represent the change in the conditional probability associated with a discrete change (0 to 1) in each of the binary variables listed in column one, with the exception of SAT score where we present the effect of a 100-point increase. Bolded effects are statistically significant at the 10 percent level or better.

Source: Authors' calculations using NLS-72.

Table 6: Actual and Predicted Outcomes

| | | Quintiles | | | | |
|---|---|---|---|---|---|---|
| *Outcome* | *Proportion* | First | Second | Third | Fourth | Fifth |
| Do not enroll in college | Predicted | 0.141 | 0.338 | 0.501 | 0.642 | 0.804 |
| | Actual | 0.138 | 0.335 | 0.494 | 0.652 | 0.824 |
| | Difference | 0.003 | 0.003 | 0.007 | -0.010 | -0.020 |
| Enroll in HBCU, no degree | Predicted | 0.001 | 0.039 | 0.073 | 0.092 | 0.122 |
| | Actual | 0.000 | 0.011 | 0.035 | 0.040 | 0.038 |
| | Difference | 0.001 | 0.028 | 0.038 | 0.052 | 0.084 |
| Enroll in HBCU, degree, no adv. health | Predicted | 0.000 | 0.040 | 0.061 | 0.080 | 0.121 |
| | Actual | 0.000 | 0.031 | 0.041 | 0.077 | 0.112 |
| | Difference | 0.000 | 0.009 | 0.020 | 0.003 | 0.009 |
| Enroll in non-HBCU, | Predicted | 0.025 | 0.048 | 0.060 | 0.069 | 0.079 |
| two-year, no degree | Actual | 0.035 | 0.051 | 0.065 | 0.091 | 0.112 |
| | Difference | -0.010 | -0.003 | -0.005 | -0.022 | -0.033 |
| Enroll in non-HBCU, | Predicted | 0.012 | 0.026 | 0.037 | 0.048 | 0.055 |
| two-year, degree, no adv. health | Actual | 0.010 | 0.013 | 0.025 | 0.030 | 0.037 |
| | Difference | 0.002 | 0.013 | 0.012 | 0.018 | 0.018 |
| Enroll in non-HBCU, | Predicted | 0.023 | 0.040 | 0.049 | 0.057 | 0.064 |
| four-year level 2 (lowest), no degree | Actual | 0.011 | 0.014 | 0.020 | 0.026 | 0.030 |
| | Difference | 0.012 | 0.026 | 0.029 | 0.031 | 0.034 |
| Enroll in non-HBCU, | Predicted | 0.010 | 0.025 | 0.042 | 0.057 | 0.068 |
| four-year level 2 (lowest), degree, no adv. health | Actual | 0.009 | 0.025 | 0.046 | 0.061 | 0.074 |
| | Difference | 0.001 | 0.000 | -0.004 | -0.004 | -0.006 |
| Enroll in non-HBCU, | Predicted | 0.030 | 0.057 | 0.078 | 0.094 | 0.108 |
| four-year level 3, no degree | Actual | 0.012 | 0.022 | 0.039 | 0.050 | 0.043 |
| | Difference | 0.018 | 0.035 | 0.039 | 0.044 | 0.065 |
| Enroll in non-HBCU, | Predicted | 0.015 | 0.043 | 0.085 | 0.141 | 0.201 |
| four-year level 3, degree, no adv. health | Actual | 0.010 | 0.042 | 0.089 | 0.154 | 0.226 |
| | Difference | 0.005 | 0.001 | -0.004 | -0.013 | 0.025 |
| Enroll in non-HBCU, | Predicted | 0.002 | 0.008 | 0.011 | 0.016 | 0.024 |
| four-year level 3, degree, adv. health | Actual | 0.002 | 0.010 | 0.018 | 0.009 | 0.020 |
| | Difference | 0.000 | -0.002 | -0.007 | 0.007 | 0.004 |
| Enroll in non-HBCU, | Predicted | 0.004 | 0.012 | 0.020 | 0.029 | 0.040 |
| four-year level 4, no degree | Actual | 0.004 | 0.009 | 0.019 | 0.021 | 0.023 |
| | Difference | 0.000 | 0.003 | 0.001 | 0.008 | 0.007 |
| Enroll in non-HBCU, | Predicted | 0.003 | 0.013 | 0.032 | 0.071 | 0.150 |
| four-year level 4, degree, no adv. health | Actual | 0.002 | 0.012 | 0.032 | 0.069 | 0.140 |
| | Difference | 0.001 | 0.001 | 0.000 | 0.002 | 0.010 |
| Enroll in non-HBCU, | Predicted | 0.001 | 0.008 | 0.011 | 0.017 | 0.030 |
| four-year level 4, degree, adv. health | Actual | 0.001 | 0.011 | 0.024 | 0.013 | 0.027 |
| | Difference | 0.000 | -0.003 | -0.013 | 0.004 | 0.003 |
| Enroll in non-HBCU, | Predicted | 0.003 | 0.008 | 0.011 | 0.014 | 0.022 |
| four-year level 5 (highest), no degree | Actual | 0.003 | 0.006 | 0.012 | 0.021 | 0.028 |
| | Difference | 0.000 | 0.002 | -0.001 | -0.007 | -0.006 |
| Enroll in non-HBCU, | Predicted | 0.002 | 0.011 | 0.024 | 0.056 | 0.178 |
| four-year level 5 (highest), degree, no adv. health | Actual | 0.002 | 0.012 | 0.021 | 0.046 | 0.155 |
| | Difference | 0.000 | -0.001 | -0.003 | 0.010 | 0.023 |
| Enroll in non-HBCU, | Predicted | 0.001 | 0.009 | 0.013 | 0.030 | 0.057 |
| four-year level 5 (highest), degree, adv. health | Actual | 0.000 | 0.013 | 0.012 | 0.021 | 0.040 |
| | Difference | 0.001 | -0.004 | 0.001 | 0.009 | 0.017 |

Notes: Quintiles are based on predicted probabilities of each educational/career outcome listed in the first column.

Table 7: Chi-Squared Goodness-of-Fit Test Statistics

| | **Quintiles** | | | | | **Overall** |
|---|---|---|---|---|---|---|
| *Outcome* | First | Second | Third | Fourth | Fifth | |
| Do not enroll in college | 0.23 | 0.11 | 0.27 | 0.35 | 1.48 | 2.44 |
| Enroll in HBCU, no degree | 4.24 | **7.76** | **7.87** | **11.74** | **22.30** | **53.92** |
| Enroll in HBCU, degree, no adv. health | **9.17** | 0.66 | 2.36 | 0.03 | 0.29 | 12.51 |
| Enroll in non-HBCU | | | | | | |
| two-year, no degree | **9.03** | 0.37 | 1.10 | **29.88** | **13.70** | **54.09** |
| two-year, degree, no adv. health | 0.71 | **16.66** | **10.90** | **21.30** | **16.53** | **66.10** |
| four-year level 2 (lowest), no degree | **13.67** | **39.77** | **36.76** | **63.12** | **42.45** | **196.76** |
| four-year level 2 (lowest), degree, no adv. health | 0.23 | 0.00 | 0.88 | 0.75 | 1.11 | 2.97 |
| four-year level 3, no degree | **26.16** | **54.23** | **56.33** | **63.44** | **76.20** | **276.37** |
| four-year level 3, degree, no adv. health | 4.22 | 0.03 | 0.63 | 3.02 | 8.24 | **16.14** |
| four-year level 3, degree, adv. health | 1.58 | 0.15 | 2.64 | 2.37 | 0.48 | 7.21 |
| four-year level 4, no degree | 0.09 | 1.57 | 0.40 | 5.33 | **18.03** | **25.41** |
| four-year level 4, degree, no adv. health | 2.46 | 0.38 | 0.00 | 0.05 | 1.74 | 4.62 |
| four-year level 4, degree, adv. health | 0.06 | 0.28 | **6.80** | 0.54 | 0.12 | 7.8 |
| four-year level 5 (highest), no degree | 0.15 | 0.78 | 0.19 | 3.42 | 1.75 | 6.28 |
| four-year level 5 (highest), degree, no adv. health | 0.61 | 0.15 | 0.73 | 3.37 | 5.24 | 10.10 |
| four-year level 5 (highest), degree, adv. health | 2.27 | 0.83 | 0.00 | 1.41 | 1.74 | 6.26 |
| Overall | | | | | | **747.98** |
| Normalized* | | | | | | **60.46** |

Notes: Quintiles are based on predicted probabilities of each educational/career path listed in the first column. Critical values for quintile statistics, which have one degree of freedom, are 3.84 (5 percent significance) and 6.64 (1 percent significance). Critical values for overall statistics, which have four degrees of freedom, are 9.49 (5 percent significance) and 13.28 (1 percent significance). Test statistics above the critical value imply a rejection of the null hypothesis that predicted proportions choosing each educational/career path equal action proportions making those choices in the data. Those statistics above the 1 percent critical value are bolded.

* Converted to a standard normal test statistic.