ERROR CORRECTION

Rindskopf, D., & Everson, H. *A comparison of models for detecting discrimination: An example from medical school admissions,* Volume 8, Number 1, pp. 89–106.

The following changes should be made:

1. Page 91, column 2, 7th line of text from bottom. The phrase should be "likelihood ratio chi-square statistic," not ". . . statistics."

2. Page 91, column 2, last line. The last sentence should start "It may very well be," not "It may very will be."

3. The figure labeled Figure 3 is Figure 4, and vice versa. In the caption to those figures, replace "Percent" by "Proportion."

4. In Figure 5, the latent exogenous variables (GPA, MCAT, & RACE) should be labeled with *xi*s (ξ), not *zeta*s (ζ).

5. In Figure 6, the residual for η_3 (ACCEPT) should be labeled ζ_3, not ξ_3.

6. Page 97, column 1, 1st full paragraph. The reference to Table 3 should be Table 4.

Readers should remove this page and insert it in their copy for future reference.

A Comparison of Models for Detecting Discrimination: An Example from Medical School Admissions

David Rindskopf
City University of New York Graduate Center

Howard Everson
U.S. Office for Civil Rights and
City University of New York Graduate Center

Detecting bias in admissions to graduate and professional schools presents important problems to the data analyst. In this paper some traditionally used methods, such as multiple regression analysis, are compared with the newer methods of logistic regression and structural equations models. The problems faced in modeling decision rules in this situation are (1) a dichotomous dependent variable, (2) nonlinear relationships between independent variables and the probability of being admitted, (3) omitted variables, and (4) errors in variables. Each method used involves an attempt to solve one or more of these problems, but each has its own drawbacks. Using multiple methods, and finding several areas of agreement in the results among the methods, makes the conclusions stronger than had only one method been used.

Despite the enactment of civil rights laws and a noticeable improvement in public attitudes, equal access to selective graduate and professional schools still eludes most minority applicants, according to a report from the Ford Foundation (Middleton, 1982). Consequently, graduate and professional schools are under increasing pressure to reexamine the educational rationale underlying traditional selective admissions practices. It has been proposed, for example, that professional schools using predictive models of admissions—particularly medical and law schools—examine separately the validity of the formulae for minorities, with an emphasis on the potentially disparate impact of standardized test scores (Middleton, 1982).

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 8, No. 1, Winter 1984, pp. 89–106

0146-6216/84/010089-18$2.15

Statistical methods have played a large role in the investigation of bias in admissions. They are useful for helping to establish whether or not bias exists, and for estimating its effects. The general procedure used, regardless of which particular statistical method is implemented, is to try to model the selection process. The goal of this article is to demonstrate, both theoretically and empirically, the traps and pitfalls in modeling selection. The first section of the paper describes the admissions process as the school sees it and the variables that are available to model this process. Then, the major potential problems in developing selection models in this situation are discussed, along with some methods that theoretically hold promise for solving some of these problems. The latter half of the paper is devoted to a detailed case study in which a variety of statistical techniques are applied to the analysis of data on admissions to a medical school.

An important point to note is that the problem confronted here is different from the usual case of detecting bias in selection. In the usual case, there is a criterion variable that is a measure of performance for those selected. A primary issue in such a situation is the possible differential predictive validity of the independent variables for various subgroups. This situation is the focus of papers such as those by Gross and Su (1975), Peterson and Novick (1976), and Breland and Ironson (1976).

The present situation is different in that admissions decisions are made without reference to a criterion (for example, performance in the first year of medical school), even though such variables may be available. There is an implication, of course, that the variables used in the selection process should be related to the probability of success (graduation versus nongraduation) or level of success (as measured, perhaps, by grades), but the decision-making model is not based on any empirical evidence of the shape of such a relationship, or on any possible differential validity for various subgroups. Those familiar with the more common selection situations will find that there are many statistical problems in common to both selection-validation studies and selection-modeling studies.

Perhaps the most commonly used technique for examining these selective admissions formulae is multiple regression. This technique allows researchers to include a "dummy" variable to denote the race of the applicant in the regression equations generated to model the admissions formulae. In this way models can then be generated to test competing theories of the selective admissions process.

Multiple regression techniques, however, are not without disadvantages. In this paper several statistical methods of detecting bias in admissions (here, to medical school) are discussed; each approach has strengths and weaknesses for a problem of this type. Some of the methods, such as polynomial regression or regression with interaction terms, are well known both in the literature on data analysis (e.g., Kerlinger & Pedhazur, 1973) and on detection of discrimination (Baldus & Cole, 1980). Others are not so well known; it is shown how methods such as logistic regression, nonlinear noncompensatory selection models, and structural equations models can be used (and misused) in modeling the selection process.

Data

Description of the Medical School and Its Selection Procedures

The data analyzed in this paper come from the 1979 to 1980 applicant pool of a state university medical college located in the northeastern United States. Applicants to this medical school are required to have completed at least 3 years of undergraduate study (i.e., 90 semester hours). Moreover, according to the stated policies of the medical school, while achieving excellence in the basic sciences of chemistry, biology, and physics is essential, academic work in the humanities and in the social sciences is also important. Admissions decisions are influenced by a review of each applicant's undergraduate record, letters of recommendation, Medical College Admissions Test (MCAT) scores, and an evaluation of the personal interview.

Invitations for a personal interview are issued only to those applicants whose academic records and other qualifications appear strong enough to justify detailed study. Interviews are ordinarily conducted by two or more interviewers independently, or by a group of interviewers. The interview, according to the school, provides an opportunity for discussion and clarification of the information submitted by the applicant. In addition, the interviewers use this opportunity to appraise such personal qualifications as responsiveness, warmth of personality, ability to adjust to social situations, poise, bearing, ability to communicate ideas clearly and concisely, and soundness of motivation.

Description of the Data Set

The total applicant pool included 3,773 applicants, 747 of whom were classified as minorities (i.e., Blacks, Puerto Ricans, Native Americans, and Asians). For purposes of analysis, a sample of 1,029 applicants (534 minorities and 495 whites) was selected randomly from the applicant pool. Of these, 124 applicants (38 minorities and 86 whites) were accepted into the medical school.[1] Data were available for each applicant on the following variables: undergraduate grade-point average (GPA)—science (SCIGPA), nonscience (NONSCIGPA)—MCAT scores on each of the six subtests—biology

[1]The authors did not consider the additional problem of predicting which students would actually come to the school after being admitted.

(BIOL), chemistry (CHEM), physics (PHYS), science problems (SCIAPPL), reading (READ), quantitative analysis (QUANT)—and whether or not the applicant was accepted (ACCEPT).

Statistical Problems

Problems Caused by a Dichotomous Dependent Variable

Unlike the usual situation in which statistical analyses are used to detect possible discrimination, the dependent variable in this case is dichotomous. This means either that traditional methods must be abandoned in favor of methods which adequately deal with qualitative data or that the traditional methods must be used with extreme caution. Because multiple regression methods assume a continuous dependent variable, the predicted values will not be integers, even though each observation (person) is either admitted or not admitted, i.e., he or she receives a score of either 0 or 1 on the dependent variable. The predictions, then, are more properly considered to be a statement regarding the probability that a person with particular values on the independent variable would be admitted. The values of the multiple correlation will of necessity be lower than would be the case if the dependent variable were continuous.

Problems Caused by Nonlinearity

Another problem is related to the fact that the outcome is dichotomous: The relation between the predictors and the outcome—predicted probability of acceptance—is probably nonlinear. If a linear model is used, predicted probabilities could be less than zero or greater than one. There are, however, a number of ways to deal with the problem of nonlinearity, some of which imply the use of (what are at this time) relatively new statistical methods. The approaches fall into three general categories:

1. Using a statistical method that reflects the hypothesized nonlinear form of the function;
2. Transforming the variables so that linearity is achieved;
3. Using a method that assumes linearity, hoping that the effects of nonlinearity are minor, then checking this assumption and interpreting the results with caution.

One of the methods of analysis used here is logistic regression, in which the *logit* of the probability of acceptance is considered to be a linear function of the independent variables. The logit is simply $\ln(p/(1-p))$, where ln is the natural logarithm and p is the predicted probability of acceptance. One advantage of using the logit as the dependent variable is that logits always correspond to probabilities between 0 and 1; therefore, no "impossible" probability estimates can occur. In addition, the form of the logistic distribution function is a reasonable one for this (and many other) problems involving a dichotomous dependent variable. The presumption is made that the relation between predictors and acceptance is only linear within a certain range of the independent variables; below a certain point, the probability of acceptance would be expected to be quite low, since there are so many better qualified applicants. A number of computer programs are available that can do logistic regression; here, program LR from the BMDP package was used.

In addition to logistic regression, other approaches were used to deal with the problems of nonlinearity. Complex models were developed based on multiple cutoff points, and polynomial regression was used. One problem arises in comparing methods: The multiple correlation can be used for the traditional regression models but is not appropriate for the logistic regression. In fact, it is uncertain whether *any* measure of goodness of fit is entirely appropriate for logistic regression, although some measures have been proposed. The likelihood ratio chi-square statistics was used in selecting models for logistic regression; even though this statistic can be misleading because of the small expected values, it is the best known measure of goodness of fit for loglinear models, of which logistic regression is a special case.

Problems Caused by Complex Admissions Models

Certain forms of nonlinear models create special problems. It may very will be, for example, that there is a minimum cutoff score on one or more variables that must be exceeded before a candidate will even be considered for admission. This type of rule can be accurately modeled using multiple regression (by creating special dummy variables) *if* the rule is known. Detecting whether there is such a rule and, if so, finding the form and cutoff points for the rule may be difficult. Sometimes, though, such rules can also be modeled by including the usual interaction terms or by using quadratic, cubic, or quartic polynomials. Logistic regression can also usually do an adequate job of fitting such models.

Problems Caused by Omitted Variables and Errors in Variables

In spite of the analyst's best efforts, almost any model-building attempt in the social sciences will fall short of perfection. Two prime reasons for this failure are omitted variables and errors in variables. An omitted variable is one that is not included in a model, either because it was not considered or because it could not be measured. If the omitted variables are uncorrelated with the dependent variable, there is no problem—that is why such variables as eye color, astrological sign, and so on can be omitted without harm. If an omitted variable is related to the dependent variable, but is uncorrelated with the independent variables in the model, then its omission will reduce the predictive accuracy of the model but will not bias the results. However, if an omitted variable is correlated with the dependent variable, after correcting for its relationship with the independent variables, then bias can result from its omission. That is, the size (or even the sign) of the coefficients of the other variables in the equation can be changed because an important factor has been omitted. The literature on this problem is extensive; almost any reference on causal models or econometrics discusses the topic—which is also called specification error.

A related problem is that observed measures are generally imperfect measures of the constructs they are designed to assess. It is usually not difficult to get agreement that most measures in the social sciences are imperfect; it is much more difficult, however, to demonstrate in an intuitive fashion how such errors can bias the results of an analysis. The culprit here is an artifact known as regression towards the mean; papers by Campbell and Erlebacher (1970) and by Campbell and Boruch (1975) are good sources of information on this problem, as is the chapter by Reichardt in Cook and Campbell (1979).

Sophisticated statistical methods are available for dealing with the problem. The general area known as structural equations models (Jöreskog, 1977) deals with causal models that can involve unobservable variables, each of which is measured by observable variables containing errors of measurement. A series of computer programs, all of which are based on Jöreskog's LISREL (*LI*near *S*tructural *REL*ationships model) approach, are available to analyze data for which such models are constructed. In this study, two different versions of the model were used—one that analyzes the whole sample as a group and the other that models separate groups (here, minority and white applicants were the two groups).

Problems with Structural Models Approaches

There are several problems in applying the structural models approaches with this data set. In the first place, they are not meant to deal with dichotomous variables. In the one-group model there are two dichotomous variables: acceptance decision (the dependent variable) and race (one of the independent variables). In the separate-groups form of the model, there is only one dichotomous variable. Because the tests for goodness of fit of the model assume multivariate normality of the variables, they could be inaccurate indicators of how good the models are. These problems may be minor if the distributions of the variables are not too skewed. Because of the way the sample was selected, there are about the same number of minorities and whites, but the acceptance ratio was only about 12%, which is rather low. An even more important problem is the lack of linearity, which not only can cause problems for the goodness-of-fit test but can make the interpretation of the model difficult or impossible. Since the authors' suspicions of nonlinearity of the data were confirmed, this is not just a theoretical problem here.

Analysis and Results

Descriptive

The correlation matrix, means, and standard deviations of the variables used on the analyses are presented in Table 1. The data for both whites and minorities are presented. A larger proportion of whites than minorities were accepted, but whites have much higher (GPAs) and MCAT scores (on the average) than do minorities. In addition, the smaller standard deviations for whites indicates that they are more homogeneous in terms of ability than the minority applicant pool.

Linearity Checks

To examine the relationship between the independent variables (GPAs and MCAT scores) and the dependent variable (admission), plots of percentage of students admitted versus the independent variables were constructed. MCAT scores ranged from 0 to 15 and were used without grouping. The GPAs ranged from 0.00 to 4.00, and were divided into intervals .10 units wide (e.g., 2.95 to 3.05). The plots for the whole sample are shown in Figure 1 for the MCAT CHEM scores. The expected nonlinearity can easily be seen, and a strong resemblance to the lower part of the logistic curve is evident. Logistic regression is, therefore, likely to prove a good method. The shapes of the curves might also be well approximated by polynomials, in particular, second- and fourth-degree polynomials.

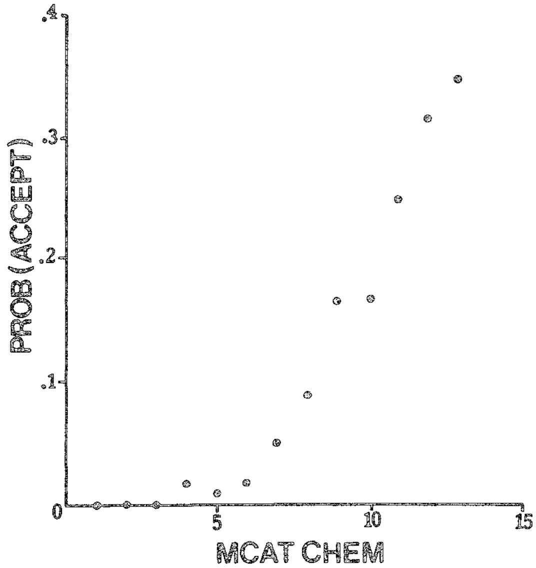
One point that is not evident from the plot is how many applicants are in the lower part of the curve where the nonlinearity occurs. In fact, approximately 90% (depending on the independent variable) are in the upper part of the curve, which is very nearly linear. For this reason, the transformations that were used in an attempt to induce linearity (as well as the polynomial models) did not produce as much benefit as was intended. An additional problem with attempts to transform the variables is that although a transformation might straighten out the relationship between the independent variable being transformed and the dependent variable, it often made the relationship among the independent variables nonlinear. That is, the relationships among the independent variables were very nearly linear before the variables were transformed, and the transformations more often than not made some of these relationships nonlinear. This had deleterious consequences for some of the structural models.

Although it was noted that the relationship between the independent variables and probability of

Table 1
Means, Standard Deviations, and Correlations for Whites
(N = 495, Below Diagonal) and Minorities (N = 534, Above Diagonal)

Variable	SCIGPA	NON-SCIGPA	BIOL	CHEM	PHYS	SCI APPL	READ	QUANT	ACCEPT	Minorities Mean	Minorities SD
SCIGPA	—	.503	.353	.445	.448	.408	.187	.358	.177	2.73	.66
NONSCIGPA	.542	—	.306	.282	.290	.281	.214	.254	.150	3.09	.54
BIOL	.331	.117	—	.723	.681	.791	.550	.608	.210	6.93	2.30
CHEM	.442	.191	.673	—	.803	.879	.479	.651	.192	7.11	2.26
PHYS	.377	.161	.608	.725	—	.859	.482	.688	.188	6.75	2.35
SCIAPPL	.405	.171	.770	.832	.801	—	.536	.679	.209	6.80	2.29
READ	.212	.162	.514	.500	.489	.533	—	.667	.160	6.06	2.28
QUANT	.308	.144	.502	.589	.614	.627	.645	—	.157	5.89	2.42
ACCEPT	.344	.191	.233	.269	.229	.235	.180	.228	—	.07	.25
Whites											
Mean	3.28	3.44	9.22	9.29	9.10	9.21	8.78	8.82	.18		
SD	.47	.41	1.86	1.89	2.14	1.86	1.89	2.04	.39		

Figure 1
Percent of Applicants Accepted by
MCAT CHEM Score



acceptance looked nonlinear, it appeared linear for a large range of the independent variables. The nonlinearity, however, is evident for the whole sample—the quadratic trend was significant at the .009 level in the CHEM subtest, for example—but not for the minorities and whites separately. This is because the distribution of the independent variables was so different for the minorities and whites. The minorities were heavily concentrated in the lower and middle parts of the distribution, and the whites in the middle and upper parts of the distribution. Table 2 shows the number of applicants in each race by SCIGPA and MCAT CHEM Score. Within each group the relationships were approximately linear, but the sample as a whole showed a nonlinear relationship.

This situation creates one problem for model construction and interpretation while it solves another. Because there are comparatively few minorities at the upper end of the distribution of independent variables, and so few whites at the lower end, it cannot be determined with certainty whether the relationships in either group would remain linear if there were a wider range on the independent variables. The probable case is that both would be curvilinear, following the trend evident in the whole sample. This implies that extra care must be used in interpreting the results of analyses, since extrapolation beyond the range of independent variables in which many applicants of each race are found will probably be misleading.

This can easily be seen by drawing one straight line through the lower part of the curve in Figure 2 (indicated by the broken line labeled A) and by drawing another straight line through the upper part of the curve (the broken line labeled B). If these straight lines are extended beyond the range on which they were calculated, then in the upper part of the range of independent variables, it would be concluded (falsely) that since Line B lies above Line A whites have a better chance of being accepted than minorities of equal ability. Looking at the lower part of the range of the independent variables, it can be seen that there are two problems: the prediction of a negative probability of acceptance for whites and the conclusion that minorities have a better chance of being accepted than whites. These false conclusions result from the extrapolation of results beyond the observed range of the independent variables.

Search for Complex Models

In order to obtain a better idea of possible complexities in the relationships between the independent and dependent variables, there was an examination of the probability of being accepted, broken down by both a grade-point-average variable (SCIGPA) and an MCAT variable (CHEM) at the same time. Because of the extremely high intercorrelations among MCAT scores, almost any one of the subtest scores would be an excellent proxy for the whole series of subtest scores. SCIGPA (rather than NONSCIGPA) was used because of the importance placed on performance in science courses by the institution in the selection process. In order to obtain a large enough number of observations in each SCIGPA by CHEM cell, each variable was categorized (SCIGPA into four categories and CHEM into five categories).

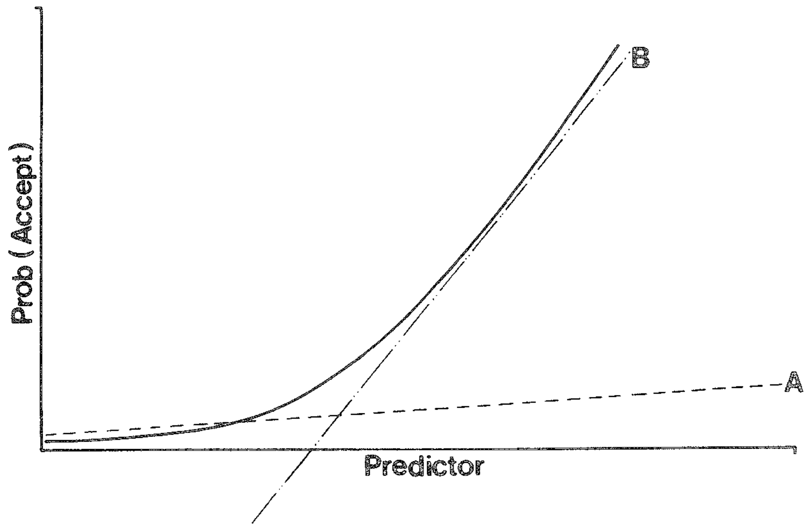
Table 2
Number of White and Minority Applicants
by Science GPA and MCAT Chemistry Score

Science GPA and Race	MCAT Chemistry Score 1-5	6-7	8-9	10-11	>11
<2.00					
White	3	3	1	0	0
Minority	27	26	11	3	1
2 to 3					
White	7	28	46	24	6
Minority	90	103	62	12	0
3 to 3.49					
White	3	27	84	75	15
Minority	14	42	32	30	3
3.5 to 4					
White	2	10	50	72	39
Minority	8	6	27	27	10

The results are displayed in Table 3. Within each cell the percentage of the sample who were accepted is listed for the whole sample, for minorities, and for whites. These results are consistent with the possibility that a multiple cutoff procedure may be in effect: If a person scores below 5 on CHEM or has a SCIGPA below 2.0, he or she will not be admitted. There are only two applicants who are exceptions to this apparent rule. (Of course, because of the high correlation among MCAT scores, almost any one of them might be used, or some combination, but the authors were unable to detect this.) If the data for whites is examined, the rule is *apparently* even more stringent: A white applicant will be considered for admission (apparently) only if he or she has a score of 8 or more on the CHEM subtest *and* has a 3.0 or better average in science courses. Only 2 white and 14 minority applicants were admitted who did not meet these standards. This represents an overall rate of less than 3% of the total sample of 1,029 applicants. In spite of appearing to favor minorities at the very low levels of MCATs and SCIGPAs, the *percentages* of whites and minorities in these categories who were accepted were not significantly different ($\chi^2=2.05$, $df=1$, $p>.05$). This is because there was such a large percentage of minorities falling into this area. As will be seen later, however, in the middle ranges of GPA (2.0 to 3.0), minorities do seem to have an advantage.

To return to the question of the apparent multiple cutoff procedure used and its implications for analysis, note that there are several ways to deal with the problem. First, the procedure could be explicitly modeled using dummy variables that would, in effect, build one model for the applicants who are minimally qualified (a model that itself could be complex) and another for those who are not (which could restrict their probability of acceptance to zero or allow it to be some small value which is estimated in the regression). However, a polynomial regression model, or possibly a model with interactions, might also be able to capture adequately the rule being used. Another approach would be to use—instead of a transformation of the independent variables, as is usually considered—a procedure similar to Winsorization: If a person received an MCAT score below the apparent cutoff value, the score would be set equal to the cutoff value. This would remove most, if not all, of the

Figure 2
Example of Curvilinear Relationship with Nearly Linear Parts



nonlinearity. Obviously, there is a wide variety of approaches to modeling the decision rule; and if more than one fits well, there is no way to say which is ''really'' the rule being used in admitting students.

Regression Models

The first stage in fitting regression models was to attempt to build models separately for both white and minority applicants using one GPA variable and one MCAT subtest score. The purpose of this was to investigate the possibility of the need for polynomial models without having to include all of the many interactions with race that might have been necessary had the sample not been split. The important result of this step was finding that polynomials were not necessary but that an interaction between MCAT scores and SCIGPA was. There was very little difference, in fact, between the polynomial models and the model with an interaction in this case because of the high correlation between MCAT scores and SCIGPA. Because of this, adding squared terms for MCAT or GPA would have approximately the same effect as adding the interaction between them, which is simply their product. The high degree of multicollinearity implies that many models might be found that fit the data about equally well; and this did, in fact, turn out to be the case.

In fitting a model for the whole sample, the first model included many variables that turned out to be either unnecessarily complex or just plain unnecessary. In order to make use of the information from all of the MCAT subtests without suffering problems due to their high intercorrelations, the six subtest scores were averaged[2] to create a variable called MCAT. In addition to SCIGPA and NONSCIGPA, race was included in the model, as well as all two-way interactions among variables. With all terms in the equation, the multiple correlation (R) was .394; dropping out all terms except MCAT, SCIGPA, Race, and the two interactions MCAT × SCIGPA and Race × SCIGPA kept the multiple correlation very high, at .393. Thus, grade-point average in nonscience courses (and its interactions) seems to contain no useful information in addition to the MCAT scores and SCIGPA.

Next, a model was fit using SCIGPA, MCAT CHEM, race, and their interaction terms. This model

[2]There is overlap among the science tests; the science applications ''subtest'' is made up of biology, chemistry, and physics problems, which are also counted in the individual subject-matter subtests.

Table 3
Percent of Applicants Accepted by
MCAT CHEM and SCIGPA

SCIGPA and Group	MCAT CHEM Score 0-5	6-7	8-9	10-11	>11
0 to 1.99					
Total	0	0	0	0	0
White	0	0	0	0	0
Minority	0	0	0	0	0
2 to 2.99					
Total	0	5	3	6	0
White	0	0	0	4	0
Minority	0	7	5	8	0
3 to 3.49					
Total	12	3	16	15	17
White	0	3	13	17	13
Minority	14	2	22	10	33
3.5 to 4.0					
Total	0	0	26	32	43
White	0	0	32	38	49
Minority	0	0	15	19	20

is even simpler, using only one of the MCAT scores instead of the whole test. With all terms in the model, the multiple correlation was .384—only slightly lower than that using the more complete model described above. After removing the three-way interaction and the Race × CHEM interaction, the multiple correlation was still .382. Therefore, the final model with SCIGPA, CHEM, race, and two interactions resulted in a multiple correlation only .01 lower than the most complete model considered.

The predicted probabilities of being accepted computed from this model are presented in Table 3 for selected combinations of CHEM and SCIGPA. Note that for whites, the predicted probabilities are negative for SCIGPA values less than 2.5; this demonstrates one of the problems with a dichotomous dependent variable. There are too few whites with low GPAs to reveal the nonlinearity that must exist for whites with low GPAs. Therefore, these numbers cannot be trusted—not just because they are negative, which is impossible—but because they are in an area that requires extrapolation. Extrapolating a linear function into an area that is suspected to be nonlinear can lead to incorrect (and in this case impossible) results. For most of the other values of CHEM and SCIGPA, there is more certainty that the results indicate the true state of affairs. It appears as if the response surfaces for the minorities and whites cross, so that at moderate levels of SCIGPA (2.0 to 3.0), minorities have a better chance of being admitted than whites, whereas at very high levels of SCIGPA, the opposite is true. A plot of the response surfaces, which clearly shows this result, was constructed from the results in Table 4 and is shown in Figure 3.

The negative predicted probabilities were worrisome, so some polynomial models were tried—even though it was known that they would not change the fit significantly—to see if there was enough information in the data to adjust the response surface. The attempt failed; no negative

Table 4
Predicted Probability of Admission for Whites and Minorities by MCAT CHEM Score and SCIGPA

SCIGPA and Race	MCAT CHEM Score 5	7	9	11
2.0				
White	−.14	−.15	−.15	−.16
Minority	.03	.03	.02	.02
2.5				
White	−.07	−.05	−.03	−.01
Minority	.03	.04	.06	.08
3.0				
White	.01	.05	.10	.14
Minority	.02	.06	.11	.15
3.25				
White	.05	.10	.16	.22
Minority	.01	.07	.13	.18
3.50				
White	.08	.15	.22	.29
Minority	.01	.08	.15	.22
3.75				
White	.12	.20	.28	.37
Minority	.01	.09	.17	.25

Note: The regression was

$$\hat{y} = .330 - .053 \times \text{CHEM} - .141 \times \text{SCIGPA} - .503 \times \text{RACE} + .165 \times \text{RACE} \times \text{SCIGPA} + .025 \times \text{SCIGPA} \times \text{CHEM}$$

predicted probability changed more than .01.

The final regression model fit was one in which there was a complex model allowing one prediction for those people below a cutoff point and another prediction for those above the cutoff. In this case, an applicant was considered to be above the cutoff if his or her SCIGPA was above 2.5 *and* MCAT CHEM score was 7 or above. (These cutoff points were chosen from examining the relationships between acceptance and each of the variables). For those above the cutoff, a model with interactions similar to that described in preceding paragraphs was fitted, whereas those below the cutoff were estimated to have an unknown (but constant within race and presumably small) probability of being selected. The multiple correlation was .394, exactly the same as the most complete model described above.

To summarize the regression models, there were models with polynomials and interactions, and models that, in addition, incorporated multiple cutoff scores. A fairly simple model consisting of one GPA variable, one MCAT subtest score, race, and two interactions did almost as well as the most complex model in capturing the selection process. (If these multiple correlations seem low, remember that with a dichotomous dependent variable there is an upper limit to the possible correlation. In this case, because the dependent variable is so skewed, the upper limit was approximately .62, which makes a multiple correlation of .38 seem much better than it would normally.)

Logistic Regression

Using logistic regression, the predicted probability of being admitted was computed for minorities and whites with varying scores on the MCAT test and SCIGPAs; the response surfaces were plotted and are in Figure 4. The variables initially used in the model were SCIGPA, MCAT SCIAPPL subtest score, race, and all two-way interactions. The terms retained in the final model were GPA, MCAT, race, and the GPA × Race interaction. The results were similar to those for the multiple regression models.

The surfaces clearly differ, and the fact that they cross indicates that for some combinations of GPA and MCAT whites are predicted to have a higher probability of being admitted. The advantage for whites appears to be greatest for students with high GPAs and MCAT scores but is still true for those with high GPAs and lower MCAT scores. For minorities, the advantage lies in the area of moderate GPAs and high MCAT scores. One explanation for this is that GPA may not be taken as seriously for minorities as for whites. It is as if a moderate GPA for minorities is not as likely to result in automatic rejection as it appears to be for whites; on the other hand, a high GPA for minorities is also not taken as seriously as for whites. A minority applicant with a very high SCIGPA (say 3.80) has a *lower* chance of being admitted than a white applicant with the same GPA, no matter what level of MCAT score the two applicants have.

Figure 3
Predicted Percent of Applicants Accepted by MCAT CHEM Score and Undergraduate GPA, for Whites and Minorities (Multiple Regression)

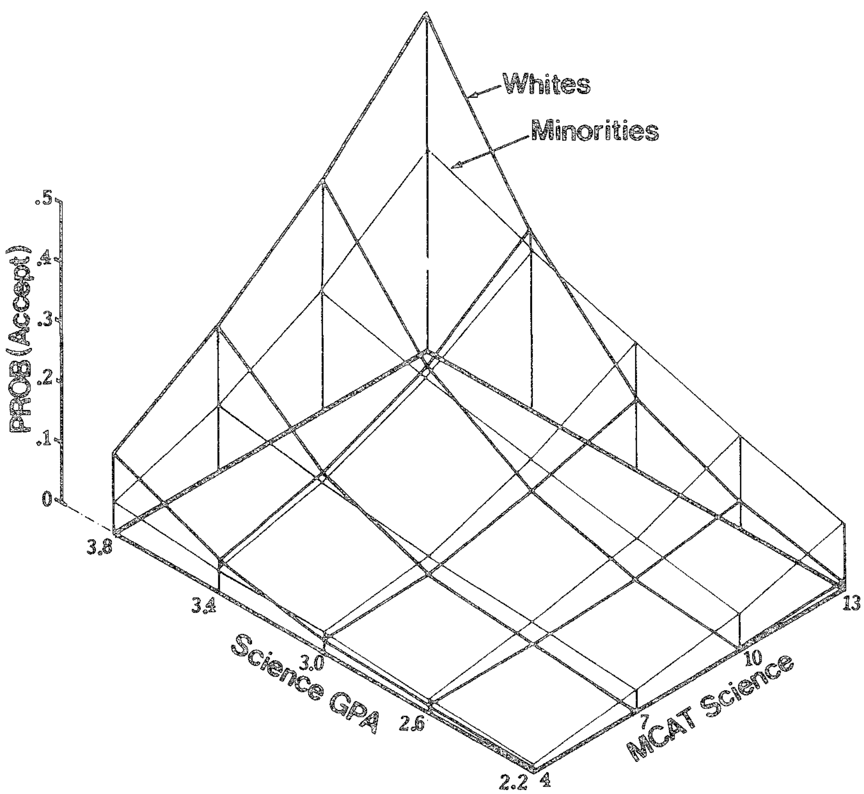
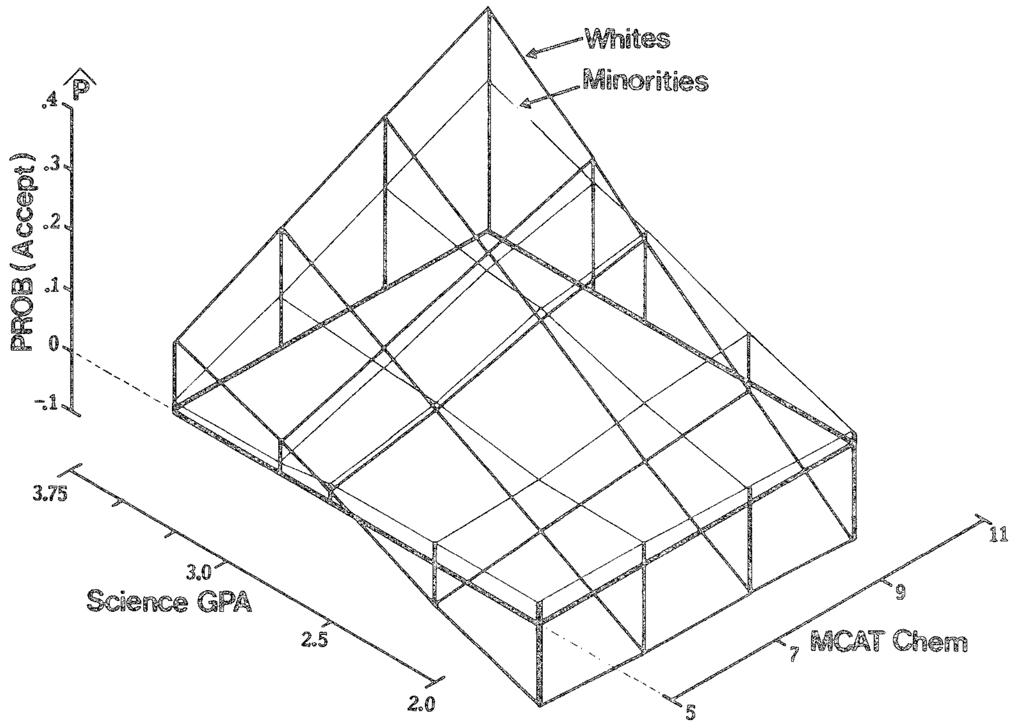


Figure 4
Predicted Percent of Applicants Accepted by
MCAT SCI Score and Undergraduate GPA,
for Whites and Minorities (Logistic Regression)



Factor Analysis of the Independent Variables

As a preliminary to developing structural models, a series of factor analyses were done to determine the factorial structure of the set of independent variables (GPAs and MCAT scores). The details will be omitted, since this phase is merely an intermediate one, of little consequence by itself. After an exploratory analysis, a number of two-, three-, and four-factor confirmatory models were tested, none of which fit the data. Several of the variables did not seem to fit the factor structure well, and a simplified model was tested with five variables: SCIGPA, NONSCIGPA, and MCAT BIOL, CHEM, and PHYS scores. A confirmatory model with two factors was tested; one factor was grades, upon which SCIGPA and NONSCIGPA were allowed to load, and the other factor was MCAT performance, on which BIOL, CHEM, and PHYS were allowed to load. The chi-square for testing goodness of fit was 4.52, with 4 degrees of freedom (p = .34), so the model fit very well and was accepted for use in building structural models.

Structural Model I: One-sample analysis

The first structural model tested hypothesized that there were three latent variables that might be affecting acceptance into medical school: GPA, MCAT performance, and race. The only dependent variable in the model is acceptance into medical school. The GPA and MCAT factors are measured by the five observed variables discussed in the preceding section on factor analysis, while race and acceptance are, of course, assumed to be perfectly measured. The model is diagrammed in Figure 5. This model did not fit well (χ^2 = 24.48, df = 10,

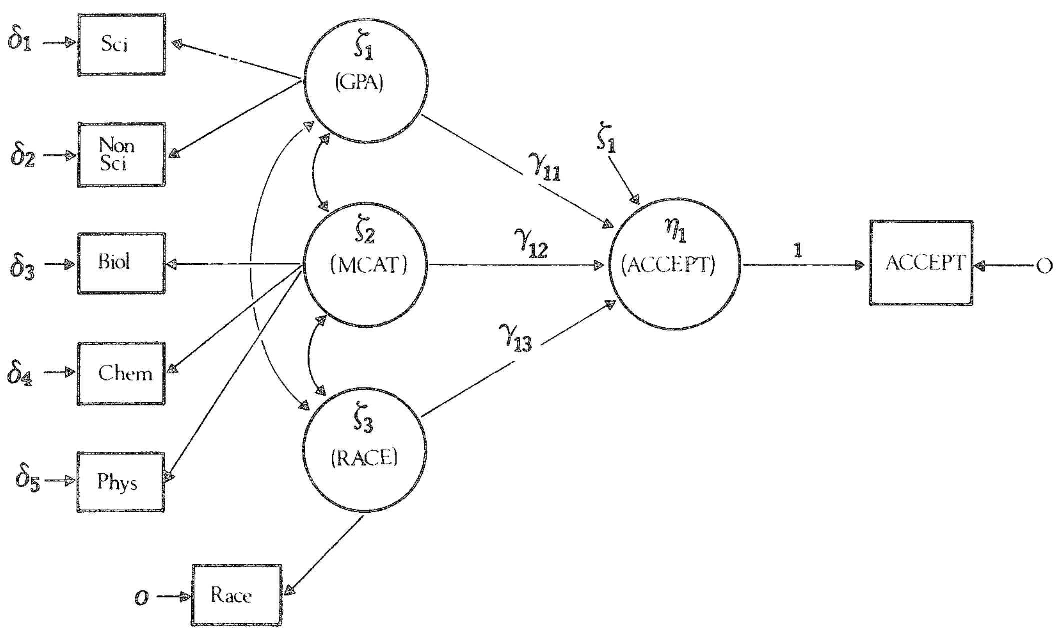
p = .006). All of the residuals were small, however, indicating that there might not be extensive problems if some small alterations were made. Examination of the residuals led to the hypothesis that two variables—NONSCIGPA and MCAT BIOL—may have different measurement characteristics for whites and minorities. Therefore, these variables were allowed to load on the race factor; this model fit well (χ^2 = 9.56, df = 8, p = .31), and the loadings of NONSCIGPA and BIOL on race were small (.065 and .080).

The meaning of the two loadings that were added is that for people of equal standing on the latent variables (GPA and MCAT), whites get slightly higher observed NONSCIGPAs and MCAT BIOL scores. The parameter estimates for the model are not presented here, but they indicated that although both GPA and MCAT factors had a large influence on admission, race had none. This seems to contradict earlier analyses. In fact, the previous analyses show why this problem occurred: The relationship between the independent variables and the dependent variable is different for whites and minorities. The one-group structural model cannot easily accommodate this; it assumes homogeneity of regression slopes, just as in an analysis of covariance with observed variables. The surprising part of this is that the model fit extremely well: The unequal slopes were not detected, and the apparent conclusion of no effect of race on admissions corresponds generally to the overall conclusion that would be made by averaging the results of a regression analysis. That is, in some instances whites are favored, whereas in others minorities are favored. The important lesson here is that a structural model may be incorrect and still fit the data very well.

Structural Model II: Separate-groups model

The second type of structural equation model tested involved fitting prediction equations separately for whites and minorities. This allows the relationships between the independent and dependent variables to be different for the two groups, thus solving one problem with the previous model. The problem of nonlinearity, however, is not solved by this approach; its probable effects are discussed

Figure 5
Structural Model for the One-Group Model

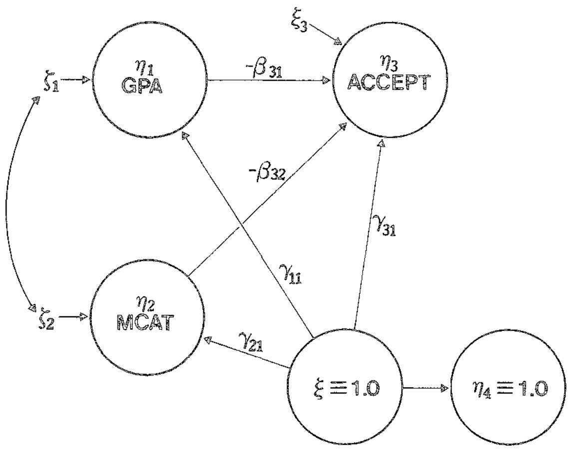


below. The analysis was done with the LISREL IV computer program (Jöreskog & Sörbom, 1978) using the method of moments about zero in order to estimate the means on the latent variables. This method is somewhat complicated, so it is not discussed in detail here, but a (somewhat terse) description can be found in the appendix to the LISREL IV manual; the LISREL V manual (Jörekog & Sörbom, 1981) contains a more detailed description of the method, including an example.

The structure of the model is shown in Figure 6 and 7. The basic structure is the same as in the one-group model in that two latent variables, GPA and MCAT, are presumed to influence acceptance into medical school. Because the races are modeled separately, no dummy variable for race is necessary. The "extra" latent variables in the model are included as "trick" variables, which will allow the model to include the means as well as variances and covariances of all of the variables (latent and observed).

As with the one-group model, the initial two-group model did not fit well ($\chi^2 = 36.52$, $df = 20$, $p = .013$). It was suspected that this lack of fit had similar origins for this model as for the one-group model—that is, the measurement models may differ for the two groups. One factor loading was allowed to differ for the groups; that model fit very well ($\chi^2 = 24.37$, $df = 19$, $p = .18$). The parameter estimates for this model are presented in Table 5, where it can be seen that the only difference in the measurement model (Λ_y matrix) for the groups is the intercept of the MCAT BIOL test score. Whites get higher MCAT BIOL test scores than minorities who are at the same level on the MCAT factor.

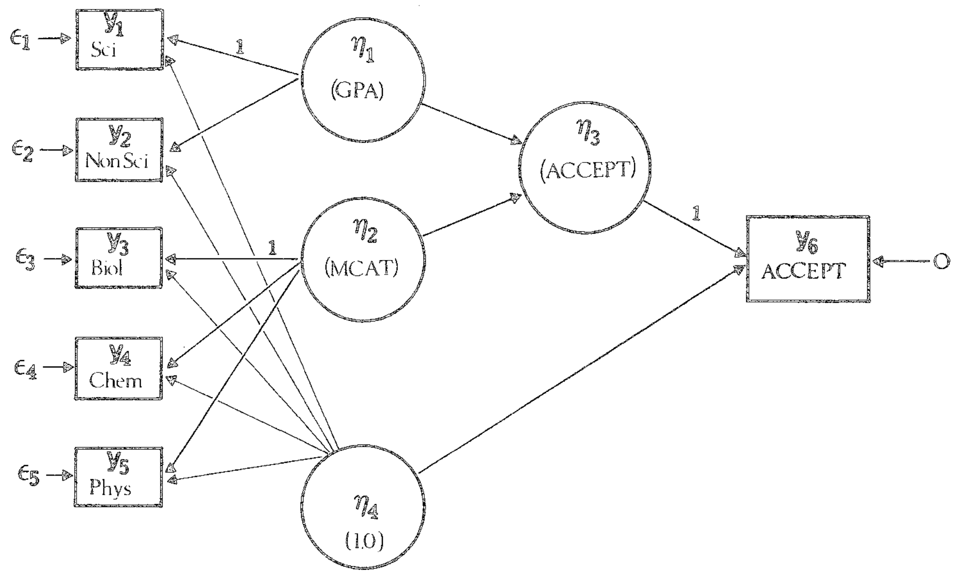
Figure 6
Structural Model for the Two-Group Model



The equations relating GPA and MCAT to probability of admission for each group are listed at the bottom of the table. The equations are very different for the two groups. (The coefficients in the equations are the parameters $-\beta_{31}$, $-\beta_{32}$, and γ_{31}.) The results of estimating the probabilities of acceptance for various combinations of values of η_1 and η_2 (the GPA and MCAT factors) for whites and minorities gives results almost identical to those obtained with the regression models and are therefore omitted. This result is not surprising, since the preliminary factor analyses indicated that SCIGPA and CHEM are very reliable measures of their respective factors, ensuring that the effect of measurement error in the regression results would be small. Whites have a higher probability of admission than minorities with equal standing on the GPA and MCAT factors for those with high GPAs, whereas the reverse is true for those with low GPAs.

In examining these results, notice that one of the problems discussed previously is making itself known. The nonlinearity cannot be easily modeled by the LISREL procedure; it is designed for *linear* structural equation systems. Although it may be possible in some instances to develop polynomial models, they are not simple to specify or to interpret. However, in this case the effects of nonlinearity are not catastrophic. As was seen in the figures showing the shape of the nonlinearity between the observed independent variables and the probability of acceptance, and from the heavy concentration of whites at the top and minorities at the bottom of the distributions of the independent variables, the response surfaces probably do an adequate job of summarizing the situation in the ranges of the independent variables for which they were calculated. Because there were few whites at the low end of the distribution, the negative predicted probabilities of acceptance can be ignored. The

Figure 7
Measurement Model for the Two-Group Model



predictions for minorities at the upper ends of the distribution should not be taken too seriously, although the situation is not as bad as for low-scoring whites. In spite of these problems, if care is taken not to extrapolate too far, it can be seen that the same conclusion would be drawn with this analysis as with the regression and logistic regression analyses: at the lower GPAs, minorities have a better chance of being accepted, whereas for high GPAs, whites have a better chance.

It was noted that one way to deal with the nonlinearity of the model is to be extremely careful about generalizing outside a region of the independent variables in which there are a reasonable number of observations. Another approach is similar to that taken earlier in the discussion of complex models. Instead of analyzing two groups, the method could be applied to three groups: whites above some assumed minimum cutoff point on the observed variables, minorities above the cutoff, and all applicants below the cutoff. If there is a large enough pool of applicants, the last group could also be split into whites and minorities to investigate the differences at the low end of the distribution of GPA and MCAT scores. It was felt that an accurate enough picture of the data was available without attempting to fit this model.

Discussion

This paper has discussed many problems involved in modeling the selection process: a dichotomous dependent variable, nonlinearity, possible complex models involving (for example) multiple cutoffs, errors in variables, and omitted variables. A wide variety of statistical techniques were applied—multiple regression using polynomials and special variables to model complex rules, logistic regression, and two forms of structural equation models. Each method has its strong and weak points, so that using any one alone might lead to false conclusions. In this case, using several methods made the weak points obvious and reinforced the strong points, making the conclusions more compelling.

The important finding, which was fairly consistent across methods of analysis, was that for students with moderate science GPAs, minorities had a better chance of being admitted than whites with the same MCAT scores, whereas for students

Table 5
Solution for 2-Group Structural Model

Nonzero Factor Loadings (Λ_y)

SCIGPA	1.0			2.727	
NONSCIGPA	.551			3.120	
BIOL		1.0		7.320	(6.938)*
CHEM		1.171		7.086	
PHYS		1.198		6.780	
ACCEPT			1.0	.067	

Other Parameter Values

	$-\beta_{31}$	$-\beta_{32}$	γ_{11}	γ_{21}	γ_{31}
White	.241	.044	.558	1.896	−.104
Minority	.058	.021	0	0	0
	ψ_{11}	ψ_{12}	ψ_{22}	ψ_{33}	
White	.200	.307	2.059	.127	
Minority	.322	.561	3.074	.059	
	$\Theta_{\varepsilon 11}$	$\Theta_{\varepsilon 22}$	$\Theta_{\varepsilon 33}$	$\Theta_{\varepsilon 44}$	$\Theta_{\varepsilon 55}$
White	.018	.114	1.491	.712	1.593
Minority	.109	.179	2.008	.823	1.250

*The value in parentheses is for minorities.

Note. The regression equations for the two groups were

$$\hat{\eta}_3 = -.104 + .241\eta_1 + .044\eta_2 \quad \text{(Whites)}$$

$$\hat{\eta}_3 = .058\eta_1 + .021\eta_2 \quad \text{(Minorities)}$$

with very high science GPAs, the reverse was true. The only method that did not find this was the one-sample structural equations model; its flaws prevented such a result from occurring.

What are some possible reasons for this result? The first possibility is that it is the applicant's race which is used in deciding how seriously to consider his or her GPA. However, an equally plausible theory (until further evidence can be gathered) is that minorities might attend schools judged to be of lower quality, and this is the characteristic that those who are deciding about admissions are really considering. That is, school quality is an omitted variable of possible importance. This hypothesis depends on the assumption that for students attending schools judged to be of low quality, GPA is not taken as seriously as for students attending schools judged to be of high quality. If, on the other hand, it were the case that any GPA from a "poor" school is judged to be worth less than an equivalent GPA from a "good" school, there would be constant bias against minorities, which was not observed. All of these hypotheses could be further complicated if there is a relationship between judged quality of undergraduate college and GPA which differs for whites and minorities. Since there is no data on school quality, the authors will refrain from elaborating on any of these complex hypotheses.

Another possibility is that however the effect works, it does so through influencing whether or not a student is invited for a personal interview. Because the admissions process is a two-stage procedure, a student who is not invited to interview has been rejected, whereas one who is invited still has some chance (at least theoretically) of being admitted. It may be that some minority students with GPAs and MCAT scores that would eliminate whites automatically are nonetheless invited for an interview. If so, they have a chance to convince the admissions committee of their potential in medical school, and some proportion may then be accepted. In an era of emphasis on equal opportunity, it is not an unlikely scenario to find medical school admissions committees making decisions in the following way. Highly qualified white applicants are easy to find, so there is no need to look very far down into the applicant pool. Therefore, many whites who could easily be expected to be successful in medical school may not even get interviewed. On the other hand, qualified minorities are harder to find, so that some (or many) must be invited to interview (and eventually be accepted) who would not even obtain an interview had they been white. To investigate these hypotheses, data on who gets interviewed would have to be examined.

Another alternative hypothesis is that minorities and whites have different patterns of performance on the personal interview (another omitted variable). This could have complex effects, particularly if the previous hypothesis about differential probability of being invited for an interview is true. If minorities are judged to perform more poorly than (otherwise qualified) whites, then minorities with high GPAs and MCAT scores will be less likely to be admitted than whites. If this is combined with a greater opportunity for minorities with moderate GPAs and MCAT scores to get a chance to interview, the results observed would be expected. Note that Milstein, Burrow, Wilkinson, and Kessen (1980) found evidence that discrepancy between the characteristics of the interviewer and the applicant had an effect on the probability of being admitted to medical school. The criteria they used did not include race, which may have been irrelevant considering the probable composition of the sample they used. In the present case, however, race (as a proxy for cultural background) could be expected to be an important factor and might make the effect of discrepancy in backgrounds even more salient than in the Milstein et al. (1980) study.

Suppose that these alternative hypotheses could be investigated and all of them eliminated except the hypothesis that race has a direct effect on probability of admissions. Would this bias mean that the procedure is unfair? Here is a new problem area. If predictors such as GPA and MCAT scores are not as valid in predicting success in medical school for minorities as they are for whites, then admissions decisions would be expected to reflect this. This would be particularly true if there are other predictors (e.g., interview performance) that were more strongly related to success in medical school for minorities than whites. In this case, just what was found might be expected: that the predictors examined were more strongly related to admissions for whites than for minorities.

Although it cannot yet be said what the admissions rule is, the results can be used to say what it is not. It is certainly not true that minorities are selected at random to satisfy equal opportunity criteria. For minorities as well as whites, the probability of being accepted does vary with qualifications, even if the relationship is not so strong for minorities. Also, it can be concluded that minorities do not get a "bonus" in admissions solely for being minority, because at the upper levels of GPA and MCAT scores they are less likely to be accepted than whites. However, *moderate-scoring* minorities may have an advantage over *moderate-scoring* whites, either directly in admission chances or in the likelihood of being interviewed. If there is a bias, it is not a consistent bias either for or against minorities.

The results of these analyses indicate that whites and minorities are treated differently, but the hypothesis cannot be ruled out that there is a reason for this which has nothing to do with race as such, but with a possibly valid omitted variable which may be correlated with race. This means that the search has not ended. Substantial knowledge has been gained about modeling the selection process; it is known that some variables have great influence, and some have no influence, as well as (at least approximately) how these variables are combined into a decision rule. There are other variables that need to be included to make the model more complete: Interview performance is particularly important, as well as information on regional, personal, and political factors in admissions.

References

Baldus, D. C., & Cole, J. W. L. *Statistical proof of intentional discrimination*. New York: McGraw-Hill, 1980.

Breland, H. M., & Ironson, G. H. *DeFunis* reconsidered: A comparative analysis of alternative admissions strategies. *Journal of Educational Measurement*, 1976, *13*, 89–99.

Campbell, D. T., & Boruch, R. F. Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations tend to underestimate effects. In C. A. Bennett & A. A. Lumsdaine (Eds.), *Evaluation and experiment*. New York: Academic Press, 1975.

Campbell, D. T., & Erlebacher, A. E. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), *Compensatory education: A national debate* (Vol. 3). *Disadvantaged child*. New York: Brunner/Mazel, 1970.

Cook, T. D., & Campbell, D. T. *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally, 1979.

Gross, A. L., & Su, W. Defining a "fair" or "unbiased" selection model: A question of utilities. *Journal of Applied Psychology*, 1975, *60*, 345–351.

Jöreskog, K. G. Structural equation models in the social sciences: Specification, estimation and testing. In P. R. Krishnaiah (Ed.), *Applications of statistics*. Amsterdam: North-Holland, 1977.

Jöreskog, K. G., & Sörbom, D. *LISREL IV: Analysis of linear structural relationships by the method of maximum likelihood*. Chicago: National Educational Resources, 1978.

Jöreskog, K. G., & Sörbom, D. *LISREL V: Analysis of linear structural relationships by maximum likelihood and least squares methods*. Chicago: International Educational Services, 1981.

Kerlinger, F. N., & Pedhazur, E. J. *Multiple regression in behavioral research*. New York: Holt, Rinehart, & Winston, 1973.

Middleton, L. Colleges urged to alter tests, grading for benefit of minority-group students. *Chronicle of Higher Education*, February 3, 1982, pp. 1;10–14.

Milstein, R. M., Burrow, G. N., Wilkinson, L., & Kessen, W. Prediction of interview ratings in a medical school admissions process. *Journal of Medical Education*, 1980, *55*, 451–453.

Peterson, N. S., & Novick, M. R. An evaluation of some models for culture-fair selection. *Journal of Educational Measurement*, 1976, *13*, 3–29.

Acknowledgment

The authors thank Alan Gross for a helpful discussion of some of the issues discussed in this paper. Work on this paper was supported by Grant No. 13631 from the Professional Staff Congress, City University of New York, Research Award Program.

Author's Address

Send requests for reprints or further information to David Rindskopf, Education Department, CUNY Graduate Center, 33 West 42nd Street, New York NY 10036, U.S.A.