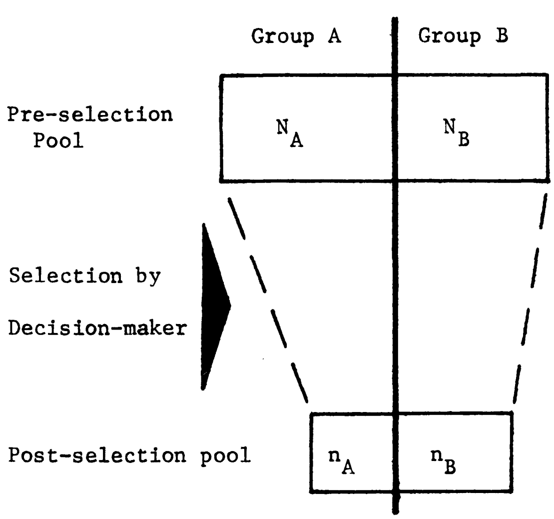
STATISTICAL MODELING TO SUPPORT A CLAIM OF INTENTIONAL DISCRIMINATION

James W. L. Cole, Pittsburgh, PA
David C. Baldus, University of Iowa

In the context of discrimination cases the courts have acknowledged that "Figures speak, and when they do, courts listen."[1] Indeed, conclusions drawn from statistical analyses often form an important and valued element of the evidence supporting claims of discrimination against minorities. But when one reads the cases in which statistics have been applied, one observes that the courts have left unresolved a number of methodological issues. Our purpose here is to examine some issues that arise in a particular class of discrimination cases and to pursue their resolution. The cases to be considered have the following defining properties in common:

1. The plaintiffs belong to a group (say, Group A) which is constitutionally or statutorily protected.
2. The practice under challenge is a selection process which assigns to each candidate for selection one of two outcomes, namely selection or rejection. Thus candidates are effectively winnowed out in a process represented schematically as in Figure 1 which highlights
   a) the <u>pre-selection</u> pool of applicants containing Group A (protected minority) and Group B (majority or other) components.
   b) the similarly constituted <u>post-selection</u> pool of those selected, and
   c) the decision maker, the individual or group actually making the choices.

FIGURE 1



3. The essence of the challenge is that the decision-maker was covertly applying a policy that used group membership to influence the applicants' selection chances and hence put Group A applicants at a disadvantage relative to others. However, those elements of the policy governing the selection process that are open to public view do not explicitly refer to group membership or any equivalent criterion as a factor influencing chances of selection.
4. The policy governing the selection process has left the decision maker ample room to use personal discretion in making choices, in that choices are not determined substantially by qualification criteria that are open to public view.

Selection processes susceptible to challenges with these properties will include those found in the substantive areas of employment selection, promotion, school admissions, some instances of criminal sentencing, and jury selection. The very early landmark case <u>Yick Wo</u> v. <u>Hopkins</u>[2] involved such a claim, and a long and rich line of challenges against jury selection processes provide numerous other examples.[3]

What is being claimed in cases having these properties is <u>intentional</u> discrimination of the type prohibited in particular under the due process and equal protection clauses of the 14-th Amendment of the U. S. Constitution. In two recent opinions, namely <u>Washington</u> v. <u>Davis</u>[4] and <u>Arlington Heights</u> v. <u>Metropolitan Housing Corporation</u>[5], the Supreme Court has made it clear that, to be successful, such challenges raised on constitutional grounds require proof of two facts. These are the existence of discriminatory impact, relative disadvantage falling to the plaintiff as a result of the suspect practice, and the existence of an intent to discriminate underlying the practice.

The dual nature of this requirement has been unclear in many decisions against jury selection systems because in these decisions a single piece of evidence has been found to establish a prima facie case of intentional discrimination. Specifically, the plaintiffs have prevailed on showing that over a suitably long period of time citizens with the same group membership have been substantially or consistently under-represented on the jury lists. However, in its discussion of requirements for proof of intentional discrimination in <u>Castaneda</u> v. <u>Partida</u>, the Supreme Court observed that the sufficiency of a single piece of evidence in such cases simply reflects the fact that the one fact can have two implications: the substantial under-representation speaks directly to the question of discriminatory impact while at the same time suggesting by its magnitude the presence of illicit motive.[6]

Selection processes can be classified usefully according as the decision-maker's choices are or are not guided by consideration of overt and verifiable qualification criteria. If the decision-maker's choices are influenced in part by such qualifications we will call the process a 'guided' discretionary process. In such processes, because equal protection doctrine is concerned with the equal treatment of equally situated candidates, such qualification variables that legitimately divide the candidates into equivalence classes of 'equal situation' must be taken into account before either a relevant measure of impact can be obtained or a valid inference of motive can be drawn.

If no such qualifications are cited, we will call the process 'purely' discretionary. With respect to those aspects of the governing policy that are open to public view, and hence for purposes of measuring the discriminatory impact of the discretionary aspects of the policy, each candidate who exceeds a basic threshold level of eligibility can be considered to be as qualified for selection as any other such candidate.

Because purely discretionary processes will be easier to model, we will consider them first. Fortunately, they do provide a suitable context for discussing interesting questions.

Modeling in Purely Discretionary Processes

For challenges of purely discretionary selection processes the four frequencies, N_A, N_B, n_A, and n_B found in the accompanying four-fold table contain the information usually considered necessary to establish a prima facie case of either discriminatory impact or discriminatory intent.

Table 1

Group	Numbers Selected	Numbers Rejected	Pre-selection Pool Totals
Group A	n_A	N_A-n_A	N_A
Group B	n_B	N_B-n_B	N_B
Totals	n	N - n	N

Usually the information in these quantities is summarized by a number or statement that compares one measure that reflects how the minority group was actually treated with a corresponding measure constructed to show how the group should have or would have been treated absent any discriminatory behavior. The following measures of actual treatment are frequently used:

a) the selection rate (or pass rate), $P_A = n_A/N_A$,

b) the rejection rate (or fail rate), $1 - P_A$,

c) the inverse selection rate, $1/P_A$,

d) the minority representation rate in the post-selection pool, $r_A = n_A/n$

(to be compared with $R_A = N_A/N$, the representation rate in the pre-selection pool), or even

e) the actual number of minority candidates chosen, n_A.

These rates or numbers can be compared in a variety of ways. Usually comparisons are made in terms of arithmetic differences or ratios, but other formulas have been suggested.[7] The courts have had some problems with the variety of measures available, the most obvious being a lack of consistency and direction in the choice of formulas for summarizing the numerical information bearing on the issue at hand. However, this lack of direction is not in itself the most troublesome aspect. As the courts move away from cases with clear-cut factual bases and encounter those that are closer, they are more inclined to compare the numbers in the case at hand with those of precedent cases.[8] To do this without adopting some function of the tetrad (n_A, n_B, N_A, N_B) that evaluates the evidence in terms of a single number is to trust the reliability of subjective judgment. On the other hand, any function chosen to evaluate the tetrad in the context of a particular question of fact should have an essentially monotonic (strictly) increasing relationship with the actual legal significance of the tetrad as it relates to the question of fact. For example, if in a given situation an arithmetic difference of 10 percentage points between R_A and r_A is much more strongly suggestive of unlawful motive when $R_A = 11\%$ than when $R_A = 91\%$, the blind dependence on R_A-r_A to support an inference of motive will invite the drawing of erroneous conclusions.

We propose that a primary determinant in choosing a measure should be the purpose of the measure--whether it is intended to measure discriminatory impact or to support an inference of motive. In measuring discriminatory impact, we are inquiring as to the relative harm done to members of the minority group as they go through the selection process, and this suggests that the underlying modeling be motivated by a utility theoretic approach. On the other hand, in inferring motive, we are seeking to identify an aspect of the decision-maker's behavior and principles of behavioral modeling should dominate.

Measuring the Discriminatory Impact. In a purely discretionary selection system, each Group A applicant entering the process can be said to have the same probability of being selected, say P_A, as any other such applicant. If we assume further that each candidate has the same (positive or negative) utility, say u, of being selected and utility 0 (zero) of being rejected[9], then the expected utility for a minority candidate is clearly

$$E_A(U) = u \cdot P_A \ .$$

Finally, assuming that in the absence of discrimination minority candidates would have the same probability of selection as that which the majority candidates have, say P_B, the expected utility for a minority candidate would be

$$E^*(U) = u \cdot P_B \ .$$

Therefore a measurement of harm should clearly be based on some comparison between $u \cdot P_A$ and $u \cdot P_B$. However, to be consistent with the principles of utility theory, a given arithmetic shortfall in the expected utility must represent the same degree of harm to the applicant regardless of the value of the expected utility that would obtain in a non-discriminatory process. That is to say, the appropriate measure of harm is the arithmetic difference

$$E^*(U) - E_A(U) = u(P_B - P_A).$$

Finally, since the minority and majority probabilities of selection are estimated without statistical bias by the corresponding observed selection rates, these results clearly suggest that a measure of discriminatory impact based on the difference between selection rates is preferable.[10]

A significant exception to this argument applies when the selection process under challenge is that of selecting the venire from which a jury will be chosen when that jury is to decide the fate of a minority criminal defendant. For in this situation, the courts must be concerned primarily with the impact on the rights of the defendant, not those of the prospective veniremen. Moreover, in this situation, the composition of the post-selection pool (that is, the venire from which the jury will ultimately be chosen) is the only aspect of the venire-selection process that bears on the defendant's rights. Consequently, while the difficulty of modeling the subsequent jury selection may make the application of the utility-theoretic argument impractical and its results of doubtful acceptability to the courts, it is clear that the minority representation rates that result from the venire selection process are the pivotal quantities in establishing the impact on the defendant's rights.

Modeling to Infer Discriminatory Motive. When we turn to the goal of modeling to infer motive, the focus of the model shifts from the treatment of the applicant to the behavior of the decision-maker. The first step in this modeling process is to pin-point as nearly as possible the kind of non-discriminatory selection mechanism the decision-maker might have used or claims to have used. The second is to adopt a model that appears to simulate that mechanism adequately. The third is to consider how discriminatory behavior might manifest itself in the context of the given selection mechanism. The fourth is to decide how that form of discriminatory behavior is to be incorporated into the model. Finally, we apply the conclusions drawn in the first four steps to the pursuit of our current objective, namely to choose a measure best suited to reflecting behavior suggestive of a discriminatory motive in the eyes of the court.

For example, the selection of prospective jurors from the citizens in a district is often supposedly done by application of an explicitly random mechanism. The appropriate model for such a process will be more or less obvious depending on the complexity of the mechanism and the care taken to adhere to it. For example, in straightforward situations the model of simple random sampling from the pool of all citizens meeting specified eligibility requirements may suffice, but when allowances are made for various forms of hardship, the model may require modification.

In the context of such a process, two forms of discrimination would seem to be most likely. The first is the ever-present possibility that the number of Group A applicants was held below some tacit quota. The second is the exclusion from consideration of some fixed portion, consisting say of πN_A individuals, from the pool of eligible Group A candidates.

If the process is being influenced by a desire to limit Group A selections to a quota, it is more likely that the decision-maker is concerned with controlling the size of the Group A representation rate than the Group A selection rate. Hence this desire is more likely to be reflected in the representation rate than the selection rate. Specifically, if the nominal model for the process is one of simple random sampling, the distribution of the representation rate would be expected to differ from the nominal binomial or hypergeometric distribution in two ways. Naturally, the mean of the distribution will be reduced from the expected N_A/N. But also the distribution would likely be truncated, especially at the upper end. This suggests that when results from several applications of the suspect process are available, as is often the case in jury selection challenges, the entire empirical distribution of representation rates may be relevant.[11]

If the Group A participation in the selection process is limited by exclusion of a fixed proportion of eligible candidates, an estimate of the excluded proportion based on the representation rates has been suggested.[12] This estimate is found by simple algebra to be

$$\hat{\pi} = \frac{R_A - r_A}{R_A(1 - r_A)}$$

Purely random mechanisms are less common in employment selection, where typically some form of evaluation or ranking will be used. If all the evaluation criteria are strictly subjective and hence inaccessible for verification or challenge, the process will still fall in the 'purely discretionary' class. However, the five-step process of analysis and modeling should reflect the dependence on the criteria if possible.

Such selection mechanisms provide more interesting modeling challenges. The particular model chosen will depend, first, on the nature of the criteria on which the informal evaluation is based; second, on what can be postulated as reasonable distributions for these criteria in the populations of candidates; and finally, on the kind of reasoning that was used to combine these criteria into a single score.

In order to develop an illustration, albeit more valuable for insights produced than for realism, suppose the following. First, for each candidate there exists a vector of qualification scores $\underset{\sim}{X}_{ij}$ $(i = 1, \cdots, N_j;\ j=A,B)$. Second, these vectors are p-variate normally distributed within groups--

$$\underset{\sim}{X}_{ij} \sim N_p(\underset{\sim}{\mu}_j, \underset{\sim}{\Sigma})$$

--with the same covariance matrix but possibly different means. And finally, the decision-maker looks at these criterion scores, constructs some weighted sum of them, say

$$Y_{ij} = \underset{\sim}{\lambda}_j' \underset{\sim}{X}_{ij},$$

and selects those candidates for whom Y_{ij} exceeds a cut score y_j^*.[13]

This model will reflect an absence of intentional discrimination only if $\underset{\sim}{\lambda}_A = \underset{\sim}{\lambda}_B$ and $y_A^* = y_B^*$. Any difference between the weight vectors or the cut scores would indicate that group membership was being used to influence the decisions.

According to this model the weighted sum Y_{ij} will be normally distributed,

$$Y_{ij} \sim N(\underset{\sim}{\lambda}_j'\underset{\sim}{\mu}_j;\ \underset{\sim}{\lambda}_j'\Sigma\underset{\sim}{\lambda}_j \equiv \sigma^2(\underset{\sim}{\lambda}_j)).$$

Thus, the probability that a candidate i, being randomly drawn from group j, will meet the standard for selection is

$$P(Y_{ij} \geq y_j^*) = \Phi\left(\frac{\underset{\sim}{\lambda}_j'\underset{\sim}{\mu}_j - y_j^*}{\sigma(\underset{\sim}{\lambda}_j)}\right).$$

Therefore we find that $\Phi^{-1}P(Y_{ij} \geq y_j^*)$ will be in the form of a linear model--

$$\Phi^{-1}\, P(Y_{ij} \geq y_j^*) = \alpha_0 + \alpha_1 Z_{ij}$$

--where $Z_{ij} = 1$ for candidates in Group A, and 0 otherwise; and the unknown coefficients have the following form:

$$\alpha_0 \equiv \frac{\lambda_B' \mu_B}{\sigma(\lambda_B)} - \frac{y_j^*}{\sigma(\lambda_B)}$$

$$\alpha_1 \equiv \frac{1}{\sigma(\lambda_B)}\, \lambda_B'\, (\mu_B - \mu_A) + \left(\frac{1}{\sigma(\lambda_A)}\, \lambda_A' - \frac{1}{\sigma(\lambda_B)} \lambda_B'\right) \mu_A + \frac{y_A^*}{\sigma(\lambda_A)} - \frac{y_B^*}{\sigma(\lambda_B)}$$

In this linear model, the parameter α_1 contains all the evidence of discriminatory behavior. This suggests using an appropriate estimate of α_1 as our measure from which to infer motive. From the methodology of probit analysis[14] we find the maximum likelihood estimate of α_1 to be $\Phi^{-1}(p_B) - \Phi^{-1}(p_A)$, where p_A and p_B are the observed selection rates for the two groups.

In practice, this measure would likely be resisted as unfamiliar and based on unverifiable assumptions. Thus we seek a more familiar and intuitive measure that would produce similar results in the case-to-case comparisons. We turn first to the logit function, $L(p) \equiv \ln\left(\frac{p}{1-p}\right)$. Since, as can be shown,

$$0.83 < \frac{(L(P_B) - L(P_A))/1.9}{\Phi^{-1}(P_B) - \Phi^{-1}(P_A)} < 1.16 \quad (.05 \leq P_A, P_B \leq .95)$$

the difference $L(p_B) - L(p_A)$ suggests itself as an alternative with the virtue of being easily explained in terms of betting odds. Finally, we note that if p_A and p_B are both small, the simple ratio p_B/p_A will give case-to-case comparisons most consistent with those of $\Phi^{-1}(p_B) - \Phi^{-1}(p_A)$ among the measures now commonly used.

As a last word it should be noted that α_1 will reflect the arguably innocent effect of the group mean differences in qualification score inextricably confounded with the influences of intentional discrimination. This fact would suggest an obvious defense to the claim, and hence it raises legal and methodological questions that are serious and most interesting, but beyond the scope of this discussion.

Modeling in Guided Discretionary Processes

In the following discussion of guided discretionary processes we will assume that there is no challenge to the legitimacy of the overt qualification criteria guiding the decision-maker, but that in making the final selection decisions, the selector has covertly used group membership as a factor. Thus it is alleged that, within the 'equivalence classes' of candidates defined by virtue of being similarly situated with respect to the overt criteria, the selector operated to the disadvantage of the Group A members. Consequently, we shift our focus, both in measuring discriminatory impact and in detecting evidence of unlawful intent, to comparisons of treatment observed within the equivalence classes.

Measuring the Discriminatory Impact. Let $\underset{\sim}{W}_{ij}$ denote the vector of overt qualifications and let $P_j(\underset{\sim}{w})$ denote the probability of selection for a group j candidate whose $\underset{\sim}{W}_{ij}$ vector equals $\underset{\sim}{w}$. Then invoking the same assumptions of utility structure and the same argument as in the previous section, the suggested measure of impact for a minority candidate with qualification vector equal to $\underset{\sim}{w}$ would be an appropriate estimate of the difference $P_B(\underset{\sim}{w}) - P_A(\underset{\sim}{w})$.

This difference can be estimated in two general ways: (1) directly from selection outcomes observed for groups of candidates having $\underset{\sim}{W}_{ij} = \underset{\sim}{w}$, if there exist such groups of sufficient size, or alternatively, (2) adopting models for $P_A(\underset{\sim}{w})$ and $P_B(\underset{\sim}{w})$ as functions of $\underset{\sim}{w}$ and estimating the parameters therein by appropriate methods.

If the latter approach is chosen, it will again be apparent that the model to use will again depend on the particular situation. But in this application of modeling, the final result will be judged purely on the reliability of the estimates that it provides for $P_B(\underset{\sim}{w}) - P_A(\underset{\sim}{w})$ as indicated in part by measures of goodness-of-fit of the model. If such measures indicate the need for a model in which $P_B(\underset{\sim}{w}) - P_A(\underset{\sim}{w})$ varies with $\underset{\sim}{w}$, the lack of a single number summarizing the magnitude of the impact for the whole case will clearly complicate case-to-case comparisons. These complications are intrinsic to situations in which some portions of Group A may have been treated more adversely than others, and it would be unwise to confine the choice of models artificially to those which have an additive term corresponding to group membership.

Modeling to Infer Discriminatory Motive. When modeling to lay the basis for an inference of motive in a guided discretionary selection process, the same multi-step analysis of the mechanism should be applied as for a purely discretionary process, with the recognition that observed values for the overt qualification criteria will permit empirical goodness of fit testing of some aspects of the resulting model.

To illustrate, we will extend the informal evaluation model of the completely discretionary process section to incorporate the qualification variables $\underset{\sim}{W}_{ij}$. Thus we will assume that adjoining the q variable vector $\underset{\sim}{W}_{ij}$ to $\underset{\sim}{X}_{ij}$ produces a vector distributed (p+q)-variate normally--

$$\begin{pmatrix} \underset{\sim}{W}_{ij} \\ \hline \underset{\sim}{X}_{ij} \end{pmatrix} \sim N_{p+q}\left(\begin{pmatrix} \underset{\sim}{\nu}_j \\ \hline \underset{\sim}{\mu}_j \end{pmatrix}; \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right), \quad \begin{pmatrix} i = 1, \cdots, N_j \\ j = A, B \end{pmatrix}$$

And, as before, we assume that the decision-maker constructs some weighted sum of all the qualification scores, say

$$Y_{ij} = \underset{\sim}{\delta}_j' \underset{\sim}{W}_{ij} + \underset{\sim}{\lambda}_j' \underset{\sim}{X}_{ij}$$

and selects those candidates for whom Y_{ij} exceeds y_j^*. In the context of this model, intentional discrimination would be implied by any of the following findings:

$$\underset{\sim}{\lambda}_A \neq \underset{\sim}{\lambda}_B, \quad \underset{\sim}{\delta}_A \neq \underset{\sim}{\delta}_B, \quad \text{or} \quad y_A^* \neq y_B^* .$$

According to this model, the distribution of Y_{ij} conditioned on a particular value for $\underset{\sim}{W}_{ij}$ is normal with mean $\underset{\sim}{\delta}_j' \underset{\sim}{w} + \underset{\sim}{\lambda}_j'[\underset{\sim}{\mu}_j + B(\underset{\sim}{w} - \underset{\sim}{\nu}_j)]$ and variance $\sigma^2(\lambda_j) \equiv \underset{\sim}{\lambda}_j' \Sigma_{22\cdot1} \underset{\sim}{\lambda}_j$, where $B \equiv \Sigma_{21}\Sigma_{11}^{-1}$ and $\Sigma_{22\cdot1} \equiv \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Thus the conditional probability that a candidate i in group j will be selected given $\underset{\sim}{W}_{ij} = \underset{\sim}{w}$ is

$$P(Y_{ij} \geq y_j^* | W_{ij} = w) = \Phi\left(\frac{w'(\delta_j + B'\lambda_j) + \mu_j^{*\prime}\lambda_j - y_j^*}{\sigma(\lambda_j)}\right),$$

where $\mu_j^* = \mu - B\nu_j$ is the vector of X intercepts in the regression functions on w. Again we find that $\Phi^{-1}P(Y_{ij} \geq y_j^* | W_{ij} = w)$ will be in the form of a linear model--

$$\Phi^{-1}P(Y_{ij} \geq y_j^* | W_{ij} = w) = \alpha_0 + \alpha_1 Z_{ij} + \alpha_2' w + \alpha_3' Z_{ij} w,$$

--where Z_{ij} is again an indicator function for membership in Group A, and the unknown coefficients have the following form:

$$\alpha_0 = \frac{\lambda_B' \mu_B^* - y_B^*}{\sigma(\lambda_B)}$$

$$\alpha_1 = \frac{-1}{\sigma(\lambda_B)} \lambda_B'(\mu_B^* - \mu_A^*) + \left(\frac{1}{\sigma(\lambda_A)} \lambda_A' - \frac{1}{\sigma(\lambda_B)} \lambda_B'\right) \mu_A^* - \left(\frac{y_A^*}{\sigma(\lambda_A)} - \frac{y_B^*}{\sigma(\lambda_B)}\right)$$

$$\alpha_2 = \frac{1}{\sigma(\lambda_B)} (\delta_B + B'\lambda_B)$$

$$\alpha_3 = \left(\frac{1}{\sigma(\lambda_A)}\delta_A - \frac{1}{\sigma(\lambda_B)}\delta_B\right) + B'\left(\frac{1}{\sigma(\lambda_A)}\lambda_A - \frac{1}{\sigma(\lambda_B)}\lambda_B\right)$$

In this model, the parameters α_1 and α_3 contain the evidence of discriminatory behavior, and are the center of our interest. Obtaining estimates of α_1 and α_3 using the methods of probit analysis will present real problems unless the values of W for the candidates are concentrated in not too small groups at just a few different points. However, even for the case where the values of W are spread thinly over many points a method described by Walker and Duncan[16] offers a method of estimating the parameters, provided that we are willing again to substitute the logit function for the inverse normal probability integral function.

Non-zero values in α_3 clearly indicate intentional discrimination in the form of differential recognition of qualifications in different groups. However, α_1 is again a sum reflecting both innocent and suspect influences and raising the same legal and methodological questions alluded to earlier.

In closing, we note that the complexity of having to consider legitimate qualification variables again results in substantial difficulty in defining a single summary measure to permit simple case-to-case comparison.

Footnotes

[1] Brooks v. Beto, Federal Reporter, 2[d] Series 366: 1 at p. 9 (1966).

[2] United States Reports 118: 356 (1886).

[3] For a recent example see Castaneda v. Partida, Supreme Court Reporter 97: 1272 (1977)

[4] Supreme Court Reporter 96: 2040 (1976)

[5] Supreme Court Reporter 97: 555 (1977)

[6] Supreme Court Reporter 97: 1272 at pages 1279-80. This distinction between the functions that evidence of under-representation on jury panels can fulfill had been drawn earlier in Swain v. Alabama (United States Reports 380: 202 at pp.208-9).

[7] One measure frequently suggested is the proportional shortfall in the representation rate, $(R_A - r_A)/R_A$. Another is the estimate of an excluded proportion discussed subsequently. See text at note 13.

[8] For example note comparisons made in Casteneda v. Partida (see note 3 above) at page 1281.

[9] This assumption is not critical for this analysis, for which each Group A candidate is assumed to have the same probability of selection. Eliminating this assumption would change the form but not the import of the argument.

[10] The persistence of u as a factor in $E^*(U)-E_A(U)$ raises a legitimate concern about the comparison of measures of impact for selection processes involving very different rewards. Thus the courts might well be more concerned with a seven percentage point difference between death sentencing rates applied to equally situated murder convicts than for a ten percentage point difference in hiring rates among equally qualified candidates. However, the appearance of u should not invalidate comparisons between results from situations in which the utility constants are similar.

[11] A discussion of hypothesis tests sensitive to under dispersed distributions of representation rates suggestive of quota limiting is found in M. O. Finkelstein, "The application of statistical decision theory to jury discrimination cases." Harvard Law Review 80: 338 (1966) at P. 365ff.

[12] J. Kirk in R. H. Amidon, et al., Mexican-Americans and Administration of Justice in the Southwest. (A report prepared for and issued through the U. S. Commission on Civil Rights) (1970) at pp. 132-3.

[13] The change in assumptions from fixing the number chosen to fixing the threshold of selection is made at some cost in credibility but with a large gain in simplicity of the argument.

[14] See, for example, D. J. Finney, Probit Analysis, Cambridge University Press (1964) at pp. 48-51.

[15] There are pitfalls in using p_B/p_A when both P_B and P_A are small as the value of the ratio will be very sensitive to sampling variability of p_A and to errors in the definition of the candidate pool for Group A. The first difficulty can be controlled by construction of an approximate confidence interval for P_B/P_A and the use of the most conservative value in the interval. The second requires close inspection of the assumptions and data gathering for the pre-selection pools.

[16] S. H. Walker and D. B. Duncan. "Estimation of the probability of an event as a function of several independent variables." Biometrika 54: 167 (1967).