# Getting it Straight When Statistics Can Lie

## JAMES P. SCANLAN

**Legal Times**
Week of June 28, 1993

Last November, in *Fisher v. Transco-Services-Milwaukee*, 979 F.2d 1239, the Seventh Circuit reversed a grant of summary judgment to the defendant in an age discrimination case challenging a computerized system for monitoring the rate at which warehouse workers filled orders.  In reinstating a claim that the system had an unjustified disparate impact on older workers, the appeals court was impressed by two things.  First, the performance standard seemed unreasonably high.  Second, the disparate impact appeared especially severe.  Noting that 10 of 27 workers aged 40 or above had been terminated compared with 1 of 25 workers under age 40, the court concluded:  "It does not take expertise in differential equations to observe that an adverse ratio of approximately 10 to 1 is disproportionate."

The court was correct that an understanding of differential equations was unnecessary in this case.  Yet, an understanding of certain more elementary statistical principles might have caused the court to see the case in a somewhat different light.

There are few things in mathematics that are at once so fundamental and so widely misunderstood as the relationship between disparities in rates of succeeding and disparities in rates of failing.  Anyone who thinks about it for a moment would probably understand that the greater the difference in the average performance of two groups, the greater will be disparities in the groups' success rates and the disparities in the groups' failure rates.

What is more difficult to understand is that when standards are raised or lowered, disparities in success rates vary inversely with disparities in failure rates.  That is, if the standard for successful performance is set quite high, disparities in success rates will be relatively great while disparities in failure rates will be relatively small.  By contrast, if the standard is set quite low, disparities in success rates will be relatively small while disparities in failure rates will be relatively great.

For example, assuming a substantial difference in average performance, if the cutoff is set where only 10 percent of the better-performing group succeeds, only 2 percent of the poorer-performing group might succeed.  This five-to-one disparity in success rates would seem quite large; yet the percentage difference in failure rates (90 percent compared with 98 percent) would not be deemed particularly great.  But if the cutoff is lowered to a point where 95 percent of the better-performing group succeeds, 80 percent of the poorer-performing group might succeed.  That disparity in success rates might not seem very great; but the disparity in failure rates (5 percent compared with 20 percent) would have risen to four-to-one.

In evaluating the impact of employment tests for selection, courts have typically looked at disparities in pass rates.  Lowering of a cutoff score has been universally regarded as a means of reducing the impact of a test because it diminishes the disparity in pass rates, even though lowering the cutoff also increases the disparity in failure rates.

The tendency for disparities in success rates and disparities in failure rates to vary inversely as cutoffs are raised or lowered will usually be observed wherever distribution of a characteristic are at all normally distributed.  As reflected in the *Fisher* case, however, in the evaluation of discipline or termination policies, disparities in failure rates are commonly examined with little recognition that lowering standards will increase disparities in failure rates.

Under the standards that had been applied by the employer in *Fisher*, the older workers' success rate (63 percent) was 66 percent of the success rate of the younger workers (96 percent), while the older workers failure rate (37 percent) was, as the court observed, close to ten times the failure rate of the younger workers (4 percent).  The expected tendency of a reducing the the standard, however, would in the long run be to reduce the disparity in success rates – for example, by raising the older workers' success rate to 80 percent while raising the success rate of younger workers to 99 percent, thus making the success rate for older workers 81 percent of the rate for younger workers.

But this would make the failure rate for older workers (20 percent) twenty times the failure rate for younger workers (one percent).  Easily, in fact, a reduction in the standard would result in the termination solely of older workers.  It is the raising of the standard that would reduce the disparity in failure rates, though at the expense of increasing the disparity in success rates.

In basing its decision on the view that the performance standards were very high and that the disparity in termination rates was very great, the Seventh Circuit failed to understand that a reduction of the standard to a level it might deem more reasonable would actually tend to increase the disparity between failure rates that it had found so disturbing.  Indeed, had the employer on its own chosen

to adopt a lower performance standard – had it done so even out of concern for its older workers' ability to meet an exacting standard – it would only have subjected itself to increased scrutiny under the Seventh Circuit's approach.

The misunderstanding of these tendencies is especially common in the analysis of racial disparities. Racial disparities in employee discipline rates are being increasingly studied. Unfortunately, they are studied with little appreciation of how the severity or leniency of the employer's discipline policies affects the observed disparities or of the likely correlation between high disparities in failure rates and low disparities in success rates.

Recently, the discipline practices of the Internal Revenue Service were subjected to intense scrutiny because of widely disparate rates at which blacks and whites were disciplined for workplace infractions. An extensive report was produced attributing the disparity largely to race-neutral factors, and recommending largely race-neutral approaches to address the situations causing the discipline problems. If such approaches are effective, they should reduce the racial disparity in avoiding discipline problems. But they may well increase further the racial disparities in discipline rates that attracted attention to the situation in the first place.

A recent study of racially disparate termination rates among Postal Service workers explored whether such disparities would be as great at a quasi-federal agency with an excellent reputation as a fair employer as they were in the private sector. Yet the authors failed to consider that the greater protections afforded public sector workers, by reducing overall termination rates, and hence reducing disparities in keeping one's job, lead to greater disparities in losing one's job.

Ironically, whether it be the protection s typically offered in civil service environments or those provided by union grievance procedures, the same measure that by eliminating arbitrariness in discipline policies reduce racial discrimination tend also to increase any racial disparities in discipline rates that are not the result of discrimination. Those measures also diminish disparities in avoiding discipline, but usually that goes unnoticed.

There are numerous other areas where a lack of understanding of this tendency has led to a mistaken focus solely on rates of failure. Consider the recent attention to racial disparities in home lending. A number of studies have shown that minorities seeking mortgages are rejected much more often than whites, even when efforts have been made to control for income level of the applicants. Last summer the Office of the Comptroller of the Currency indicated that banks where minority rejection rates are twice those for whites would face intense scrutiny.

It is hard to know the extent to which these disparities actually reflect discrimination in the sense that banks are treating identical loan applicants differently because of race. In any broad income grouping, the poorer group will tend to be disproportionately concentrated at the lower end of the grouping. Moreover, the disparity in wealth between minorities and whites earning the same income is well-documented. So one would expect these studies to reveal some racial disparities in bank lending practices whether or not similarly situated people are treated differently because of their race. But in measuring the size of such disparities it would seem a mistake to look at rejection rates.

A bank's lending practices operate just like tests. At banks with relatively lenient lending criteria, the black approval rate will be closer to the white approval rate than at other banks, while the disparity in rejection rates will be greater than at other banks. Thus, so long as the focus is on disparities in rejection rates, the banks whose credit "tests" would be deemed to have the least discriminatory effect, as that concept usually is understood, instead will be deemed to have the most discriminatory practices. Moreover, the banks deemed to have the most discriminatory practices will tend to be those at which blacks, like whites, are most likely to get loans.

The tendency was illustrated by the data first released by the Federal Reserve in October 1991, which were divided according to four income groupings. Among applicants for conventional mortgages, the higher the income grouping, the greater was the overall acceptance rate, and the closer was the black acceptance rate to the white acceptance rate; but, the higher the income, the greater was the racial disparity in rejection rates.

To be sure, this will not always happen. Other factors, including such discrimination as a particular institution actually may engage in, often outweigh the statistical tendencies. But the statistical tendencies are essential parts of the picture, and without understanding them, no one can make heads or tails out of the data.

Even in the selection context difficult issues can arise as to how to measure a disparity. The federal government's Uniform Guidelines on Employee Selection Procedures speak generally of a "four-fifths rule" whereby federal enforcement action usually will be limited to situations where one group's rate of satisfying a selection criterion is less than 80 percent of the rate of another group. Relying on the Guidelines, many courts also have applied the four fifths rule to limit disparate impact claims to situations where the impact can be deemed serious.

But interpretations accompanying the Guidelines have noted a particular exception to the focus on selection rates in the case of the use of arrest or conviction records as disqualifying criteria. Since usually a large enough majority of members of all races will satisfy the requirement of having no arrest or conviction record, such policies do not often violate the four-fifths rule. In such cases, the interpretations have stated, the appropriate focus is upon disparities in disqualification rates. If difficult to justify logically, the approach seems entirely sensible.

A problem arises however, when one seeks a less discriminatory alternative to a policy barring hire of persons

with arrest or conviction records.  The obvious less discriminatory alternative to a rule barring hire of persons with arrest records is a rule barring hire of persons with conviction records, and the obvious alternative to a rule barring hire of persons with any arrest or conviction record is a rule barring hire of persons only with arrests or convictions for serious crimes.  But the probable tendency of these alternatives is to reduce the disparate impact only as measured in terms of rates of meeting the requirement of not having such a record; they are likely to increase the disparity in the rates at which members of two groups are disqualified, which the interpretation indicates ought to be the focus.

Moreover, like banks with very liberal lending policies and warehouses with low performance standards, employers who voluntarily limit their disqualifying criteria – for example, to convictions for serious crimes – could find themselves more likely candidates for such claims.

There is no obvious solution to the arrest/conviction problem.  But, as with each of the other situations described above, it can be better addressed with a full understanding of the seemingly paradoxical relationship between disparities in success and disparities in failure.