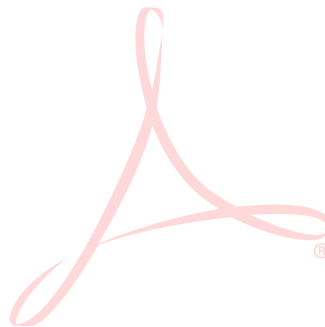


**LIES, DAMN LIES AND STATISTICS:  
A VIEW FROM A STATISTICAL EXPERT**

**DWIGHT STEWARD, Ph.D.  
M. DOUGLAS BERG, Ph.D.**



**State Bar of Texas  
ADVANCED EMPLOYMENT LAW  
February 20-21, 2003  
Houston, Texas**

**CHAPTER 13.1**

After completing high school in San Antonio, Texas in 1986, Dr. Steward obtained a B.A. in economics from the University of Texas at Austin. After brief service as a U.S. Army Field Artillery Officer during Operation Desert Storm, he obtained a Ph.D. in economics from the University of Iowa with field specializations in labor economics, applied econometrics (statistics), and banking in 1995. During graduate school, Dr. Steward had the honor of receiving a full Federal Reserve System fellowship and serving as a resident dissertation scholar at the Federal Reserve Bank of Atlanta.

After graduate school, Dr. Steward was employed as an economist and statistician at Welch Consulting in College Station, Texas from 1995 to 1997. Welch Consulting is an industry leading firm that specializes in the statistical analysis of discrimination issues. In 1997, Dr. Steward left the firm and began teaching at the University of Texas at Austin (UT) and founded Steward Research Group. Steward Research Group is a firm that provides economic and statistical research on employment and discrimination issues to the legal, business and governmental communities.

As independent economist, has provided research in dozens of issues including employment discrimination, racial profiling, and economic damage calculations in major legal cases involving catastrophic injuries and death. Dr. Steward has also participated in a number of public forums and testified in state court, federal court, and before the Texas State Legislature on these issues. As a Senior Lecturer in the McCombs School of Business and the Department of Economics at UT from 1997 to 2002, Dr. Steward taught numerous courses in upper division statistics, microeconomics, financial institutions, and corporate finance. In 2002, Dr. Steward was honored by UT's largest student professional business organization as teacher of the year.

## LIES, DAMN LIES AND STATISTICS: A VIEW FROM A STATISTICAL EXPERT

### I. FORM HYPOTHESIS CONCERNING STATISTICAL EMPLOYMENT DISCRIMINATION

Generally, the statistical expert will conduct statistical tests that address one or both of the following employment discrimination questions: (1) if there was no discrimination against the protected group members, what would be the probability of observing the employment related disparity by random chance alone? (2) after accounting for relevant employment factors, does the individual's protected class status remain a statistically important factor?

The first question is typically used when assessing the statistical importance of an overall difference in selection rates or salary by protected group status. For instance, if it is observed at a particular company that on average African-American employees earn less than White employees, then the statistical expert will use generally accepted principles to calculate the probability that sheer random chance generated the observed racial salary differential. The probability that is calculated from the expert's statistical test is commonly referred to as a **chance probability**.

In the above example, a small chance probability, for instance a 1% chance, would mean that if there was no discrimination against African-American employees, then we would only expect to see the observed racial salary differentials in 1 out of every 100 instances. Conversely, if the calculated chance probability was large, say 50%, then we would expect to see the observed probability in 1 out of every 2 instances. A chance probability that is small and equal to or less than a pre-specified threshold is referred to as **statistically significant**. In most studies, an employment disparity finding that has less than a 1% chance of occurring by sheer random is viewed as statistically significant. The actual methodology used to calculate the chance probability depends on the type of employment question being addressed by the statistical expert.

The second question is typically used by the statistical expert when assessing the importance of the employee's protected group status in the selection or pay process when there are observable employment related differences among the individuals being analyzed. For instance, it is generally recognized that an individual employee's salary will tend to be related to factors such as the type of job performed, seniority, education, as well as company specific factors such as time in salary grade. In these analyses, the expert will develop statistical models, such as **linear regression** models, to account for the effect that these employment factors and the protected group status collectively have on the selection or compensation process. In these types of models, a finding that the individual's protected group member status remains a statistically

important factor even after accounting for the observable employment related differences among the individuals, is typically viewed as suggestive of a non-neutral employment process.

It is important for the attorneys on both sides of the case to understand the limitations of the employment analyses performed by the statistical expert. These limitations are not the result of the statistical expert, but are result of the statistical foundations upon which the employment analyses are based.

A. The statistical expert's *a priori* assumptions, or **null hypothesis**, concerning the defendant's employment processes must be that the employment processes being studied are non-discriminatory. This is because statistical employment analyses are founded on statistical principles that rely on the simple calculation of random chance events. The random chance calculation that is performed in employment analyses is in principle similar to the calculation of finding the probability of 'drawing a red ball from an urn filled with red and green balls' or finding the probability of 'getting 7 heads when flipping a fair coin 10 times'.

B. A statistical expert can never state that the statistical test that she or he performed is statistical proof of employment discrimination. As discussed above, the statistical test used by the expert only calculates the probability of that sheer random chance generated the observed employment disparity.

C. Only correlation, not causation, can be inferred from the statistical expert's employment analysis. For instance, if after accounting for relevant employment related factors, an employee's gender remains a statistically important factor in a salary regression model, the principles that underlie regression techniques, only measure the degree of correlation between the employee's salary and gender across the individuals being studied. The statistical models are not designed to determine if the person's gender is the reason for the observed disparity in pay.

### II. COLLECT RELEVANT EMPLOYER DATA AND INFORMATION

Regardless of who the attorney is representing, it is important to provide the statistical expert with employment information that completely describes the employment processes that are being studied. Generally this will involve providing the statistical expert not only the hardcopy or electronic employment data that describes the statistical disparity, but in addition, the background information on the employer and the employer's practices. Typically, the

background information will include information from sources such as electronic databases provided by the defendant, employer handbooks, written descriptions of the relevant selection processes, and depositions of human resource personnel and other key decision makers. If at all possible, engaging the statistical expert early in the discovery process typically will allow the expert to more adequately prepare a listing of the specific information needed in the employment analysis.

In most employment analyses the statistical expert will at a minimum require the following information about the defendant's employment processes.

A. Employee level information. The employee level information required by the statistical expert includes not only the demographic and employment information, such as date of hire, salary grade, etc., for the plaintiff but also for all the employees in the organization being analyzed. This information will allow the statistical expert to construct comparison pools of 'similarly situated' employees.

B. Employer practices information. This type of information includes information about the factors that are incorporated into the employer's selection or compensation processes. For instance, in cases involving discrimination in employee terminations, it is important to determine the specific formula or individual factors that were considered by the employer in the relevant reduction in force action.

C. Company specific information. This information generally includes company specific factors that describe the organizational differences between different divisions within the relevant analysis unit. For example, in some companies that closely tie financial performance to employee salaries, it is not uncommon to observe higher average pay levels in division that generate higher levels of revenue for the company. For larger organizations, the pay or salary grade structure for the company is also important information for the statistical expert.

**III. GENERAL CONSIDERATIONS IN STATISTICAL EMPLOYMENT ANALYSES**

There are several issues that should be considered by the attorney when reviewing the employment analyses performed by the statistical expert. First, although the issues faced by the statistical expert performing an analysis in an employment discrimination lawsuit will be unique in a number of ways, the expert's analysis should be of the same quality as non-litigation related research. In addition the statistical expert's analysis should be grounded by sound information about the defendant's employment practices, relevant professional and academic studies, and professional common-sense. For instance, as discussed above the statistical expert typically uses

proprietary employer level data to analyze the employment practices of a particular organization. In contrast, published academic labor and employment research studies tend to use publicly available labor market level data to study the issues related to the employment practices of an entire labor market. Although the level of analysis is clearly different, the methodologies and reasoning that underlie the labor market studies are frequently relevant and should generally be used in the construction of employer specific models of compensation and employee selection processes.

Second, ultimately, the expert's employment analysis should be reported in a way that allows the research to be verified by another statistical expert. In practice, this involves ensuring that the necessary information to reproduce the exact data used by the expert in his or her analysis is provided. Additionally, this requires supplying information that will allow the analysis including the results to be replicated by the opposing statistical expert.

**IV. SPECIFIC CONCERNS FROM A STATISTICAL EXPERT'S VIEWPOINT**

**A. Discriminatory Hiring Allegations**

Statistical analyses involving hiring allegations tend to present a number of special issues that should be considered by attorneys on both sides of the employment discrimination case. First, it is not uncommon for organizations to keep only cursory records on individuals who apply for a given job position. Consequently, special attention must be dedicated to the construction of the relevant applicant pool in these types of employment cases. In some instances in hiring analyses, the employer level hiring data is augmented with labor market availability information from government and other publicly available sources to provide a more probative analysis.

Second, in addition to the data issues in hiring cases, the definition of the relevant pool of qualified applicants is typically an issue of concern for the statistical expert. Generally speaking, it is important for the expert to ensure that the hiring analysis compares individuals who, but for their protected class status, would have a comparable probability of being hired. For example, in a case involving allegations of gender discrimination in faculty hiring at a university, factors such as the individual's prior research and teaching experience would be factors to potentially consider in the construction of the qualified applicant pool.

Third, it is important to consider each of the stages that comprise a typical employer's hiring process as well as the different factors that may effect the individual's decisions at each stage of the hiring process. For example, many employers' hiring processes are comprised of distinct stages that may include the application stage, interview stage, testing stage, and ultimately the selection stage. It is important

to clearly define the hiring stages, as well as to determine the number of applicants in both the protected and non-protected group who are not only qualified but are truly interested, potential job candidates for the given position.

### **B. Discrimination In Pay Allegations**

In contrast to cases involving hiring discrimination, employment cases involving discriminatory salary allegations tend to present relatively few special data issues for the statistical expert. Generally, employers tend to keep data records on compensation issues that are sufficient for most statistical employment analyses. In these types of cases, the statistical expert will generally be most concerned with the specification of the underlying compensation process. For example, in many of the cases both the plaintiff and defendant statistical experts will utilize the same employer salary and generally use the same class of statistical methodologies. The differences between the opposing statistical experts opinions frequently revolves around the specification of the compensation model and the employment actors included in the model.

### **C. Discrimination In Employment Termination**

Similar to cases involving discriminatory pay allegations, discriminatory employment termination type cases present relatively few data related issues for the statistical expert. Like pay cases, most of the differences revolve around the specification of the statistical model of the employer's practices. However, determining which protected and non-protected group members were actually eligible for a given reduction in force action does tend to produce differing opinions on the neutrality of the employer's practices.