

Measures of Discrimination and Protocols for Evidence: The Case of the Spare List

Steven M. Crunk¹, William B. Fairley²

¹Department of Mathematics, San Jose State University, San Jose, CA 95192-0103

²Analysis & Inference, Inc., 1489 Baltimore Pike, Suite 305, Springfield, PA 19064

Abstract

The selection of measures of discrimination is a critical step in determining if there is legal evidence to establish discrimination. In a case that went to trial, an individual alleged his minority status was a factor in assigning him to a less desirable job on a city police force. Each expert testifying on the statistics used different measures. The different measures resulted in sharp differences in the characterization of discrimination. Grounds for selection of one measure over another are treated. These grounds include the stage of the case in legal proceedings (e.g. “prima facie” versus “business necessity” phases), the kind of discrimination alleged, and the interest attaching to frequency vis a vis severity. Further, we discuss the importance of the consulting statistician, and argue that their status as independent professionals in the judicial process is at risk.

Key Words: discrimination; measure; legal evidence; protocol; selection; expert; permutation test

1. Introduction

In Garcia v. Springfield Police Department the Plaintiff, an officer in the Springfield Police Department, alleges that his ethnicity (Hispanic) influenced the assignments he was given as an officer. Specifically, he states that Hispanic officers were more often assigned to duty on the “Spare List”, which is used to substitute for other officers on vacation, sick, special events, etc. We understand that it is generally, though not always, viewed as a less desirable assignment than others. This assignment was alleged to be given to Hispanic officers because they were Hispanic.

The attorneys for the plaintiff engaged a professional statistician to investigate possible discrimination in the case. In response, attorneys for the defendant, the city of Springfield, Massachusetts (“city”), engaged two statisticians through the firm Analysis & Inference, Inc. (A&I), to critique the statistical evidence presented by the plaintiff. Both authors worked on this analysis on behalf of A&I. References to A&I may refer to the statisticians or to the company itself, as per context.

We used the summary data given as an appendix to the report presented by the statistician on behalf of the plaintiff, which plaintiff’s attorney provided in the form of a Xerox copy.¹

The period of time covered in the data was the three years January 1, 2002 through December 31, 2004. Plaintiff Garcia filed a complaint in early 2004 raising the issue of discriminatory work assignment. The data included a list of all officers employed by the city at any time during this period, their ethnicity (African-Americans were not included), the number of days on the roster, as well as the number of days on the Spare List for each of the three years. The statistician on behalf of the plaintiff stated that, insofar as the suit might have changed the practice of the police department, the years 2002-2003 are the more relevant period.

¹ Certain of the values were not clearly legible from the copy, so they may differ in a few places from the values used by the statistician on behalf of the plaintiff. The differences are expected to have been small, and affect the results negligibly.

2. Measures of Discrimination by the Plaintiff's Statistician

The statistician on behalf of the plaintiff provided three measures of discrimination, which we describe individually below.²

2.1 Comparison of Overall Incidence Rates

The first measure is the ratio of the incidence on the Spare List for Hispanic versus Caucasian. Taking each year separately, the statistician on behalf of the plaintiff classified each police officer as having been on the Spare List or not during that year, and also by ethnicity (only taking into consideration male Caucasian and Hispanic officers). As an example, this summary for the year 2002 is presented in Table 1 below. Tables given in this paper are similar in form, but not always identical to those given in reports by statisticians on behalf of the plaintiff and defendant. The numbers are the same. Plaintiff's statistician stated in his report that, for the data in Table 1, the rate at which Hispanics are on the Spare List is 1.85 higher than for Caucasians, which he calculates $76.5/41.3 = 1.85$. We note that the odds ratio for the table is 2.98.

Table 1: Breakdown of Spare List inclusion by ethnicity during 2002

	On Spare List at least once?		Total	Percent Yes
	No	Yes		
Caucasian	27	19	46	41.3
Hispanic	4	13	17	76.5
Total	31	32	63	

We calculated ordinary Pearson's Chi-squared test with Yates' continuity correction on the data in Table 1 provides a p-value of 0.1277, indicating that the higher proportion of Hispanics that were on the Spare List at least once during 2002 is not statistically significant at the 0.05 level (Fisher's exact test gives a p-value of 0.09518). The same test for the year 2003 shows significance, as it does also for the combined data for 2002-2003. The statistician on behalf of the plaintiff performed a permutation test for the years 2002-3 combined, and reported a p-value of 0.0060.

The statistician on behalf of the plaintiff noted that this method has a shortcoming: It does not distinguish between an officer who is only on the Spare List for a day, and one who is on the Spare List many days in the year. Another shortcoming that we observed is that officers differ in how frequently they are at work at all, i.e. "were "on the roster," thereby having different opportunities to be spare-listed.³

2.2 Comparison of the Average Number of Days on the Spare List

The second measure is the ratio of the average number of days on the Spare List between Hispanic and Caucasian. The statistician on behalf of the plaintiff takes the total number of days on the Spare List for all Hispanics and divides by the number of Hispanic officers, and identifies this as the average number of days on the Spare List for Hispanic officers, and similarly for Caucasian officers. He provided a summary table as in Table 2.

He notes that the average number of days on the Spare List for Hispanics is some 2.6 times larger than for Caucasians. He reports a permutation test p-value for the 2002-3 years combined of 0.0011. He notes that this analysis has the shortcoming that it does not take account the number of days an officer is on the roster, e.g., two officers who were on the Spare List for a total of 20 days during the year, one of whom was on the roster for 200 days and the other who was on the roster for 30 days, would each contribute equally to this analysis.

² We use the term "discrimination" loosely here, because a finding of discrimination in the law is a legal finding. Strictly speaking, statisticians are making a legal claim if they "find" or even "measure" discrimination. Statistics is relevant to such a legal claim, but there are other criteria than statistics of differences between groups to meet before a finding of discrimination is made. Thus, our use of "discrimination" here must be understood only as a shorthand way of referring to the subject at hand.

³ There are potential reasons why less time on the roster could lead to a different chance of being spare-listed, e.g., if available fewer days, may be placed on the Spare List more often; or out so often they do not have a regular position. Such differences were not taken into account in any of the analyses, though they could be.

Table 2: Average number of days on the Spare List by ethnicity during 2002

	Number of officers	Total days on Spare List	Average
Caucasian	46	1821	39.6
Hispanic	17	1741	102.4

2.3 Comparison of the Percent of Time on the Spare List

The third measure is ratio of the percentage of their work time spent on the Spare List for Hispanic vis a vis Caucasian. A table like Table 3 below was given and he concludes “Thus by this measure the average Hispanic was 2.3 times as likely to be assigned to the Spare List” (where $2.3 \approx 30.8/13.4$).

Table 3: Percent days on the Spare List by ethnicity during 2002

	Total days on roster	Total days on Spare List	Percent days on Spare List
Caucasian	13,638	1,821	13.4
Hispanic	5,649	1,741	30.8

However, as he pointed out, “All days get weighted equally. Thus a person observed for 1000 days gets weighted ten times as much as a person with only 100 days.”

Because the unit of interest is the individual officer, we do not believe that the measure “percent days on the Spare List” by ethnicity” measures the difference between individual Hispanic officers and individual Caucasian officers. By dividing both the numerator and the denominator of this measure for each group separately by the number of officers in the group on the roster, the new numerator becomes the second measure as described in section 2.2, already shown to be unsatisfactory, while the denominator becomes the average number of days officers in the group spend on the roster. Thus, the difference between the groups depends on the irrelevant ratio of average days on the roster. No claim was made that Hispanic officers were discriminated against in the number of days on the roster. In particular, a group like Hispanic officers, with somewhat fewer average days on the roster than Caucasians, will thereby be shown to have a higher value of the measure in comparison to Caucasian officers.

3. Alternative Measures of Discrimination

We gave two alternative measures of discrimination in our report. We discuss each on in turn.

3.1 Continued Presence on the Spare List

An analysis such as that presented in Table 4 below shows the propensity to stay on the Spare List for Hispanic and Caucasian combined if they once were, and to remain off it if they never were. Of those on the Spare List at some time during 2002, 60 percent were on the Spare List at some point during 2003 and 50 percent were on the Spare List at some point during 2004. Similarly, of those not on the Spare List during 2002, only 20 percent were ever on the Spare List during 2003 and only 15 percent were on the Spare List in 2004 and, similarly, are not likely to remain. That is, once an officer is placed on the Spare List, for whatever reason, they are likely to remain on the Spare List during succeeding years.

Table 4: Percent of officers on, and not on, the Spare List in 2002 and on the Spare List in 2003 and 2004

2002 status re Spare List	Percent days on Spare List in	
	2003	2004
On	60	50
Not on	20	15

The first alternative measure of discrimination we proposed is the ratio, for Hispanic to Caucasian officers, of percent days on the Spare List given in 2003, and then again in 2004, for those who were on it in 2002. Table 5 gives these percents by ethnicity, and they are close to the same for Hispanics *vis a vis* Caucasians. A test of no difference in proportion of time on the Spare List between Hispanics and Caucasians during 2002 returns a p-value of 0.48, while the same test for 2003 returns a p-value of 0.41, and so the differences, small as they are, are also not statistically significant.

Table 5: Percent of days on the Spare List in each of 2002 and 2003, by ethnicity, among those who were on the Spare List in 2002

<u>Percent of days on Spare List</u>		
<u>Year</u>	<u>Hispanic</u>	<u>Caucasian</u>
2002	60	61
2003	45	42

The closeness of these percents provides a different perspective on the differences between Hispanic and Caucasian given in Tables 2 and 3. The emphasis now must turn to an officer's initial placement on the Spare List, and if such placement is due to discrimination. The statistician on behalf of the plaintiff did not reach this question.

3.2 Comparison of Percent Time on the Spare List per Individual Officer

The measure proposed by the statistician on behalf of the plaintiff in section 2.3 above uses the total time on the spare list of the given ethnic group divided by the total number of days on the roster for that ethnic group. Although this could be called "percent of time on the spare list", it is not the percent of time for officers, or per officer, which in our view it should be. A second alternative measure of discrimination we proposed is the percent time that officers spend on the spare list is the ratio of the simple average of the times (days) Hispanic officers spend on the spare list to the simple average of the time Caucasian officers do.. Such a measure is influenced by both the frequency (number of officers ever on the Spare List) and severity of officers' appearances on the Spare List (how long they are on the Spare List). It takes into account differences in the lengths of time that individual officers spend on the Spare List. Finally, the unit of analysis in this measure is the individual officer. We see this as the most appropriate measure since discrimination is something that occurs, or does not occur, to individuals. Table 6 gives a summary of the data. The ratios are not small. The ratios of percent for Hispanic *vis a vis* Caucasian in year 2002 is statistically significant at the 0.05 level, and not for 2003 and 2004. The p-values given by a difference of means test are 0.0146, 0.1211, and 0.1337 for the years 2002, 2003, and 2004 respectively.

Table 6: Average percent of days on the Spare List per officer, by ethnicity

<u>Percent days on Spare List</u>		
<u>Year</u>	<u>Hispanic</u>	<u>Caucasian</u>
2002	19.2	12.5
2003	28.6	17.1
2004	29.0	16.4

The results in Table 6 show on their face differences in ethnicity, in contrast to those in Table 5, which do not. There is no contradiction. The measures are different. In particular the data in Table 5 include all officers whether they were ever on the Spare List or not, whereas the data in Table 5 exclude officers who were not on the Spare List in 2002.

4. Further Thoughts on This Case

Other analyses were performed by the statisticians for each side, and other differences of opinion on inclusion of data e.g., the inclusion of only Hispanic and Caucasian officers in the analysis, and the inclusion of only males by the statistician on behalf of the plaintiff were discussed in our report.

It is worth noting that the permutation tests used throughout by the statistician on behalf of the plaintiff is, at least for the first analyses, simply an approximation of Fisher's exact test. Gastwirth (1988, page 224) observes that the model assumed (in the Report this is the simple random sample model) cannot be used until other factors that could influence differences between groups are checked. Gastwirth (1988, page 318) further emphasizes that the "randomization assumption", which is the application of the simple random sample model or the permutations test, requires that other factors be the same. That is, for example, that chance that a Caucasian is placed on the Spare List should be a constant value for all Caucasians, and similarly for Hispanics. However, there could be many other factors which could influence the probability of being on the Spare List, e.g., desire to be on the list, length of employment, training status, and so on.

One could ask whether the analyses of the statistician on behalf of the plaintiff and those on behalf of the defendant would fare under what is called a "Daubert challenge." Daubert refers to the case before the U.S. Supreme Court, Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993) 509 U.S. 579, 589, The Daubert case for the first time established for federal judges a duty to ascertain for themselves that certain basic indicia of scientific method had been met by the experts testifying in a trial.

We believe both sides presented professional analyses of the data. They simply had a difference of opinion on the appropriate analyses. There were no Daubert challenges in the present case, and in our view neither challenge, if made, should have been successful. We do note that a prominent professor of the law of evidence (Berger (2005)) states that Daubert challenges are often misused and abused by lawyers and even by the courts. The issue of the American Journal of Public Health, in which this article was included, treats presentation of scientific evidence in the courtroom, with special attention to Daubert.

We now turn to a discussion of the profession of statistical consulting in the legal setting

5. On Being a Professional Statistician in the Legal Setting

5.1 Ethics

In a recent column in AmStat News (Lachenbruch, Stablein, and Wittes (2008)), the authors state that many doubt the integrity of a paid professional statistician. They reference an article in the Journal of the American Medical Association (JAMA) which as one of its conclusions states that "All journals must require a statistical analysis of clinical trial data conducted by a statistician who is not an employee of a for-profit company". Lachenbruch, et al, contend, on the contrary, that if this were true for pharmaceutical studies, it would also be true of all statistical reports whenever an individual or organization hiring a statistician has a financial interest in the statistical results. They continue by outlining why the recommendations in the JAMA article are "unworkable".

In our view, also, and setting aside the obvious ethical aspect, it is typically in the statistician's best interest to be honest in a paid statistical analysis if his work is to continue to be valued. In the legal setting, both sides may hire statisticians. If a statistician is found to be unethical either in the current case or in past work, there is a substantial risk that this will be exposed in court. This result will damage the case of the people on whose behalf they worked. Further, they may find it difficult to ever finding future employment in the legal field.

5.2 Timing of Statistical Work

We have found that attorneys, like many others, do not understand how much more useful statisticians can be when they are involved in a case at a much earlier stage. In particular, in a legal case a statistician can be invaluable in helping to formulate -- or respond to -- discovery. The expertise of a statistician can pay for itself many times over in the greater value of the information obtained, in how it is used, and in efficiency.

5.3 Data Availability and Format

All parties and the court will save time if statisticians on both sides work with a common set of data, and if the format is the same.⁴ We are reminded of another case in which we, as statisticians on behalf of the plaintiff, were provided

⁴ Michael O. Finkelstein has proposed a series of protocols for courts to follow that would require stipulation by the parties to common data sets. See Finkelstein, Quantitative methods in law (1978).

with approximately 20 reams of paper printed in what looked to be approximately 5 point font. The statistician on behalf of the defense included analysis performed using an electronic data set provided to him.

6. Conclusion

In order to maintain our status as professional independent statisticians in the support of the legal profession, we need to make sure that the lawyers understand what it is that we expect of them, and that they treat us as professionals. We need to communicate with them effectively. We need to publish regularly in the statistical and legal journals, so that lawyers and judges understand what it is that we do as professional, ethical, statisticians.

Oh yes, in the case of the Spare List, the jury returned a verdict in favor of the defense. As is usually the case, there was little information from the lawyers, the plaintiff, the jurors, or the judge to indicate the influence the statisticians on behalf of either side had in the jury coming to its conclusion.

References

- Berger, M.A. (2005), “What has a decade of Daubert Wrought?”, *American Journal of Public Health*, Vol 95, No S1.
- Box, G.E.P, and Draper, N.R. (1987), *Empirical Model-Building and Response Surfaces*, Wiley.
- Federal Judicial Center, (2000), *Reference Manual on Scientific Evidence, Second Edition*,
[http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/\\$file/sciman00.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/$file/sciman00.pdf)
- Finkelstein, M. O. (1978) *Quantitative Methods in Law*, Free Press.
- Gastwirth, J.L. (1988), *Statistical Reasoning in Law and Public Policy*, Vol. 1, Academic Press Inc.
- Lachenbruch, P.A., Stablein, D., and Wittes, J. (2008), “Too Quick to Impugn: JAMA Editorialists Doubt Integrity of Statisticians Working in For-Profit World”, *AmStat News*, July 2008.