Journal of Criminal Law and Criminology

Volume 87
Issue 2 *Winter*

Article 6

Winter 1997

Statistical Aspects of Cases Concerning Racial Discrimination in Drug Sentencing: Stephens v. State and U.S. v. Armstrong

Joseph L. Gastwirth

Tapan K. Nayak

0091-4169/96/8702-0583
THE JOURNAL OF CRIMINAL LAW & CRIMINOLOGY
Copyright © 1997 by Northwestern University, School of Law
Vol. 87, No. 2
Printed in U.S.A.

CRIMINOLOGY

STATISTICAL ASPECTS OF CASES CONCERNING RACIAL DISCRIMINATION IN DRUG SENTENCING: *STEPHENS V. STATE* AND *U.S. V. ARMSTRONG*

JOSEPH L. GASTWIRTH*
TAPAN K. NAYAK**

I. INTRODUCTION

Statistical evidence has been accepted in a wide variety of legal cases, including trademark confusion,[1] product liability,[2] and jury and employment discrimination.[3] Recently, statistics introduced on behalf of defendants who have challenged the fairness of sentencing practices have been poorly received by courts, as compared to their acceptability in equal pay and related fair employment cases. In this article we argue that the usefulness of aggregating the results of many prosecutorial decisions into a statistical analysis may not be adequately appreciated by the legal system. This lack of appreciation may be partially due to the fact that rarely is any data set perfect; one can almost always assert that information about some potentially relevant variable

* Professor of Statistics and Economics, George Washington University. It is a pleasure for the authors to thank Professor Robert J. Cottrol of The George Washington University Law School for many helpful comments and suggestions, and John J. Murphy for his editorial assistance. This research was supported in part by a grant from the National Science Foundation. The authors wish to point out that the views and opinions expressed herein are solely their own, and not those of the Foundation.

** Associate Professor of Statistics, George Washington University.

1 *See* Qualitex Co. v. Jacobsen Products Co., 115 S. Ct. 1300 (1995); Indianapolis Colts, Inc. v. Metropolitan Baltimore Football Club, 34 F.3d 410 (7th Cir. 1994).

2 *See* Glasser v. A.H. Robins Co., 950 F.2d 147, 148 (4th Cir. 1991); Kane v. Johns-Mansville Corp., 843 F.2d 636, 639 (2d Cir. 1988).

3 *See generally* Palmer v. Shultz, 815 F.2d 84 (D.C. Cir. 1987) (discussing the use of statistical evidence in employment discrimination cases).

is lacking or that the sample is not sufficiently large. In the context of two recent cases involving claims of racial disparities in drug offense sentencing[4] we illustrate how a fuller use of statistical techniques applied to the data might have strengthened the claims of the defendants.[5]

II. Overview of the Use of Statistics in Discrimination Cases

Ever since the United States Supreme Court adopted statistical testing in the *Castenada v. Partida*[6] jury discrimination case, statistical evidence has been used in many types of cases concerning discrimination regarding race, age, or sex.[7] When a plaintiff in a civil case claims disparate treatment, or a defendant in a criminal case introduces statistical evidence as part of their *prima facie* case, the purpose is to show that otherwise comparable individuals of the protected class are being treated less favorably than those from the majority group.

In *Texas Dep't. of Community Affairs v. Burdine*,[8] the Supreme Court outlined the order of proof needed to show disparate treatment in an employment discrimination case.[9] In *Batson v. Kentucky*,[10] the Court adapted this standard to criminal cases in which the defendant claims that the prosecution abused its discretion in making peremptory challenges at trial by eliminating a disproportionate number of minority venirepersons.[11] Once a defendant raises a *Batson* claim, presumably based on the fact that most if not all minority venirepersons were eliminated, the prosecution must advance a neutral explanation for their removal. Then the defendant is given the opportunity to

---

4 U.S. v. Armstrong, 116 S. Ct. 1480 (1996); Stephens v. State, 456 S.E.2d 560 (Ga. 1995).

5 Careful statistical analysis can also support the prosecution's case. In *U.S. v. Olvis*, 97 F.3d 739 (4th Cir. 1996), the Fourth Circuit interpreted the *Armstrong* decision's term "similarly situated" for purposes of selective prosecution cases as meaning that the circumstances surrounding their cases present no distinguishable legitimate factors justifying different prosecutorial decisions. *Id.* at 744. The defendants were part of a crack cocaine conspiracy and the government indicted twenty-five people, all of whom were black. There were a total of eighty participants; five were white and the rest black. The probability that no whites would be selected if twenty-five individuals were chosen from this pool of eighty is .1447, noticeably larger than the usual .05 level corresponding to the two-standard deviation criteria. See *infra* note 12 for an explanation. Thus, the data does not indicate that black conspirators were more likely to be prosecuted than whites by any statistically significant degree.

6 430 U.S. 482 (1977).

7 *See* Doan v. Seagate Technology, 82 F.3d 974 (10th Cir. 1996) (age discrimination); Melendez v. Illinois Bell Telephone Co., 79 F.3d 661 (7th Cir. 1996) (race discrimination); Palmer v. Schultz, 815 F.2d 84 (D.C. Cir. 1987) (sex discrimination).

8 450 U.S. 248 (1981).

9 *Id.* at 252-56.

10 476 U.S. 79 (1986).

11 *Id.* at 93-96.

demonstrate that the state's reasons are insufficient or pretextual. The trial court then must determine whether the defendant has established purposeful discrimination. This procedure is sensible, as only the prosecution possesses the information that it uses to decide which venirepersons to challenge.

In the typical jury discrimination case, the plaintiff uses data for a number of venires to compare the number of minority jurors to its expected number, derived from the minority fraction of persons eligible for jury service. A plaintiff in a hiring discrimination case would do much the same. This technique has been called "standard deviation" analysis, as the difference between the actual number of minorities and the numbers expected if selections were randomly chosen from the eligible pool is measured in units of the sampling variability, called standard deviation.[12]

In cases concerning the discriminatory use of peremptory challenges, *Batson* allowed the defendant to establish a prima facie case based solely on evidence concerning the prosecution's exercise of its peremptory challenges in his own case.[13] The Court noted that the peremptory challenge process enables those who have a mind to discriminate to do so.[14] The decision in *Batson* overruled *Swain v. Alabama*,[15] which required the defendant to demonstrate a pattern of discriminatory peremptory strikes in a reasonable number of similar cases. In *Batson* the Court noted that this was an extremely difficult

---

12 The standard deviation method was described by the Court in *Castenada v. Partida*, 430 U.S. 482, 496 n.17 (1977). When n people are hired from a qualified labor pool and p is the minority fraction of that pool, then one expects np of the hires to be minority members. If one assumes the hires are randomly selected from the pool, the "sampling variability" of the number of minority hires is measured by the standard deviation

$$s = \sqrt{np\,(1-p)}$$

There is about a 95% chance that the number of minority hires in a random sample will be within 2 standard deviations from the expected number np. When the observed number of minorities is not in this interval, statisticians call this a "statistically significant" result as it suggests that the observed data is inconsistent with the assumed fraction p. For example, suppose women form 25% of the qualified labor force and a company hires 100 employees. We expect $np = 100(.25) = 25$ women to be hired. The standard deviation is

$$\sqrt{np\,(1-p)} = \sqrt{25\,(.75)} = 4.33$$

If only 10 women are hired we calculate the difference between the observed and expected number of hires expressed in standard deviation units, i.e.,

$$\frac{observed-expected}{4.33} = \frac{-15}{4.33} = -3.46$$

This difference exceeds three standard deviations and implies that statistically significantly fewer women were hired than expected given their proportion of the relevant labor pool.

13 *Batson*, 476 U.S. at 96.

14 *Id.*

15 380 U.S. 202 (1966).

hurdle for defendants, and essentially impossible in jurisdictions where the requisite information—the race of the jurors—was not preserved.[16] The Court did not state that defendants must only use information from their own case; rather, statistical data on the practices of the same prosecutor or of prosecutors from the same office or in the same system is admissible evidence in conjunction with the pattern of strikes in the defendant's trial.

Before describing the criteria courts have used to evaluate the completeness of a statistical analysis, we note that Justice Powell's opinion in *McCleskey v. Kemp*[17] distinguished death-sentencing cases from employment discrimination cases with respect to the number of potentially relevant variables, and asserted that there is no common standard by which to compare and evaluate all defendants who were eligible for the death penalty.[18] Since the death penalty may be given for a wide variety of very serious offenses it is unclear that this reservation concerning statistical analysis of sentencing data should apply to a more homogeneous set of crimes.

In practice, virtually no data set is ideal; one can almost always suggest another potentially relevant variable or question the accuracy of the measurement of a factor: for example, is education adequately measured by years of school completed or does one need the grade point average too? In the context of a regression analysis, which predicts pay as a function of appropriate factors like seniority,[19] the Court in *Bazemore v. Friday*[20] stated that it is not necessary for a party offering a regression to incorporate all measurable variables relating to productivity as long as the model includes the major ones. In legal terms, omissions should affect the weight or importance of the evidence rather than its admissibility. Subsequently, in *Allen v. Seidman*,[21] Judge Posner noted that it is easy to take "pot shots" at a statistical

16 *Batson*, 476 U.S. at 96 n.17.

17 481 U.S. 279 (1987).

18 *Id.* at 294-95, n.14.

19 In equal pay cases regression analysis fits the salary data to an equation incorporating the factors related to job productivity. For example, one might believe that years of experience and years of education above high school increase one's productivity in a particular job or set of jobs. The model might then be:

$$\text{Salary} = \text{Constant} + b(\text{Experience}) + c(\text{Education}) + d(\text{Sex}).$$

The coefficients b, c and d reflect the role of the corresponding factor. If sex does not play a role in salary determination, the coefficient d should be zero. Statistical tests, which incorporate sampling variation in the estimate of d—i.e., on actual data from a fair employer the value of d might differ from zero due to random factors—are available. More details can be found in Judge Patrick Higginbotham's opinion in *Vuyanich v. Republic National Bank*, 505 F. Supp. 224 (N.D. Tex. 1980), or in 1 JOSEPH L. GASTWIRTH, STATISTICAL REASONING IN LAW AND PUBLIC POLICY ch. 8 (1988).

20 478 U.S. 385 (1986).

21 881 F.2d 375 (7th Cir. 1989).

analysis and that the defendant needs to do more than simply raise a potential flaw.[22] The critic should show that the flaw would seriously influence the ultimate inference. A number of statistical methods have been developed for assessing the potential effect of missing data,[23] omitted variables,[24] and errors of measurement.[25] These methods can be used to assist courts in evaluating statistical evidence.

III. *STEPHENS V. STATE*: RACIAL DISPARITIES IN PROSECUTORIAL REQUESTS FOR LIFE SENTENCES FOR REPEAT DRUG OFFENDERS

A. THE MAJORITY OPINION

Mr. Stephens challenged the constitutionality of a Georgia law[26] which provided for mandatory life imprisonment upon a second conviction for the sale of a controlled substance or possession with intent to distribute a controlled substance. As the law gives prosecutors the discretion to seek a life sentence, he contended that the law was applied in a racially discriminatory manner in violation of both the United States and Georgia Constitutions.

To support his claim, Stephens submitted both statewide and county-wide statistical data. A statewide study conducted by the Georgia Department of Corrections showed that of all persons eligible for a life sentence, only 1 of 168 whites sentenced for two or more convictions for drug sales is serving a life sentence compared to 202 out of 1219 blacks.[27] Apparently the statistical significance of this data was not calculated and submitted into evidence.[28] The court also found

22 Allen v. Seidman, 881 F.2d 375, 381 (7th Cir. 1989).

23 *See* RODERICK J.A. LITTLE & DONALD B. RUBIN, STATISTICAL ANALYSIS WITH MISSING DATA (1987).

24 *See* PAUL R. ROSENBAUM, OBSERVATIONAL STUDIES (1995); Joseph L. Gastwirth, *Methods for Assessing the Sensitivity of Statistical Comparisons Used in Title VII Cases to Omitted Variables*, 33 JURIMETRICS 19 (1992).

25 WAYNE FULLER, MEASUREMENT ERROR MODELS (1987).

26 GA. CODE ANN. § 16-13-30(d).

27 *Stephens v. State*, 456 S.E.2d 560, 561 (Ga. 1995).

28 The data would ordinarily be presented as a two-by-two table reporting the number of persons of each race who were serving life sentences as well as those not serving:

Race	Life Sentence	Lesser Sentence	Total
Black	202	1017	1219
White	1	167	168
Total	203	1184	1387

The corresponding proportions, .006 for whites and .166 for blacks, show that black repeat drug offenders were .166/.006 = 27.8 times more likely to be serving life terms than whites. The appropriate statistical test is Fisher's exact test, which calculates the probability that a disparity at least as large as the one observed would occur assuming the 203 drug

that 98.4% (369 of 375) of the state's prisoners serving life sentences for drug offenses were African-American, although they constituted only 27% of the state's population. In Hall County, where the defendant was convicted, all fourteen persons serving a life sentence for drug offenses were African-American.[29] The dissent noted that African-Americans form less than ten percent of the county's population, but accounted for 50% to 60% of those arrested in drug investigations.[30]

In a divided opinion, the majority held that this statistical evidence was insufficient to support Stephens' claim of an equal protection violation under the Georgia Constitution. The opinion observed that Stephens failed to present critical evidence concerning persons eligible for life sentences under the statute in Hall County, but against whom the district attorney failed to seek the enhanced sentence.[31] The majority noted that in each judicial circuit the district attorney exercises discretion in seeking the increased sentence, so that a defendant must present some evidence addressing whether the prosecutor handling a particular case engaged in selective prosecution in order to prove a state equal protection violation.[32]

Judge Fletcher went on to criticize the statistical study for ignoring other factors that may explain the disparity in sentencing. His opinion mentions the following possibly significant factors: the charge brought, concurrent offenses, prior offenses and sentences, the type of lawyer (retained or court-appointed), whether the defendant plead guilty, the circuit where convicted, and the defendant's legal status.[33] Furthermore, Judge Fletcher wrote, "[w]ithout more adequate information about what is happening . . . we defer deciding whether statistical evidence alone can ever be sufficient to prove an allegation of discriminatory intent in sentencing under the Georgia constitution."[34] Thus, the issue of the type of proof defendants need to offer to establish differential treatment of eligible individuals and the role of statistical evidence in these cases remains unsettled.

---

offenders were chosen at random from the 1387 eligibles. This probability is zero to eight decimal places, which is far more significant than the usual .05 or .01 levels used in social science and most equal employment cases. The formula appears in 1 JOSEPH L. GASTWIRTH, STATISTICAL REASONING IN LAW AND PUBLIC POLICY 217-25 (1988), and the test has been accepted in equal employment cases.

29 *Stephens*, 456 S.E.2d at 560-61.

30 *Id.* at 568 n.3 (Benham, J., dissenting).

31 *Stephens*, 456 S.E.2d at 562.

32 *Id.*

33 *Id.*

34 *Id.*

B. THE MINORITY OPINION

The dissent, written by Judge Benham, described the statistics as "numbing and paralyzing,"[35] and objected to the majority opinion's failure to state that they substantiate a need for serious inquiry. After reviewing some of the major U.S. cases on jury discrimination, he noted that *Batson* held that an inference of discriminatory intent could be drawn from certain conduct or statistical data and recognized that the crucial information about an allegedly discriminatory decision could only come from the one who made the decision.[36]

Judge Benham believed that *Batson* was more relevant to Stephens' case than *McCleskey* because *Stephens* concerned a well-defined sentencing process for a small number of related criminal activities. His dissent implied that the application of the Georgia statute allows a prosecutor's charging discretion to be exercised in a discriminatory fashion just as peremptory challenges may be used to discriminate. Thus, the prosecutor, when confronted with facts supporting an inference of discrimination, should bear the burden of explaining the data.[37] Moreover, since the life penalty is only imposed in about 15% of the cases, he observed that the state should be required to give race-neutral reasons for the "monochromatic" application of the statute in Hall County.[38]

C. A STATISTICAL VIEW OF THE EVIDENCE

1. The Hall County Data

As all opinions in the case mention the importance of Hall County, the jurisdiction in which Stephens was convicted, we begin by calculating the probability that all fourteen persons serving life sentences for drug offenses in Hall County would be black, if drug suspects were a reasonable pool from which drug offenders would come. As several investigators stated that between 50% and 60% of drug investigations involved black males, we will assume that blacks form 60% of those eligible for life sentences. By choosing the high end of the range, we favor the state and implicitly allow for a difference between the racial compositions of drug suspects and repeat offenders, although we believe that the two populations should be

35 *Id.* at 566 (Benham, J., dissenting).

36 *Id.* at 567 (Benham, J., dissenting). A similar point was made earlier by Amy G. Applegate, *Prosecutorial Discretion and Discrimination in the Decision to Charge*, 55 TEMPLE L. Q. 35, 76-78 (1982).

37 *Stephens*, 456 S.E.2d at 567 (Benham, J., dissenting); *see also* JOHN HART ELY, DEMOCRACY AND DISTRUST 175 (1980) (broad prosecutorial discretion opens the door to racial discrimination).

38 *Stephens*, 456 S.E.2d at 570 (Benham, J., dissenting).

similar. Moreover, we assume that targets of investigation are chosen in a race-neutral manner. The standard binomial test[39] is statistically significant at the usual .05 level. Indeed, the probability of the data is .0008, or less than one in a thousand. It can thus be shown that in order for the fourteen of fourteen figure not to be statistically significant at the two-sided .05 level, which corresponds to the Court's two standard deviation criteria, blacks would need to form at least 77% of those eligible for drug-related life sentences in Hall County.[40] While it is conceivable that blacks might form a somewhat higher percentage of repeat offenders than they do of drug-related investigation targets in Hall County, state officials, with better access to the data, should have the burden of showing a lack of disparity by introducing appropriate evidence demonstrating that blacks form over 77% of those eligible for a life sentence.

We also note that African-Americans, who form less than 10% of the county's population, composed over half the subjects of drug investigations. Although this fact might raise a question about the race-neutrality of selecting subjects for investigation of drug-related activity—for example, whether drugs used more heavily in the black community were the subject of more investigations than other drugs—the statistical test ignores this issue and assumes that blacks form 60% of those eligible for life sentences. If racial bias entered into the process of choosing drug suspects, the actual percentage of blacks among serious drug offenders who should be considered for life sentences would be less than 60%, and the statistical result would be more significant.[41]

2. *Are Judge Fletcher's Caveats about the Statewide Data Likely to Explain the Disparity in Life Sentencing Requests?*

Earlier we listed a number of factors that the majority opinion suggested might reduce the *Stephens* disparity to a legally acceptable one. Now we examine them using a statistical technique developed by Cornfield during the smoking/lung cancer controversy, and used by Gastwirth[42] to analyze employment discrimination data. The result gives numerical criteria for an omitted factor, which is either present, e.g., regular smokers, or absent, e.g., non-smokers, to explain an ob-

---

[39] *See supra* note 12, for an explanation of the standard binomial test. Further examples are given in 1 GASTWIRTH, *supra* note 19, at 167-77.

[40] This approach was accepted in *Capaci v. Katz & Besthoff, Inc.*, 711 F.2d 647, 653 (5th Cir. 1983), in which the evidence was that men would have needed to form 98% of the labor pool in order for the zero number of female hires not to be statistically significant.

[41] A result is more statistically significant than another if the probability of it occurring due to chance or a process of random selection is smaller.

[42] *See* Joseph L. Gastwirth, *Employment Discrimination: A Statistician's Look at Analysis of Disparate Impact Claims, in* II LAW & INEQ. 151, 155 (1992).

served ratio of two proportions, e.g., the lung cancer rates of smokers and non-smokers. Because it examines whether a particular "missing factor" could account for the observed disparity or ratio of two proportions, it focuses on the known influence of the specific factor and the data from the actual case.

First, the omitted factor or factors must be at least as strongly related to the response, that is, multiply one's chance of producing response (such as requesting a life sentence) as the ratio of black to white life sentencing requests in the data. Secondly, the prevalence (proportion of individuals meeting the factor) of the factor among black eligibles needs to be much larger than the prevalence amongst whites. For example, if we have 1000 black defendants, 250 of whom use firearms, the prevalence of firearm use in that population would be .25.

To state the result we need to define the relative risk as the ratio of two proportions. In the state data, the relative risk, R, of a black facing a life sentence is (202/1219)/(1/168) = .1657/.0059 = 27.8 times that of a white eligible. The formal statement of Cornfield's lemma[43] is:

In order for an omitted factor (X) to explain an observed relative risk (R),

a) the relative risk (R_X), associated with factor X must exceed R, and

b1) the prevalence ratio, $\theta = f_2/f_1$, *must be at least*

$$R + \frac{R-1}{R_x-1} \cdot \frac{1}{f_1}$$

or equivalently that

b2) $f_2 \geq R f_1 + (R-1)/(R_x-1)$.

In our application of the lemma, we allow for sampling variability by using the lower end of a 95% confidence interval[44] for R, instead of the 27.8 calculated from the raw data. This value is 5.01, which implies that in order for a factor to explain the data it must multiply a drug criminal's chance of having the prosecution ask for a life sentence by five-fold and its prevalence amongst black repeat drug offenders must be at least five times its prevalence amongst whites.

Suppose one or several of the factors Judge Fletcher listed would justify the prosecution requesting a life sentence. This implies that

---

[43] *Id.* at 156.

[44] If one considered that the offenders given a life sentence were a random sample from those eligible, there would be a range of observed relative risks. A 95% confidence interval is a commonly used range of plausible values for the true relative risk, R. Technically, if one took many samples 95% of the time, the true R will lie in the interval obtained from the sample.

they would increase an offender's probability of being given a life sentence by a factor, R_x. For illustrative purposes we take R_x=10, although our experience suggests that this is quite large. If only one-half of one percent of the white defendants had this factor, then equation b2) implies that 47.1% of the black defendants would need to possess this factor.[45] As we are comparing blacks and whites convicted for drug crimes, it does not seem realistic to believe that the prevalences of any of the factors mentioned by Judge Fletcher differ so widely in the two groups. Moreover, no evidence was cited indicating that any of the variables mentioned, either individually or collectively, increase the probability a prosecutor could justifiably ask for a life sentence by a factor of 5, much less the value 10 used in the above calculation.

While we believe that the state should be asked to demonstrate that the factors listed by Judge Fletcher satisfy the criterion for reducing the disparity to a non-significant magnitude,[46] we examine whether the named factors satisfy the conditions of the Cornfield lemma. First, it seems quite implausible that some of the factors could meet the prevalence condition, that is, be five times more prevalent among black eligibles than whites. For instance, could the type of legal counsel or the rates of entering guilty pleas of eligible black and white defendants differ by a factor of five?

We next review some studies which looked at jury decisions or racial differences in sentencing to see whether some of the explanatory factors accepted by the majority had a sufficiently strong effect and were so differently distributed in the white and black groups that Cornfield's condition could reasonably be met. Spohn, Gruhl, and Welch[47] studied the sentences imposed on 2366 individuals convicted of crimes in a large city. There were 1939 black defendants and 427 whites in their sample.[48] They observed that most of the disparity in sentences received by black and white criminal defendants was explained by the charge brought and the existence of a prior record. However, even after controlling for legal and extra-legal factors, black males had a higher probability of receiving a prison sentence than similar whites.[49] Because the statistical technique, multiple linear regression, used by those authors is more appropriate for continuous variables such as length of sentence than binary or yes-no variables we

---

[45] This follows by substituting the values R=5.0, f^1=.005, and R^x=10 in the right side of equation (b2), yielding (5.01)(.005)+(4.01)/9 = .471.

[46] *See supra* note 33 and accompanying text.

[47] Cassia Spohn et al., *The Effect of Race on Sentencing: A Re-Examination of an Unsettled Question*, 16 L. & SOC'Y REV. 71 (1981).

[48] *Id.* at 74.

[49] *Id.* at 85.

will not discuss the details of their study.[50]

The classic study by Kalven and Zeisel[51] of jury verdicts also noted that the type of attorney had a minor effect on the verdict. However, a recent reanalysis by Gastwirth and Sinclair[52] showed that the odds of a jury being more lenient than a judge in serious crimes were doubled if the judge thought the defendant's lawyer was superior to the prosecution's.[53] While the drug crimes involved in *Stephens* would be serious, the potential effect of a better lawyer is not even close to the factor of 5 required by the Cornfield lemma.[54] Moreover, no evidence of a difference in type or quality of attorney by black and white defendants was cited in the opinion.

Thus, the legal factors cited by Judge Fletcher in the *Stephens* opinion do not appear sufficiently strongly related to the probability of imprisonment to explain the difference in life sentences requested by Georgia state prosecutors. Of greater importance is that the major explanatory variables found in the Spohn et al. study—type of charge and the existence of a prior record—are quite similar for repeat drug offenders. The charge is always a drug offense and all individuals eligible have a prior record of such an offense.[55] While it is conceivable that blacks might have more prior offenses, be involved in different types of drug-related criminal activity than whites, or have a much higher frequency of violence associated with their crimes, courts

---

50 To illustrate, the inappropriateness of the correlation measures used in linear rgression for assessing the strength of relationship between binary variables, we examine their reporting an ordinary Pearson correlation, r, of race and imprisonment of .144. The correlation is commonly used to describe the relationship between continuous variables such as LSAT score and law school GPA, where correlations of .3 to .5 are usual. Thus, one might believe that their result indicates a small relationship between race and imprisonment. The following table shows that an r=.144 can reflect quite a meaningful difference in proportions:

	Prison Sentence	Not	Total	% Incarcerated
Black	2249	5751	8000	28.11%
White	251	1749	2000	12.55%

While the Pearson correlation between race and receiving a prison sentence is .144, blacks are more than twice as likely to receive such a sentence than whites.

51 HARRY KALVEN, JR. & HANS ZEISEL, THE AMERICAN JURY 351-72 (1966).

52 Joseph L. Gastwirth and Michael D. Sinclair, A Re-Examination of the Kalven-Zeisel Study of Judge-Jury Agreements Using Recent Contingency Table Analysis Techniques (January, 1997) (unpublished Technical Report, on file with the Department of Statistics, The George Washington University).

53 *Id.* at 9.

54 Technically, the odds ratio is a different measure than the relative risk. For the data in the *Stephens* case, however, the odds ratio exceeds the relative risk. Hence the value 2 may overestimate the effect of a better lawyer in the Cornfield lemma.

55 Spohn et al., *supra* note 47, at 83.

should require the presentation of such evidence. This is especially true in *Stephens*, where a study of drug offenders in Georgia during the period from January 1977 through May 1985 was available.[56]

Myers found that the probability of being given a prison sentence depended on the defendant's involvement in the drug trade, and was .2 for users, .4 for sales offenses, and .6 for trafficking.[57] From Table 1 in Myers we calculated that 63.57% of black offenders were users, 35.64% were selling drugs and only 0.79% were traffickers. Correspondingly, 72.83% of whites were users, 26.25% were selling drugs, and 0.92% were traffickers. Thus, the prevalences are similar, and the largest ratio, 1.36, occurs for the seller category (35.64/26.25). If these prevalences are even roughly applicable to more current Georgia data, then the second part of Cornfield's criterion, which requires a five-fold difference in the proportions of black and whites in the categories of drug offenders deserving the most serious punishment, cannot be satisfied. Since the statute allowing life imprisonment focuses on the distribution of drugs, the ratio of the probabilities of incarceration for individuals who were traffickers or sellers to that of users should indicate the increased likelihood of a severe penalty. With Myers' data, these ratios were 3.0 (traffickers to users) and 2.0 (sellers to users), again substantially less than the factor of 5 required by Cornfield's lemma. This implies that, by itself, the degree of involvement in the drug trade cannot fully explain the racial disparity in the state-wide data.

Furthermore, a regression analysis relating the probability of receiving a prison sentence to race and related factors such as the nature of the offense and the seriousness of the crime estimated that blacks had a 13% higher probability of receiving a prison sentence than similarly situated whites.[58] That analysis did not include data on prior arrests or incarcerations, but the number of prior arrests was negatively related to the degree of involvement in the drug trade. Indeed, the average number of prior arrests was 3.3 for users, 2.8 for salespersons, and 2.2 for traffickers. Myers observed that for both blacks and whites the probability of imprisonment increased with the degree of involvement in the drug trade. The racial disparity also increased with the level of involvement. Indeed, black users were 12.5% more likely to receive a prison term than whites.[59] The corresponding

---

[56] Martha A. Myers, *Symbolic Policy and the Sentencing of Drug Offenders*, 23 L. & SOC'Y REV. 295 (1989).

[57] *Id.* at 302 tbl.1.

[58] *Id.* at 305 tbl.3.

[59] *Id.* at 310 tbl.5.

disparities for sales and trafficking offenders were 18.6% and 24.6%.[60] This implies that the seriousness of the drug-related offense is unlikely to justify the racial disparity in *Stephens*, as the racial disparity in imprisonment rates increased with the seriousness of drug activity. Had the racial disparities in incarceration rates decreased with the level of involvement, then this factor might help explain the data in *Stephens*. If the prevalences of the various offenses or the racial disparities in imprisonment rates changed in the direction of greater fairness in the time between Myers' study and the trial, one could have expected the state to submit this information in the Stephens case.

An earlier study by Unnever,[61] of 313 drug offenders in Miami during 1971, illustrates the wisdom of Judge Posner's remark in *Allen v. Seidman* that it is possible for a full statistical analysis incorporating alleged confounding factors suggested by the party opposing a statistical presentation to actually strengthen the original conclusion.[62] In a logistic regression[63] of the odds of receiving a prison sentence, as more variables, such as the number of charges, bail status, and type of attorney were added, the odds of a black defendant receiving a prison sentence relative to a similar white actually increased slightly from about 2.1 to 2.5.[64] While receiving a prison sentence from a judge is different than having a prosecutor request a life sentence, the fact that adjusting for these factors in an analysis of drug offenders in Miami did not reduce the racial disparity raises serious doubts that they could explain the substantial disparity in life sentence requests in the Georgia data.

It should be emphasized that we are not claiming that racial disparities as large as the one in the statewide data on prosecutorial requests for life sentences in Georgia could never be explained by a legitimate race-neutral factor or set of such factors. Cornfield's criteria, however, indicate that these factors would need to be extremely highly related to a defendant deserving a life sentence, and the fraction of black repeat drug offenders possessing these factors would need to be much larger than the corresponding fraction of white repeat drug offenders. The relevant literature we examined indicates that these conditions are not likely to be satisfied by the factors listed by Justice Fletcher, as they could not come close to "explaining" simi-

---

60 *Id.*

61 *See* James D. Unnever, *Direct and Organizational Discrimination in the Sentencing of Drug Offenders*, 30 SOC. PROBLEMS 212 (1982).

62 Allen v. Seidman, 881 F.2d 375, 380 (7th Cir. 1989).

63 Logistic regression is a variant of regression, *supra* note 19, in which the response is a yes-no or 0-1 variable. It is appropriate for the examination of hiring and promotion data. *See* 1 GASTWIRTH, *supra* note 19, at 442-55.

64 Unnever, *supra* note 61, at 219 tbl.2.

lar racial disparities in incarceration rates of drug offenders.

IV. *U.S. v. Armstrong*: How Much Evidence of Discrimination is Needed to Obtain Discovery?

A. THE MAJORITY OPINION

In *U.S. v. Armstrong*,[65] the defendants were indicted for conspiring to distribute cocaine base, and some were charged with selling crack and using a firearm in connection with drug trafficking. They alleged that the government discriminated against them as blacks in deciding to prosecute them for violating federal rather than state law.[66] The sentence for such charges is noticeably greater under federal law.[67] The defendants moved for discovery on their claim but the government chose not to comply with the court order, arguing that there was no evidence that the government had acted unfairly.[68]

To support their claim of selective prosecution the defendants cited a study of the twenty-four cases prosecuted under the relevant laws.[69] All twenty-four defendants were black. On the basis of the data the district court ordered the government to:

1) provide a list of all cases it prosecuted for cocaine base and firearms offenses;
2) identify the race of the defendants;
3) identify which law enforcement agency investigated the case; and
4) explain the criteria the U.S. Attorneys used to decide to bring cocaine base cases to federal court.[70]

While the government did not supply all the information, it did contend that seven non-black defendants had been prosecuted, although all were members of minority groups, and that many blacks had been prosecuted in state, rather than federal, court. The government also referred to four specific factors it considered in bringing the cases to federal court:

1) strength of the evidence;
2) the deterrent value of bringing a charge;
3) the federal interest in bringing the charge; and
4) the criminal history of the suspect.

More importantly, the government explained, the defendants had all

---

65 116 S. Ct. 1480 (1996), *rev'g* 48 F.3d 1508 (9th Cir. 1995).

66 *Armstrong*, 116 S. Ct. at 1483.

67 *Armstrong*, 48 F.3d at 1511. The sentence is a minimum of 10 years to a maximum of life under Federal law, and a minimum of three years to a maximum of five years for a comparable offense under California law. *Id.*

68 *Armstrong*, 116 S. Ct. at 1484.

69 *Id.* at 1483 & n.1. *See* 21 U.S.C. § 841, 846.

70 *Armstrong*, 116 S. Ct. at 1484.

been involved extensively in drug dealing.[71]

In response, the defendants provided information that blacks and whites were equally represented among cocaine base addicts at a particular treatment center, and that many whites were prosecuted in state court.[72] The Court of Appeals upheld the dismissal of the cases ordered by the district judge in response to the government's refusal to obey the discovery order.[73] The appellate majority agreed with the district judge's decision that the data was sufficient in number and time frame to require the government to provide some explanation of its procedures.[74] The dissent argued that the data did not compare the twenty-four blacks prosecuted with a group of similarly situated individuals of other races.[75] It cited data showing that 87.9% of all those convicted for cocaine-based offenses nationwide are African-American.[76]

The Court granted *certiorari* to determine whether the data submitted to the Court were sufficient to support discovery in selective prosecution cases.[77] The Court also considered whether Federal Rule of Criminal Procedure 16 was applicable. As this second issue is purely a legal one, the statistical evidence had no role in that part of the decision.

After noting that the showing necessary to obtain discovery should be a significant barrier to the litigation of insubstantial claims,[78] the Supreme Court held that in order to meet the legal requirements entitling discovery, the defendant must produce evidence that similarly situated defendants of other races could have been prosecuted but were not.[79]

The *Armstrong* majority pointed out that the Court of Appeals reached its earlier decision by assuming that people of all races commit all types of crimes, and that no type of crime is the exclusive province of any particular group. The opinion cites data showing that 91% of individuals convicted of pornography or prostitution were white, while over 90% of those convicted of crack cocaine trafficking were black.[80]

The Court did not accept the defendants' data on the percentage

---

71 *Id.*
72 *Id.*
73 *Id.*; *see also Armstrong*, 48 F.3d 1508, 1520 (9th Cir. 1995).
74 *Armstrong*, 48 F.3d at 1516.
75 *Id.* at 1521-22, 1524 (Rymer, J., dissenting).
76 *Id.* at 1524 n.4 (Rymer, J. dissenting).
77 *Armstrong*, 116 S. Ct. at 1485.
78 *Id.* at 1486.
79 *Id.* at 1488-89.
80 *Id.*

of black persons in drug treatment, or the attorney's affadivit that many non-blacks were prosecuted in state court for crack offenses, holding that such evidence did not meet the required showing of different treatment of similarly situated persons.[81] The Court observed that the defendant had not identified non-black individuals who could have been prosecuted for the offenses for which respondents were charged but were not prosecuted.[82]

B. THE MINORITY OPINION

Justice Stevens' dissent criticized the majority opinion for discounting the affadavit of an experienced criminal defense lawyer.[83] Justice Stevens noted that the presumption that some whites are prosecuted in state court is not "contradicted" by the statistics on the percentage of blacks convicted. Indeed, they are consistent with selective prosecution of blacks. The most relevant comparison therefore would use the racial composition of individuals who commit the crime.[84]

In a footnote[85] Justice Stevens refers to the Government's own data as showing that of 3500 defendants charged with Federal narcotics violations during the previous three years, all but eleven were black. Furthermore, those eleven were all from other minority groups. No white defendant was among the 3500.

C. STATISTICAL ANALYSIS OF THE DATA

In this section we first examine the implication of the defendants' data and then consider the larger data base described in the dissent. Recall that the defendant suggested that the percentage (50%) of black individuals receiving drug treatment was a reasonable estimate of those involved in drug dealing, while the dissenting judge at the appellate level used the 87.9% figure derived from convictions in the nation.

In Table 1 we provide the results of the statistical test used by the court in *Castenada v. Partida*[86] for various black fractions (p) ranging from .5 to .879. If the defendants' value of $p = .5$ is correct, then the probability that all twenty-four individuals prosecuted would be black would be less than one in one million, corresponding to a difference of 4.9 standard deviations from expected. This disparity is far in excess of the two to three standard deviation criteria set forth in *Cas-*

81 *Id.* at 1489.
82 *Id.* at 1489.
83 *Id.* at 1494 (Stevens, J., dissenting).
84 *Id.* (Stevens, J., dissenting).
85 *Id.* at 1495 n.6 (Stevens, J., dissenting).
86 430 U.S. 482 (1977).

tenada.[87] On the other hand, if the nationwide data is appropriate, $p = .879$, and the disparity is 1.8 standard deviation units, which is less than the two standard deviation criteria adopted by the Court. This criterion corresponds to the .05 (or five percent) level of significance that is typically used in scientific journals.

TABLE 1. THE PROBABILITY OF OBSERVING TWENTY-FOUR BLACKS IN A SAMPLE OF TWENTY-FOUR INDIVIDUALS CHOSEN FROM A POPULATION WHERE BLACKS FORM THE FRACTION P OF THE POPULATION.

(1) Fraction (p)	(2) Exp. Number	(3) STD. DEV	(4) Z-score	(5) one-sided prob.
.50	12	2.4495	4.899	6×10^{-8}
.60	14.4	2.40	4.0	4.74×10^{-6}
.70	16.8	2.245	3.207	1.92×10^{-4}
.75	18	2.1213	2.828	.0010
.80	19.2	1.9596	2.449	.0047
.85	20.4	1.7493	2.058	.0202
.879	21.096	1.5977	1.818	.0453

Note: The Z-score is the statistic used in *Castenada* v. *Partida* which is the normal approximation to the exact binomial test. It is the ratio of the difference between the actual data (24) and the expected number (col. (2)) divided by the std. deviation (col. (3)).

It seems clear, however, that the .879 figure may well be too high as it refers to persons convicted, who first must be prosecuted. If the government does selectively prosecute blacks, then their proportion amongst persons convicted is likely to reflect the alleged discrimination.[88] Moreover, the applicability of national data to a single district is questionable, especially if the racial make-up of the district differs from that of the nation. When a firm's hires are compared to a proxy labor force in equal employment cases courts require data from the local labor market, often adjusting for skills and commuting patterns.[89] On the other hand, the defendants' assumption that the ad-

[87] *Id.* at 512 n.17.

[88] There is data suggesting this may be the case. *See* Memorandum from Richard Berk and Alec Campbell to Paul Rochnes, *Preliminary Data on Race and Crack Charging Practices in Los Angeles*, 6 FED. SENT. REP. 36 (1993) (over a two-year period, blacks formed 58% of arrests, but 83% of those charged at the Federal level).

[89] *See* DAVID C. BALDUS & JAMES W.L. COLE, STATISTICAL PROOF OF DISCRIMINATION 61-67 (Supp. 1987); Sheldon E. Haber & Joseph L. Gastwirth, *Specifying the Labor Market for Individual Firms*, 101 MONTHLY LAB. REV. 26 (August 1978), for discussions of weighted labor markets. These authors note that it is important to use weights obtained from residence data from actual applicants rather than from employees. Indeed, in *Markey v. Tenneco Oil Co.*, 635 F.2d 497 (5th Cir. 1981), weighting each sub-area by residence of applicants yielded an availability of 42% while weighting by residence of employees at their time of hire yielded a figure of 32.61%. *Id.* at 500 n.3. Thus, the area or region used to determine the black fraction of potential employees or, as in *Armstrong*, their fraction of those accused for cocaine related crimes is quite important. Just as unrefined national data should not be used in an equal employment case, it should not be accepted as providing a reliable

dict population is a proxy pool for any traffickers is also questionable as addicts often are users rather than dealers.

From Table 1 we can obtain the minimum value of the fraction of black individuals eligible for prosecution that would yield a figure of twenty-four blacks out of twenty-four prosecutions as a plausible result of sampling from this eligible population. If we select a level of significance of .001, corresponding to 2.88 standard deviations, we see that $p = .75$. Thus, unless the fraction of blacks involved in cocaine base drug violations in the district is at least .75, the data would be statistically significant at nearly the three standard deviation level. If one believed that statistical significance at the two standard deviation level would be sufficient for the plaintiff to establish a prima facie case, then the fraction of black persons accused of cocaine base related drug crimes would only need to exceed $p = .854$. The defendants' data would satisfy the two standard deviation criteria if the government's national figure ($p = .879$) that was accepted by the dissenting appellate judge overestimates p by just .03.

An amicus brief filed in the *Armstrong* case[90] reported the following racial data on defendants charged in federal crack prosecutions as a result of sting operations in the Central District of California during 1992-94: black (109), Hispanic (28), Asian (8), and white (1).[91] Thus, blacks formed 74.7 percent of these defendants.[92] As this data pertains to the same District, it is more relevant than the national data. Had this fraction ($p = .747$) been used in Table 1, a result very close to the one for $p = .75$ would have been found—i.e., the data would almost reach significance at the three standard deviations level.

The *Armstrong* opinion used the relative risk measure in describing the discriminatory effect of a section of the Alabama Constitution examined by the Court in *Hunter v. Underwood*.[93] The Court noted that in that case, by modest estimates blacks were at least 1.7 times as likely to be disenfranchised under the constitutional provision in question.[94] We now calculate a statistically conservative estimate of the risk an eligible black drug defendant had of being prosecuted in federal court, relative to a white defendant in the Central District of California. If we consider the twenty-four offenders prosecuted for crack offenses in 1991 as a sample of all prosecutions over a period of

estimate of the black fraction of accused criminals eligible for federal prosecution in the Los Angeles area.

90 Brief for the National Association of Criminal Defense Lawyers as Amicus Curiae, U.S. v. Armstrong, 116 S. Ct. 1480 (1996) (No. 95-157).

91 *Id.* at 16.

92 *Id.*

93 471 U.S. 222 (1985).

94 U.S. v. Armstrong, 116 S. Ct. 1480, 1487 (1996) (quoting *Hunter*, 471 U.S. at 227).

years, there is (as we noted earlier) a high degree of confidence that the fraction of black defendants prosecuted is at least .8575.[95] As blacks formed only 74.7% of those eligible for federal prosecution, the risk of a black drug offender relative to a white drug offender of being prosecuted in federal court is shown by the equation[96]
(.8575/.1425) / (.747/.253) = 6.0175/2.9526 = 2.038.
As 2.04 exceeds 1.7, the value the Court accepted as demonstrating discriminatory effect in *Hunter*, this value should support a finding of discriminatory effect here.

Since the criteria for obtaining discovery should be less than that required to establish a prima facie case, and as over 85% of individuals eligible for prosecution under Federal law would need to be black in order for the data not to be statistically significant, it seems to us that some further investigation of the racial composition of eligibles is appropriate.

We now examine the implication of the second data set cited by Justice Stevens. A similar calculation shows that in order for no whites to be among the 3500 persons prosecuted for Federal narcotics violations *not* to be statistically significant, whites would need to form *less* than *two-tenths of one percent* of those eligible for prosecution for drug offenses. Given the widespread illegal use of narcotics in the nation and in California, we doubt that the white percentage of eligibles could be so low. Indeed, none of the large racial imbalances amongst persons convicted of various types of crimes cited in the majority opinion[97] are this lopsided.

D. RELATED ISSUES

While our re-analysis of the data in the *Armstrong* case supports Justice Stevens' dissenting opinion that discovery is justified, it is quite possible that a careful study would show that the Government also was justified in prosecuting the defendants in Federal court. According to an affidavit cited in the opinion[98] the defendants participated in multiple sales of cocaine base that exceeded twice the threshold necessary for a mandatory ten year sentence and the evidence against them was

---

95 *See supra* tbl.1 and accompanying text.

96 The formula is equivalent to equation (5B.2) in 1 GASTWIRTH, *supra* note 19, at 241, with $p = .747$, and $a/t = .8575$. A simple illustrative calculation, using the same percentages of eligibles (.747) and selected for prosecution (.8575) may be helpful. Suppose there are 100,000 eligible drug offenders (74,700 black, 25,300 white). Ten thousand (8575 black, 1425 white) are prosecuted in Federal court. The relative risk of a black to a non-black offender would be (85700/74700) / (1425/25300) = 2.038. An equivalent formula is discussed in BALDUS & COLE, *supra* note 89, at 85 n.18.

97 *Armstrong*, 116 S. Ct. at 1489.

98 *Id.* at 1484.

quite strong. Only an examination of the files under the control of the government, however, could demonstrate that all drug offenders so highly involved in the drug trade were prosecuted under the federal statute, and that race did not influence the prosecution decisions against such serious offenders.

The need for discovery in this case is similar to that in employment discrimination cases where the relevant information is entirely in the files of the employer. As Judge Posner's opinion in *Riordan v. Kempiners*[99] observed, discrimination is difficult to prove, as the best workers will be retained. Thus, an employer's discriminatory practices would more likely manifest themselves in the treatment of average or merely satisfactory employees.[100] By analogy, one would expect that prosecutors who might be biased against a minority group would be harder on minority offenders whose level of criminal activity slightly exceeded a threshold for a more serious penalty than on comparable whites.

The *Armstrong* opinion suggests that the Court was concerned with placing a large administrative burden upon the Government, which would need to assemble information from many files to perform this analysis.[101] From a statistical view, a random sample of about three hundred files would probably suffice to estimate the white proportion[102] of drug offenders eligible for prosecution under federal law. If whites formed less than twenty-five percent of the sample, then from Table 1 it follows that the defendants' data would not show a statistically significant bias against blacks at the .001 level, corresponding to 2.8 standard deviations. Sample surveys are routinely used in Lanham Act cases,[103] and using a sample of individual cases in large class action tort cases to determine a schedule for awarding damages has been used by one court, and recommended in the legal literature.[104] Selecting a random sample of records for the purpose of discovery when unfairness issues are raised in criminal cases may prove

---

99 831 F.2d 690, 697-98 (7th Cir. 1987).

100 *Id.* at 1360.

101 *Armstrong*, 116 S. Ct. at 1488.

102 To determine the sample size, n, needed to estimate a fraction, p, to within .05 with 95% confidence, one uses the formula

$$n = (1.96)^2 p(1-p)/(.05)^2.$$

See LYMAN OTT ET AL., STATISTICS: A TOOL FOR SOCIAL SCIENCE RESEARCH 202 (3rd ed. 1983). In our problem, the white fraction, p, of eligible drug offenders is about $p = .25$. Using this value in the formula yields $n = 289$, or about 300.

103 *See, e.g., supra* note 1.

104 *See* John C. Coffee, Jr., *Class Wars: The Dilemma of the Mass Tort Class Action*, 95 COLUM. L. REV. 1343 (1995); Michael J. Saks & Peter D. Blanck, *Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts*, 44 STAN. L. REV. 815, 815, 851 (1992); Cimino v. Raymark Indus., Inc., 751 F. Supp. 649, 659-65 (E.D. Tex. 1990).

useful and efficient. It would give defendants a reasonable opportunity to discover similarly situated defendants of other races who were not similarly prosecuted as required to prevail on the merits under *Armstrong*.[105]

V. CONCLUSION

The re-examination of the data in the *Stephens* and *Armstrong* cases utilizing Cornfield's result and related statistical tools suggests that black drug offenders have a basis for questioning the fairness of both state and federal prosecutions. The noticeable racial disparity in the data from the *Stephens* case does not appear to be fully explicable on grounds other than race.

In the *Armstrong* case, the data cited by Justice Stevens appears to be much stronger evidence of a racial disparity in the treatment of drug offenders than the small study introduced by the defendants. Although the sample size, twenty-four in the defendants' study, is relatively small, the fact that all twenty-four individuals prosecuted were black implies that the data was as extreme as possible. As an experienced attorney submitted evidence that some white offenders were eligible for prosecution, the data would appear to justify some discovery. We suggest that a reasonably-sized random sample might well suffice for this purpose.

An advantage of sampling is that it reduces the burden on state and federal governments while enabling defendants who have produced some credible evidence of selective prosecution, perhaps based on a small study, to pursue their claim. This should achieve the goal of more liberal criminal discovery advocated by some commentators,[106] while requiring the defendants to produce some support for their claim and preventing defendants from flooding the prosecutors with requests for large numbers of files.

---

105 *Armstrong*, 116 S. Ct. at 1487.

106 *See, e.g.*, Tobin Romero, Note, *Liberal Discovery on Selective Prosecution Claims: Fulfilling the Promise of Equal Justice*, 84 GEO. L.J. 2043 (1996).