Statistical measures and methods for assessing the representativeness of juries: a reanalysis of the data in *Berghuis* v. *Smith*

JOSEPH L. GASTWIRTH† AND QING PAN

Department of Statistics, George Washington University, Washington, DC 20052, USA

To assess the fairness of a jury system, courts compare data on the race–gender–ethnic composition of a number of jury venires (pools) to that of the jury-eligible population of the jurisdiction. The recent U.S. Supreme Court opinion in the *Berghuis* v. *Smith* case discussed the criteria a defendant’s evidence must meet to establish a *prima facie* case that the jury pools were not representative. The decision relied on the three criteria set previously in *Duren* v. *Missouri*: name a distinct subgroup of the population, demonstrate their underrepresentation with statistics and identify the aspect of the jury system, which produced the disparity. In addition to formal statistical testing, courts have considered the absolute and comparative differences between the percentages of minorities in the age-eligible population and on the jury venires. In *Berghuis*, data for the period when the defendant’s trial was held indicate that 6.03% of the potential jurors in the circuit (felony) courts were African American but they formed 7.28% of the age-eligible residents of the County. The 1.25% absolute difference seems small. The comparative difference or percentage shortfall is 18%, suggesting underrepresentation. The defendant argued that the disparity in the circuit courts arose because, at the time of his trial, jurors were first assigned to local courts and African Americans were concentrated in Grand Rapids. Shortly after the trial, the County changed the system, however, the comparative disparity (CD) only declined by 3%. This small change failed to convince the Court that the juror assignment system caused the underrepresentation. The available data in *Berghuis* v. *Smith* differed from most jury discrimination cases as the racial identification of the venire members were unavailable. The defendant’s expert estimated the number of African Americans on the venires from their addresses and compared it to their expected number derived from their fraction of the county’s age-eligible population. He did not test whether the difference was statistically significant. The two analyses available to the Court treated the estimated number of minority jurors on the venires as though it was the actual number, which is statistically incorrect. This article derives the appropriate standard deviation analysis. Applying it to the data in *Berghuis*, it is shown that there was a statistically significant disparity between the percentages of African Americans on the venires and in the age-eligible population. On the other hand, the difference between the CDs in the two periods was not near significance. Thus, the first two *Duren* criteria were satisfied but the third was not, supporting the Court’s decision.

**Keywords:** absolute disparity; comparative disparity; decision theory; fair representation; jury discrimination; measures of disparity or discrimination; selection ratio; statistical hypothesis testing.

1. **Introduction**

Courts in the USA often examine the race, ethnic or gender composition of venires (jury pools) to determine whether the state provides criminal defendants with a fair jury, which is guaranteed in two different parts of the U.S. Constitution. Defendants can argue that their right to equal protection
under the law is violated when individuals of their race–ethnic group are either excluded or substantially underrepresented in the jury pool. They can also argue that their Sixth Amendment right to a fair trial is violated when a distinct subgroup of the population is substantially underrepresented. In fair representation cases, a defendant may challenge the representativeness of the jury pool on grounds that members of any distinct subgroup of the community, including ones that they do not belong to, are underrepresented. In both types of cases, courts have used several statistical measures or methods to examine data on the race–ethnic–gender composition of venires (jury pools). Which statistical measure of the disparity between the percentages of a subgroup in the jury pool and the jury-eligible population is adopted may determine whether a defendant has satisfied one of the criteria needed to demonstrate a \textit{prima facie} case of a nonrepresentative jury pool. These issues arose in the fair representation case, \textit{Berghuis v. Smith}, recently decided by the U.S. Supreme Court. The Michigan state supreme court was not convinced that the statistical underrepresentation of African Americans on the venires of the state court where Smith was tried was large enough to violate his right to a fair trial. It noted that the difference between the 6.0% of 929 venire members during the term of the trial who were African American in the county was not sufficiently less than their 7.28% of the age-eligible population of the county. This absolute disparity (AD) of 1.28% was less than the value most other federal appellate courts had found inconsequential. Instead of the absolute measure, however, the Sixth Circuit Federal Court of Appeals adopted the comparative disparity (CD) measure and found that the magnitude (18%) of underrepresentation did violate the defendant’s Constitutional right to a fair trial. That court also focused on the fact that the comparative underrepresentation was 34.8% on the particular venire from which the defendant’s jury was chosen. The U.S. Supreme Court reversed the Sixth Circuit and upheld the Michigan Supreme Court’s decision. The Court noted that there are advantages and disadvantages associated with the three types of statistical comparisons currently used to examine whether the data reporting the race–ethnic–gender composition of jury pools were consistent with a representative cross section of the community. The Court did not state a preference for any one method, so lower courts can consider all three approaches and decide which is most appropriate in each case.

This paper reviews the various measures and statistical analyses courts have used in jury discrimination cases and compares their merits from a statistical viewpoint. A brief review of the two types of legal challenges is given in Section 2. The next section describes the various statistical measures and tests and illustrates their use on data from several important jury discrimination cases. Section 4 discusses the statistical evidence submitted by the defendant and the lower courts’ interpretation of it. The last subsections are devoted to the arguments made by both parties to the U.S. Supreme Court and the Court’s resolution of the case. The statistical data and analyses submitted in \textit{Berghuis v. Smith} are reexamined in Section 5. Consideration of the power of a statistical test will be seen

---

1. See \textit{Strauder v. West Virginia}, 100 U.S. 303, 310 (1879) (holding that a West Virginia statute that only allowed whites to serve on juries violated the rights of African Americans under the equal protection clause) and \textit{Neal v. Delaware}, 103 U.S. 370 (1880) (finding a denial of equal protection when no African Americans had ever been called for jury service). In \textit{Castaneda v. Partida} 430 U.S. 482 (1977), the Court concluded that a defendant’s equal protection rights can be violated when their minority group is substantially underrepresented in the venires of the jurisdiction. This decision was the first to rely on the results of a formal statistical hypothesis test comparing the minority proportion on venires to their proportion of the population. The analysis is discussed in Section 3.1.


3. 615 N.W.2d 1, 3 (2000) (stating that the defendant’s evidence failed to establish a legally significant disparity under either measure).

4. 543 F. 3d 326 (6th Cir. 2008).
to clarify the usefulness of the “standard deviation” analysis adopted in the Castaneda v. Partida\(^5\) equal protection case in the fair representation context. The two calculations of the “standard deviation analysis” in the record, one by a judge on the Michigan Supreme Court and the other in an amicus brief are questioned. The formula they used did not account for the fact that the number of African Americans on the venires needed to be estimated from the home addresses of the members of the jury pool. In Section 5.6, the data are reanalysed using the correct standard deviation of the statistic comparing the estimated number of minority jurors to its expected value. It yields a statistically significant underrepresentation at the 2-standard deviation level for the time period relevant to the case. After showing underrepresentation, a defendant needs to identify the specific aspect of the jury selection system that led to it. In Berghuis, the defendant suggested that the assignment process of jurors to different courts created the disparity. A statistical comparison of data on jury composition when the old system was in operation to data after it was changed indicates that the effect of the assignment system was not statistically significant, which supports the decisions of the U.S. and Michigan Supreme Courts.

2. **The two types of legal challenges that can be brought by a defendant**

The Sixth Amendment of the U.S. Constitution guarantees a defendant a fair trial and requires states to provide jury pools (venires) that contain a fair and reasonable cross section of the community. The “fair cross-section” requirement is also codified in federal law.\(^6\) The statute states that registered voters are the main source of potential jurors but jury selection plans should utilize additional sources of names when necessary to implement the policy.\(^7\) The seminal cases are Duren v. Missouri\(^8\) and Taylor v. Louisiana\(^9\) that established three criteria a defendant must satisfy in order to establish a *prima facie* case that an important subgroup of the population was underrepresented in jury pools. These are (1) the group alleged to have been excluded is a “distinctive” group in the community, (2) the group’s representation in the source from which juries are selected is not fair and reasonable in relation to the number of such persons in the community, and (3) this underrepresentation results from a systematic exclusion of the group in the jury selection process. Typically, the second requirement is established or rebutted by examining statistical data on the composition of the pool of individuals called for jury service by the state. Statistical analysis may also be helpful in establishing the third requirement when data on the composition of venires over several years show that the same subgroup of the community is consistently underrepresented\(^10\) or that minority representation on juries decreased substantially after a change in the jury selection process. As the Supreme Court

\(^6\) Jury Selection and Service Act of 1968 28 U.S.C. Sec. 1861–1869 (1988). Courts apply the same criteria to check that a jury selection process satisfies the Act or the Sixth Amendment. See *U.S. v. Miller*, 771 F.2d 1219, 1227 (9th Cir. 1985).
\(^8\) 439 U.S. 357 (1979).
\(^10\) Since the individuals called to serve on juries are supposed to be similar to random sample of the population, intuitively one would expect that in any year the probability that a particular subgroup would be underrepresented would be about one-half. Since the individuals chosen in each year or even each quarter can be regarded as independent random samples, it is very unlikely that random selections from the community would yield jury pools having fewer members from the same subgroup of the community in a string of consecutive quarters or years.
noted in *Duren v. Missouri*\(^{11}\) once systematic disproportion is shown, the only remaining question is whether the state can show that the underrepresentation arises is a consequence of advancing a significant state interest.\(^{12}\)

The “Equal Protection” clause of the U.S. Constitution also prohibits purposeful or intentional discrimination by the state. The criteria a defendant needs to satisfy to prevail in this type of case are similar, but not identical to those in a “fair representation” claim because in equal protection cases, the defendant must show that the state intended to discriminate against a distinct group. They are (1) the defendant belongs to a group that can be singled out for discriminatory treatment, (2) members of this group were substantially underrepresented on the venire, and (3) that the venire was selected under a practice providing an opportunity for discrimination.\(^{13}\) While similar statistical evidence is used in the second step of both types of claims, courts may require a greater degree of underrepresentation in the “Equal Protection” context as a discriminatory purpose needs to be inferred from data indicating a discriminatory effect or impact; this is not needed in cases brought under the Sixth Amendment.\(^{14}\) As in “fair representation” cases, statistical evidence supporting an “equal protection” claim should demonstrate meaningful underrepresentation for a reasonable time period. Of course, the state can rebut statistical and other evidence by showing that a discriminatory purpose was not involved. In *Castaneda v. Partida*,\(^{15}\) the Court accepted the use of hypothesis testing to determine whether the proportion of venire members of the defendant’s subgroup was statistically significantly less than their proportion in the jury-eligible population. Since racial discrimination in jury selection also violates the equal protection rights of individuals excluded from possible jury service, in *Powers v. Ohio*\(^{16}\) and *Campbell v. Louisiana v. Campbell*,\(^{17}\) the Court allowed a defendant to assert their rights and challenge the exclusion of jurors of a different race. In both cases, white defendants had legal “standing” to challenge the exclusion of blacks on grand juries (*Campbell*) or by the prosecution’s use of its peremptory challenges (*Powell*).

### 3. Statistical measures and tests used in the evaluation of data on the composition of jury pools

This section describes the statistical measures and procedures that have been used to examine the degree of underrepresentation of a distinct subgroup of the population in jury discrimination cases. They are illustrated on data from *Castaneda*, brought under the equal protection clause, and *Duren*, which concerned fair representation. Then, a few of the conflicting lower court opinions are discussed. Potential problems inherent in each of the different measures and in the available data will be discussed.

#### 3.1 Definitions and application to the data from *Castaneda v. Partida*

The first step in the process is to determine the fraction (\(\pi\)) of the jury-eligible population belonging to the protected subgroup. This fraction is usually obtained from voter registration rolls and/or

---

\(^{11}\) *Supra* n. 8 at 367–368.

\(^{12}\) Ibid. at 370 where the opinion indicates that reasonable exemptions based on special hardship, incapacity or community needs would not substantially reduce the representativeness of the remaining pool of jurors.

\(^{13}\) See *U.S. v. Grisham*, 63 F.3d 1074, 1081 (5th Cir. 1995).

\(^{14}\) *Duren*, 439 U.S. at 368 n. 26. See Beale (1983) for a discussion of the interplay between statistical methodology and legal theory underlying the claim.

\(^{15}\) 430 U.S. 482 (1977).


A REANALYSIS OF THE DATA IN BERGHUIS V. SMITH

Census data. Some courts consider the minority fraction of the general population but most refine the Census data by eliminating people less than 18 years of age and senior citizens in jurisdictions where they can request to be excused. Often, the composition of venires or juries for several years are examined with data on the total number \( n \) of individuals called for jury service and the number \( x \) of them belonging to the subgroup in question.\(^\text{18}\) The minority fraction, \( p = x/n \), of the jury pool is then compared to their fraction, \( \pi \), of the jury-eligible population. In Castaneda, several values of the Mexican American fraction \( (\pi) \) were considered. They formed 79.1% of the general population, 72% of the adult population and 65% of the population over 25 with some schooling.\(^\text{19}\) Both the 79.1% and 65% figures will be used to illustrate the various measures. During the relevant 11-year period, there were 338 Mexican Americans out of 870 grand jurors chosen to serve by the County. In 2.5 of the 11 years, when the same state district court judge supervised the process, 100 of the 220 grand jurors were Mexican American.

The AD is the difference \( p - \pi \). For the full data set, \( p = 338/870 = 0.3885 \) or 0.39. If one uses the Mexican American proportion of the total population in the county, \( \pi = 0.791 \) and the AD = -0.403 indicating that Mexican Americans received 40% fewer appointments to the grand jury than expected. If one uses the 65% figure, which included age and education, \( \pi = 0.65 \) and the AD = -0.261, which implies that Mexican Americans received 26% fewer grand jury appointments than expected. For the smaller data set \( = 100/220 = 0.455 \). The corresponding values of the AD are \(-0.336 \) when \( \pi = 0.791 \) and \(-0.195 \) when \( \pi = 0.65 \).

Many courts have required plaintiffs to demonstrate an AD of at least 0.10 (courts ignore the sign as it is always negative in these cases). This standard has been criticized because it allows a jurisdiction to totally exclude a small distinct subgroup forming less than 10% of the age-eligible population. Although this problem is well known,\(^\text{20}\) some courts rigorously adhere to it\(^\text{21}\) even when the subgroup is small.\(^\text{22}\)

The absolute impact (AI) measure is the number of additional jurors from the subgroup who would have been in the pool of jurors or grand jurors had the jury selections reflected the eligible population. It is obtained by multiplying the AD by the number of jurors serving on the relevant panels. For the Castaneda data, when \( \pi = 0.791 \), the AI = 0.403 \times 870 = 351 for the 11-year period and 0.261 \times 220 = 57.4 or 57 for the shorter period. Using the more restrictive value, 0.65, for \( \pi \) yields an AI of 169.7 = 170 for the 11-year period and 42.9 or 43 for the shorter period. In statistical terms, the AI is the difference between the actual number of minority jurors and the

\(^{18}\) There are a few cases where the fair representation of several subgroups of the population was of concern. For example, in United States v. Biaggi, 909 F.2d 662 (2d. Cir. 1990), one defendant argued that the jury selection process discriminated against African Americans and Hispanics.

\(^{19}\) The majority opinion by Justice Blackmun adopted the 79.1% figure but also showed that using the 65% figure would not change the ultimate statistical conclusion. The 65% figure incorporated some factors suggested by Chief Justice Rehnquist in his dissent, 430 U.S. 505–506.

\(^{20}\) This problem is noted in a report of Judge Gewin, reproduced in an Appendix to Foster v. Sparks, 506 F.2d. 805 (1975). Many commentators have also criticized the use of the AD because of this problem; see Kairys \textit{et al.} (1977) at 793–794, Gastwirth (1988) at 156, Williams (1990) and Detre (1994).

\(^{21}\) See United States v. Grisham, 63 F. 3d 1074, 1078–1079 (11th Cir. 1995) and United States v. Carmichael, 560 F. 3d 1270 (11th Cir. 2009).

\(^{22}\) The AD was considered in United States v. Orange, 447 F.3d 792 (10th Cir. 2006) even though each of the four subgroups, African Americans, American Indian, Asian and Hispanic, composed less than 10% of the voting age population of the district. The opinion, Ibid. 798 n. 7, cites cases from several other circuits that approved of systems with ADs in the range of 2–11%.
number expected if they were chosen like a random sample of the eligible population.\(^{23}\) Thus, when \(\pi = 0.791\), out of 870 jurors, one expects 79.1% or 688 to be Mexican Americans. When \(\pi = 0.65\), one expects 566 to be Mexican American.

An obvious problem with the AI is that it depends on the size of the venire or venires under scrutiny.\(^{24}\) Thus, a standardized version, which divides the AI by the expected number of minority jurors, is also used. While the AD measures the disparity in absolute terms, this standardization expresses the minority shortfall as a percentage loss of venire appointments. It is often referred to as the CD (Beale, 1983; Detre, 1994) or percentage shortfall. For the Castaneda data, when \(\pi = 0.791\), CD = \((338 - 688)/688 = -0.509\), implying that the Mexican American community of Hidalgo County served as jurors slightly less than half their expected number of times. When \(\pi = 0.65\), the CD becomes \((338 - 566)/566 = -0.403\). An alternative formula for the CD is \(AD/\pi\) or \((p - \pi)/\pi\).

Thus, when \(\pi = 0.791\) and \(p = 0.3885\) the CD = \(-0.403/0.791 = -0.509\).\(^{25}\)

The statistical measure used to assess the disparate impact of an employment practice is the selection ratio (SR); the ratio of the probability an eligible member of the distinct subgroup is in the jury pool or venires under study to that of an eligible majority person. This ratio measure is the well-studied relative risk, used in epidemiology and public health. In our context, it is calculated as the ratio of the odds of choosing a member of the distinct subgroup from the jury-eligible pool, i.e. \(p/(1 - p)\) to the odds, \(\pi/(1 - \pi)\) of choosing a member of the distinct subgroup from the actual venire(s) or juror(s), i.e.

\[
SR = \frac{p/(1 - p)}{\pi/(1 - \pi)}.
\]  

(1)

Apparently, an equivalent measure was proposed by the U.S. Civil Rights Commission when it suggested that a disparity of 20% or more between the proportion of eligible whites selected for the master jury wheel and the proportion of eligible minority persons selected be remedied by supplementing the source list.\(^{26}\) The SR estimates the ratio of the probability a jury-eligible minority member has the opportunity to serve on a jury to the probability a similar majority member can serve on a jury. Kaye and Zeisel (1997, p. 180) note that this quantity is of greater legal interest than the AD. Like the CD, SR is a relative measure. They differ because the CD measures the diminished probability a minority member has of being on a venire compared to that of a member of the overall population; the SR compares the probability a minority member has of participating on a jury to that of a majority individual. If a minority forms a relatively sizeable percentage of the population,

\(^{23}\) Suppose the fraction, \(\pi\) of a large population belongs to a subgroup, then the probability that a randomly selected member of the population is from the subgroup is \(\pi\). If a sample of size \(n\) is randomly chosen, then we expect \(n\) times \(\pi\) to be from the subgroup.

\(^{24}\) Beale (1983, p. 280) observes that decisions indicating that a change of one or two panel members is not meaningful have not explained their reasons. She notes that the views of minority jurors might affect the others and that even one juror can have an impact in criminal cases requiring a unanimous vote to convict.

\(^{25}\) The slight difference in the fourth decimal place of the two calculations of the CD when \(\pi = 0.791\) is due to rounding.

\(^{26}\) See Foster v. Sparks, 506 F.2d. 805, 818 (1975). The measure the Commission proposed is described by Kairys et al. (1977) at 791 as the proportion of eligible’s standard. Routine algebra shows that the measure proposed by the Commission is \(1 - SR\), which is why they suggested that the minority shortfall is less than 20%. This suggestion is consistent with an SR of at least 80%, mirroring the statistical criterion used in disparate impact cases.
e.g. 30–40%, its low rate of participation will also lower the rate of the overall population, thereby decreasing the impact of minority underrepresentation as measured by the CD.\footnote{27}

The measures SR and CD are relative measures and the CD has been criticized because it appears to overstate the disparity when the underrepresented group forms a very small percentage of the population.\footnote{28} The point is often illustrated by considering a community with only one jury-eligible minority member out of 100,000. A wheel of 1000 is taken at random and the minority individual is not on it. Even though \( \pi \), the minority fraction of the eligible population is 0.00001, both relative measures equal zero because \( p = 0 \) as the one minority member is not on the wheel. While this example demonstrates that it is difficult to prove discrimination against an extremely small minority group in a community, most cases concern minorities forming at least half of 1% of the local population. Since courts consider the composition of the jury pool over a period of time, by increasing the size of the sample examined (increasing the number of venires or years examined) often one will be able to detect substantial underrepresentation of small, but distinct subgroups.\footnote{29}

In \textit{Duren}, the Court repeated its statement in \textit{Taylor} that states were free to prescribe relevant qualifications for jurors and provide reasonable exemptions as long as the jury lists or panels were representative of the community.\footnote{30} In almost every basic statistics text, students learn that the most reliable way of obtaining a representative panel of a prespecified population is to take a simple random sample. This is accomplished by ensuring that every member of that population has the same chance of being in the sample; i.e. the identical probability of being selected.\footnote{31}

For the \textit{Castaneda} data, when \( \pi = 0.791 \) and \( p = 0.3885 \), SR = 0.6353/3.7847 = 0.168. Thus, the probability a Mexican American residing in Hidalgo served on a jury during the 11-year period was about one-sixth that of a white. Even when one uses the more conservative value of .65 for the Mexican-American fraction of individuals meeting the age and literacy requirements, the SR = 0.6353/1.8571 = 0.342. This implies that the probability a Mexican-American eligible for jury service served during the period was only one-third that of a similar white. In equal employment

\footnote{27} For example, consider a jurisdiction with 40,000 eligible minority members and 60,000 whites. Six hundred (1%) of the whites are selected to be on the venires for the time period but only half of 1% or 200 minority members are. In the whole population, 800 or 0.8% are on the venires. The CD measure equals (200–320)/320 = −37.5%, suggesting that the minority received 62.5% of its expected number of venire members. The SR is 50%, indicating that a minority only had one-half the chance of serving as a white.

\footnote{28} See \textit{United States v. Hafen}, 726, F.2d 21, 24 (1st Cir. 1984) and the discussion in \textit{Beale} (1983, p. 274).

\footnote{29} In the situation where a minority group forms one-half of 1% of the eligible pool, if one examines venires with a total of 1000 individuals one expects to observe five minority members. The only outcome reaching statistical significance at the standard 0.05 level of a two-tailed test is observing no minorities as the probability of one tail, observing only one or no minority potential jurors already is 0.04. Yet observing only one minority when five are expected suggests that the jury selection system should be examined further. By increasing the sample to 2000, 10 members from that small subgroup are expected and the statistical test will classify pools with four or fewer as indicating a statistically significant shortfall. More precisely, the Poisson approximation to the binomial model yields a probability just under 0.03 of observing four or fewer minorities in a sample of 2000. Suppose that only two minorities appear among the 2000 individuals on the venires chosen from the eligible 100,000. The probability that this would occur in a random sample of the eligible pool is only 0.003, clearly indicating that chance is unlikely to produce the result. The CD equals (2 − 10)/10 = −80\% and the SR is just under 20\%, indicating that minority members had only one-fifth the chance of serving on a jury as whites. Such a severe underrepresentation of a small minority group will never be detected by setting a threshold of 10 or even 5\% under the AD criterion; statistical hypothesis testing of a sample of appropriate size and the relative measures indicate such a jury selection system should receive judicial scrutiny.

\footnote{30} 439 U.S. 357, 367–378, citing \textit{Taylor} 419 U.S. 538. The Court went on to emphasize that the aspects of a jury selection process that result in a disproportionate exclusion of a distinctive group need to be justified by a showing that they seriously advanced a significant state interest.

\footnote{31} See \textit{Weiss} (2005) at 11.
cases, when the SR of the minority pass rate to the majority pass rate is less than 0.80, the test or job requirement is considered to have a disparate impact and the employer needs to show it is a valid predictor of successful job performance. Thus, whichever minority fraction of the eligible jury population one adopts (the majority’s 0.791 or the dissent’s 0.65), the SR is very low and indicates that the jury selection process in the County had a strong disparate effect on Mexican Americans.

The last approach often adopted by courts is to formally test the hypothesis that the observed jury composition is consistent with a random sample of size \( n \) (the number of individuals on the venires or juries examined) from a population in which a fraction, \( \pi \), belong to the distinct subgroup in question. The \( p \) value or probability of obtaining a sample which is ‘at least as far from what is expected’ in a random sample indicates that the proportion of minorities in the data is inconsistent with the hypothesis when the \( p \) value is small (e.g. less than 0.05 or 0.01). The test statistic is based on the AI, the difference between the number \( x \) of minority jurors and the expected number \( n\pi \) when the hypothesis is true. Rather than standardizing \( x - n\pi \) by \( n\pi \) as in the CD measure this difference is measured in units of the variation from expected that typically occurs in random samples. Technically, these units are standard deviations, where 1 standard deviation is \( \sqrt{n\pi(1-\pi)} \).

Thus, the normal form of the test statistic is as follows:

\[
Z = \frac{x - n\pi}{\sqrt{n\pi(1-\pi)}}. \tag{2}
\]

When the sample size, \( n \), is reasonably large and \( \pi \) is not very small, the statistic \( Z \) is essentially a standard normal (bell-curve) distribution and the \( p \) value is obtained from standard tables. A \( Z \) value of at least +2 or less than −2 occurs about 5% of the time in random samples, while a value of at least +2.6 or less than −2.6 occurs about 1% of the time. In Castaneda, the Court noted that values of \( Z \) between 2 and 3 standard deviations are suspect to a social scientist.

Using the dissent’s value of 0.645 for the Mexican American fraction of eligible jurors and applying formula (2) to the data in Castaneda yields

\[
Z = \frac{338 - (0.645 \times 870)}{\sqrt{0.645 \times 0.355 \times 870}} = \frac{226.65}{14.1} = -15.8.
\]

A shortfall of over 15 standard deviations corresponds to a probability of less than one in a billion, clearly indicating a significant shortfall of Mexican Americans on Hidalgo County jury panels.

Comments: 1) A potential problem with the “2–3” standard deviation statistical criterion occurs in very large samples. When the minority percentage of a master list or of many venires containing thousands of names, is compared to their fraction of age eligible members of the population, almost always a statistically significant disparity will be found. 2) Using statistical testing in conjunction

\[32\] The application of the method to data in jury discrimination cases was suggested in a classic article by Finkelstein (1966) and is now the commonly accepted approach to examine data on the composition of venires (Kaye, 1985; Gastwirth, 1988; Detre, 1994).

\[33\] This issue is noted by Detre (1994). The problem arises because the standard deviation of the proportion of minorities in a sample of size \( n \) from a large population is \( \sqrt{\pi(1-\pi)/n} \). Here, \( \pi \), is the minority fraction of the population. Notice that if \( \pi = 0.10 \) and \( n = 4000 \), then 1 standard deviation equals 0.0047. Even if the minority proportion in the jury wheel was 0.09, the difference between the two percentages (10% in the population and 9% in the wheel) would be statistically significant at the 2 standard deviation or 0.05 level of statistical significance. Practically speaking, however, all the measures indicate a very modest difference (SR = 89%, CD = 10%, AD = 0.01 or 1%) and courts might well decide that disparities of this level are minimal. The relationship between statistical significance and sample size is discussed by Peresie (2009, at 787–790). As shown in Section 5.1, when the number, \( n \), of potential jurors is small, it may not be possible to detect a legally meaningful shortfall in minority panelists.
with a meaningful measure of the practical impact, e.g. the SR or CD should be less than a prede-
determined value, e.g. 0.80 or 0.85, will assist courts in determining when a statistically significant
difference between the minority fractions of the jury wheel and age-eligible population is of legal
importance. A similar recommendation in the context of disparate impact cases was recently made
by Peresie (2009).

2) If the authorities began with a master list of jury eligible members of the population, it would
be easy to check whether the resulting venires were consistent with a random sample. While the
Census data provide the age-eligible population, categorized by gender, race and ethnicity, it is prac-
tically impossible to remove those individuals in the area who are not eligible for jury service as the
Census does not collect information on felony convictions or level of ability to speak and understand
English. Thus, the venires in a fairly run jury selection process will not perfectly mirror the Cen-
sus data. It would be preferable to compare the demographic composition of the jury pool before
individuals who are not eligible for service or qualify for a legitimate exemption are removed to the
Census data. The authors have not come across a case where this type of data have been preserved
or analysed. With modern technology, the demographic composition of individuals to whom jury
questionnaires are sent or that of the respondents could be obtained and compared with the Census
data.

3.2 Application of the methods to the data from the precedential fair representation case: Duren v.
Missouri

Although the total exclusion of women from juries was found to violate a defendant’s Sixth Amend-
ment right to a fair trial in the 1940s, the issue of their underrepresentation on jury panels result-
ning from barriers to their serving created by the states was not settled by the Court until Taylor v.
Louisiana. The law in Louisiana provided that a woman should not be selected for jury service un-
less she had previously filed a written declaration expressing her desire to be subject to jury service.
No such requirement was made of men. In Taylor, women formed 53% of the jury-eligible popula-
tion but of the 1800 persons drawn to fill petit juries in St. Tammany parish during the period of the
trial, only 12 were women. No woman was on the defendant’s venire of 175. While finding that the
near total exclusion of women violated the defendant’s right to a jury drawn from a fair cross section
of the community, the Court in Taylor also said that the states can prescribe relevant qualifications
for their jurors and provide reasonable exemptions “so long as it may be fairly said that the jury lists
or panels are representative of community.”

The first case that did not involve nearly complete exclusion is Duren. The petitioner contended
that his right to a trial by a jury chosen from a fair cross section of the community was denied by
provisions of the Missouri law that granted all women an automatic exemption from jury service if
they requested it. Only men over 65 were entitled by the law to request a similar exemption. The jury

34 Pollard (2000, at 219) notes that it is difficult to determine competency in English from the available Census data and
Kaye and Zeisel (1997, at 183) observe that valid statutory disqualifications, such as conviction of a felony, are more prevalent
in some subgroups of the population.
36 See Glasser v. United States 315 U.S. 60, 85–86 (1942) and Ballard v. United States, 329 U.S. 187 (1946). In both
cases, the defendants were males who argued that excluding women violated their right to a jury that fairly represented the
community.
38 See 439 U.S. 357, 361-364 (1979) for the information summarized in this paragraph.
selection process in Jackson County at the time (1975) began with the mailing of a questionnaire to randomly selected persons on the voter registration list. The questionnaire contained a list of occupations and other categories that disqualified or exempted one from jury service. In particular, the questionnaire allowed anyone over 65 to elect not to serve. Women who did not indicate that they were willing to serve on juries were assumed to have chosen not to serve. The 1970 Census data indicated that 54% of the adults in the County were women. During the periods of June to October 1975 and January to March 1976, 11,197 persons were summoned for jury service, 2,992 or 26.7% were women. Of those summoned, 741 women of a total of 5,119 people appeared for service. Thus, women formed 14.55% of the postsummons weekly venires during the period when the defendant’s jury was chosen. In March 1976 (the time of the trial), 453 of the 1,537 or 29.5% of those summoned for jury duty were women and of the 707 appearing 110 or 15.5% were women. The petitioner’s jury was selected from a venire of 53 that included five women; however, the final jury was all male.

The Missouri Supreme Court opinion\(^\text{39}\) noted that information on the number of women who claimed an exemption and that many of the occupations, e.g. school teachers who could be excused were predominantly female. It noted that the female percentage, 15.5%, of the venires during the period clearly exceeded the less than 1% in Taylor and found that the system did not violate either the “equal protection” clause or the Sixth Amendment. In an interesting dissent,\(^\text{40}\) Judge Seiler noted that Jackson County gave women two opportunities to be excused from jury service. First, when they received the screening questionnaire and a second time when they were summoned. Consequently, women formed no more than 15% of any of the five venires of the criminal trials held from April 1975 to March 1976. In contrast, women formed just under 40% of the venires of the federal court for the Western Division of the state. That court did not automatically exempt women but did allow women taking care of minor children without adequate domestic help to be excused. In that division, women formed 53% of the master jury wheel and 39.8% of the actual jurors.

The U.S. Supreme Court found that the disproportionate and consistent of exclusion of women from those summoned and again at the venire stage was obviously due to Missouri’s system. The opinion emphasized that once the defendant has made out a \textit{prima facie} case showing his constitutional right to a fair jury is infringed, the state bears the burden of justifying the system by showing that the attainment of a fair cross section is incompatible with a significant state interest. It also needs to demonstrate that the underrepresentation arose from the exemptions related to the state interest. The Court noted that the state had not offered a substantial justification for the automatic exemption of women.\(^\text{41}\)

To quantify the underrepresentation of women, their 26.7% of those summoned will be compared with their age-eligible percentage (54%). The \(\text{AD} = 27.3\%\), the \(\text{CD} = 50.6\%\) and the \(\text{SR} = 31\%\); i.e. women had just under one-third the probability of being summoned for jury service as men. Thus, the first stage of the jury selection process already excluded a disproportionate fraction of the female population of the County. Evaluating the overall process, which led to women forming only 14.55% of the venires during the period, yields \(\text{AD} = 39.45\%\), \(\text{CD} = 73.1\%\) and \(\text{SR} = 0.145\). Thus, a jury-eligible female only had one-seventh the probability of participating in a venire as a similar male. Following the dissent in the State Supreme Court, these values should be compared to the corresponding ones for the federal court in the area (\(\pi = 0.53, p = 0.398\)). For those data,

\(^{39}\) See 556 S.W. 11, 16–18.

\(^{40}\) Ibid. at 23–25.

\(^{41}\) The material in this paragraph comes from \textit{Duren}, 439 U.S. at 367–369.
AD = 13.2%, CD = 24.9% and SR = 58.6%. These figures refer to actual jurors, rather than the venires, so some of the disparity may have arisen when potential jurors were excused for cause or peremptorily challenged. Since women who needed to take care of young children were excused, these measures may provide an idea of the potential impact of a legitimate justification on the gender composition of venires. Of course, the effect of excusing women with care taking obligations will be smaller in cases concerning the fairness of the racial or ethnic composition of venires than in cases relating to gender. Since minority groups tend to have more children and a higher prevalence of single mothers, a legitimate process of excusing women with child care responsibility might well lead to an SR less than 100%.  

3.3 Application of the measures and tests to data from other federal cases

While the AD is simple to calculate, an early use of it in *Swain v. Alabama* where the Court said that the AD needs to be in the 10–15% range in order to establish a *prima facie* case that the Constitutional right of the defendant have been violated received substantial criticism from commentators. They emphasize that a minority group composing less than 10% of the community could be systematically excluded from serving on juries if lower courts rigidly adhered to that requirement. In the equal protection context, the Court’s use of statistical hypothesis testing to assess the shortfall in minority jurors in *Castaneda v. Partida* essentially over-ruled that part of *Swain*. The briefs of the parties cite many cases concerning the circumstances in which the AD or CD is more appropriate.

An illuminating discussion of the different criteria and their use in both types of cases is given by the late Judge Kaufman’s opinion in *Alston v. Manson*. The two defendants contended that the jury selection scheme evidenced intent to discriminate against blacks. In Connecticut, each county compiles its jury arrays by obtaining potential jurors from its towns but the law favours smaller towns. The smallest town in New Haven County, Beacon Falls, contributed 4.2% of its adult population to the array, while the city of New Haven contributed only 1.1% of its adults. Since minorities tend to live in urban areas of the state, restricting the fraction of the array chosen from New Haven would lead to their underrepresentation. The opinion notes that African Americans formed 5.96% of the jury-eligible population. Out of an array of 8405, only 368 or 4.38% were African Americans.

The state argued that the AD of 1.58% should not be considered legally meaningful. The defendant’s statistician conducted a standard deviation analysis, presumably based on formula (1). The number of African Americans on the array has expected value 500.94, with a standard deviation of 21.704. Thus, the difference between the observed number (368) of African Americans and their expected number is −6.13 standard deviation units; clearly exceeding the 2–3 standard deviations criteria of *Castaneda*. Thus, the court concluded that the statistical evidence supported defendant’s claim that their “equal protection” rights under the 14th amendment were violated. It should be

---

42 An approximate calculation of the potential effect of child care responsibilities will be discussed in the context of the *Berghuis* case in Section 5.5. Because single heads of households with small children do not form the majority of the jury eligible population, by itself this factor is unlikely to create a very low SR but might reduce an SR of 0.85 for a minority racial or ethnic group to 0.80 or so.

43 380 U.S. 202 (1965)

44 See n. 20, Finkelstein (1966) at 388, Sperlich and Jaspovice (1979); see supra n. 29 for a numerical example. See DeCani (1974) for a discussion of how a statistician would have presented the results of a formal test of the fairness of the juries in *Swain*.

45 791 F. 2d 255 (2d. Cir. 1986).

46 Ibid at 258.
J. L. GASTWIRTH AND Q. PAN

noted that data in Alston have a CD = 26.5% and a SR = 0.722. Thus, the probability a jury-eligible African American member had of being selected for the array is less than three-fourths that of a white; below the 80% criterion advocated by the U.S. Civil Rights Commission.

The Alston opinion noted an apparent discrepancy between its analysis of the data under the “equal protection” clause and a previous decision, United States v. Jenkins47 in which it evaluated a “fair representation” claim by considering the AI of the underrepresentation on a typical jury pool of 60. In Jenkins, the City of New Haven, CT obtained its jurors by sending questionnaires to a sample of registered voters. This resulted in an African American representation of 3.3% of the jury pools over a 4-year period in contrast with their 5.45% of the adult population of the city. While the opinion realized that this difference appeared meaningful on a ratio scale (CD = −39.4, SR = 0.592), it observed that taking steps to supplement the voter roles to reduce the AD of 2.15% would result in only one more African American on a typical venire (as 2.15% of 60 = 1.29). Apparently, a standard deviation analysis was not submitted.

The court in Jenkins expressed concern that disparities similar to the one in that case would exist for at least one of the multitude of subgroups of the population in almost any federal district in the country.48 From a statistical view, this is a legitimate concern because the theory of hypothesis testing assumes the hypothesis being tested is specified before one examines the data. The Jenkins opinion realized that a defendant (or their lawyer) would see their venire and probably be aware of the race–ethnic–gender composition of other venires. After noticing that one or two subgroups of the general population were underrepresented, the defendant could allege the venires were not representative of the community. The usual statistical tests would not be valid because the defendant selected the subgroup(s) that had the minimum number(s) relative to their numbers expected in a random sample after looking at the data.49 When the number of comparisons that are made, either explicitly or implicitly, is known, one can correct the statistical test.50 In most “equal protection” cases, however, the defendant asserts that members of their own race, ethnic group or gender are underrepresented. Then, the hypothesis being tested is whether the proportion of venire belonging to the defendant’s subgroup equals their proportion (π) of the jury-eligible population obtained from an appropriate data source; and alternative (the subgroup is underrepresented) is clear.51 Furthermore, the number of possible comparisons is limited because not all “subgroups” of the population are deemed “cognizable.” Groups defined by race, ethnicity or religion are considered distinct.52 However, young adults, poor people and non-English speakers are not.53

47 496 F. 2d 57, 66(2nd Cir. 1974)
48 See 496 F. 2d 57 at 65, 66.
49 Statistically, one should be using the minimum of the Z statistic in (1) for all the distinct subgroups in the area. For example, in New York City, a defendant could also question the representation of Italian Americans, Asians, women, Catholics or Jews in addition to African Americans and Hispanics.
50 The simplest correction, called the Bonferroni adjustment, just multiplies the p value of the test for one group by the number of comparisons.
51 Since the Court in Powers v. Ohio, supra n. 16, allows a minority defendant to allege underrepresentation of a majority group as well as their own, the use of two-sided alternative as in Castaneda covers both possible scenarios.
52 See United States v. Gelb, 881 F.2d 1151, 1161 (2d Cir. 1989) (holding that Jews are a distinctive group because they formed a significant portion of the population) and United States v. Osorio, 801 F.Supp. 966, 977 (D. Conn. 1992) (noting that both African Americans and Hispanics are “distinctive” groups).
53 See United States v. Pichay (986 F. 2d 1239 (9th Cir. 1993) (not recognizing poor people), United States v. Fletcher, 965 F. 2d 781,782 (9th Cir. 1992) (not recognizing college students) and Hernandez v. United States 500 U.S. 352 (1991) (not recognizing people who do not speak English). People with long hair and a beard were also not considered a distinct group in Purkett v. Elem, 514 U.S. 765, 769 (1005) (per curium).
Subsequent cases in the Second Circuit have followed Jenkins rather than Alston, especially when the source of the master jury list were names from voter registration and driver’s license files, which are apparently race neutral. In United States v. Biaggi, a system with an AD of 3.6% for Blacks and 4.4% for Hispanics was approved even though the standard deviation analysis showed that the difference between the age-eligible population and composition of the jury wheel was statistically significant. The trial court in Biaggi noted that the AI on minorities in a typical venire of 50–60 implied only a shortfall of two to four African Americans and three or four Hispanics. Comment: The data in Biaggi illustrate a major disadvantage of the AI measure. If two or three more minority members are on the venire, it will be more difficult for the prosecution to eliminate all of them by peremptory challenges.

4. The Berghuis v. Smith case

4.1 Background

Mr. Smith was charged with several felony offenses related to a shooting at a bar in Grand Rapids, Michigan on 7 November 1991. He was convicted of second degree murder and one count of possessing a firearm during the commission of a felony in October 1993 but found not guilty of related charges involving an assault during a fight. Before the jury was sworn, the defendant claimed that the jury array, containing only 2 or 3 black jurors out of about 60 was unfair and challenged the venire. The trial judge rejected that claim and the trial continued. Ultimately, 37 individuals, all of whom were white, were considered for the jury and 14 (12 + 2 alternates) were seated.

The master jury list in Kent County was not based on voter registrations but combined lists of holders of driver’s licenses or identification cards issued by the Department of Motor Vehicles (DMV) to nondrivers. The potential jurors for the October 1992 to October 1993 term were obtained by the Secretary of State, who took a simple random sample of about 50,000 names from a master list of 350,000 residents of the County. The DMV file did not identify race. These individuals were sent a questionnaire that would determine whether they were eligible for jury service and also allowed them to claim a statutory exemption. About 5% of the questionnaires were returned by the Post Office as undeliverable and about 15–20% were not answered. A follow-up letter was sent to those nonrespondents; about 50% of whom then returned the questionnaires. Apparently, the areas with the highest rate of nonresponse were those with large minority populations.

The number of individuals on the “qualified list” after this first phase is not reported. The county estimated the likely needs of both the circuit (felony) and the district (misdemeanor) courts and randomly selected 12,000

---

56 Recall that the prosecutor only needs to provide a race-neutral justification for challenging a juror. Courts often examine whether similar majority members were also challenged. See Love v. Yates, 586 F. Supp. 2d 1155, 1175–1176 (N.D. Ca. 2008) finding discrimination when the prosecution challenged five of eight Blacks, while several non-Blacks who had given similar answers to the questionnaire as the Blacks who had been removed by the prosecutor were not challenged. Statistical procedures for analyzing peremptory challenge data are discussed by Barrett (2007) and Gastwirth (2005).
57 According to the Brief for the Petitioner at 5, the trial judge requested 60 jurors and thought three were African American. The defendant’s attorney estimated that the venire consisted of nearly 100 people, 2 or 3 of whom were African Americans.
58 Ibid. at 7–8, individuals were exempt if they were not a resident of the county, could not speak English or were physically or mentally unable to carry out the functions of a juror, had served on a jury during the previous year or were currently under sentence for a felony. Individuals over the age of 70 could request an exemption.
59 543 F.3d at 331.
individuals for a “second list.” These people received a summons and a personal history questionnaire. At this stage, prospective jurors could request to be excused for reasons of hardship, e.g. child care concerns, transportation problems or inability to take time off from their job. While the names of individuals placed on the potential jury list or excused from serving were kept by County, their racial identification was not.\footnote{While the original master list provided by the Secretary of State did not identify a person’s race, the County could have obtained this information on individuals who served or were excused.}

The defendant’s statistical analysis, described in more detail in the next subsection indicated that African Americans formed 7.28% of the age-eligible residents of the County; however, only 6.03% of the 929 members of jury pools during the 6-month period that included the trial were expected to be African American.\footnote{Because the precise number of African American members in the venires was not available, it was estimated from their proportion in the Census tracts in which the venire members resided. The calculated percentage varies depending whether one uses the expected number 55.4 or rounds it up to 56 or down to 55. The Sixth Circuit opinion, supra n. 4 at 337 uses 56.} A greater underrepresentation occurred on the particular venire from which the jurors in the defendant’s trial were chosen (a shortfall of 4 out of 11.5 expected if the venire was a random sample of the eligible population). In the following year, African Americans formed 6.17% of the 2252 individuals on the venires; again, below their percentage in the eligible community. The defendant suggested two aspects of the jury selection system created this shortfall.

First, during the time of defendant’s trial, members of the jury pool were first assigned to local courts before being sent for possible jury duty in the circuit court, which hears felony cases. Because the majority of African Americans lived in Grand Rapids, the largest city, the defendant claimed that this process led them to be siphoned off to the city courts, which had a large volume of cases.\footnote{Mr. Foster, the administrator of circuit court, stated that the visible number of minorities on the jury arrays seemed low (2 February 1998 Motion Tr. at 27–38, J.A. 25a) and that it was believed this was due to giving first priority to the district courts, Ibid. at 30). Mr. R. Hillary, a lawyer in the Public Defender’s Office, also testified that there were few black potential jurors (Ibid. at 129–137, J.A. 60a). He also discussed a study conducted by the City High School that indicated economic reasons such as lack of child care or transportation contributed to the higher proportion of unreturned questionnaires by the minority community.} While statistical support for this assertion was not submitted, two witnesses testified that during the relevant time period the percentage of minorities on the panels was low.\footnote{At the hearing, an expert demographer presented the Census statistics (J.A. at 79a–86a). The 1990 Census data showed that 64% of African American families with children are headed by single parents, while only 19% of white families are. The poverty rate in the County was 6.7%, while 31.5% of African Americans were in poverty. Only 27% of whites were renters, in contrast with 59% of African Americans. Finally, about 20% of the whites had moved in the last 15 months, while 32% of the African Americans had. This might lead to a higher proportion of questionnaires mailed to African Americans to be returned by the Post Office.} The administrator changed the process of first sending individuals to the local courts in October 1993. Second African Americans were far more likely to be single heads of households, live in poverty, rent their home, move in the last 15 months and not have an automobile than whites.\footnote{Answer Opposing Petition for Certiorari at 14.}

Thus, allowing such individuals to request to be excused because of the hardship jury service would reduce the potential pool of minority jurors.

While the defendant argues that excusing potential jurors for reason of hardship contributed to the underrepresentation of minorities on venires, the state asserted that any disparity resulting from economic or social factors were unrelated to the jury selection system. The state also cited a number of cases where courts accepted jury systems with AD or CD larger than those (AD = 1.28% and CD = 18%) here.
Before describing how the courts weighed the statistical evidence in the expert report, the data and analysis will be summarized in some detail. Because the race of individuals with a driver’s license or state identification card was not recorded, the statistician needed to estimate the African American proportion of the venires.

4.2 The defendant’s statistical analysis

It is useful to index each tract by \( i \), where \( i \) goes from 1 to 112. In order to estimate the number of African Americans serving on each venire, the expert calculated the minority proportion of age-eligible individuals living in each census tract from the 1990 Census. Since the tract where a venire member resides can be determined from their address, the number of venire members from each of the 112 tracts in Kent County can be obtained and the expected number of minorities calculated.

The method is similar to one used to estimate minority availability in equal employment cases, where subareas of the labour market closer to the worksite receive a higher weight, which is determined by the proportion of applicants from the area. The method of assigning race to DMV information using census tract data was successful in obtaining an appropriate control group in an epidemiologic study (Karami et al., 2010). From the Census data, one knows the total population \( N_i \) of age-eligible residents of the tract and the number \( B_i \) who are African American. Then, the fraction \( \pi_i = B_i / N_i \) denotes the fraction of potential jurors residing in the \( i \)th tract who are African American. The total age-eligible population, \( N \), in the entire county is the sum of the \( N_i \) over all 112 tracts. For each venire or for any time period, suppose \( n_i \) jurors came from the \( i \)th tract. Then, we expect \( n_i \pi_i \) to be African American. The jury pool can be regarded as a sample of \( n \) (\( n = 929 \) in the first period and 2252 in the second) from the age-eligible population. Since it is known that \( n_i \) of the \( n \) potential jurors come from the \( i \)th tract, the expected number of African Americans among the \( n \) venire members equals

\[
\sum_{i=1}^{112} n_i \pi_i .
\]  

If all the tracts are represented in the sample of potential jurors in accordance to their fraction, \( N_i / N \) of the population, then \( \sum_{i=1}^{112} n_i \pi_i = n \pi \), the expected number of African Americans in a random sample of the entire county. To assess whether African Americans were fairly represented on the venires, one can examine whether residents of the tracts they lived in were underrepresented in among the \( n \) venire members. Thus, the defendant’s expert compared the expected number of African Americans derived from formula (3) to the expected number, \( n \pi \), in a random sample from the county. The comparison considered by the expert estimates the AD by

\[
\sum_{i=1}^{112} n_i \pi_i - n \pi
\]  

65 If a tract contributes 20 members of the jury pool and minorities form 10% of the age-eligible population, then the expected number of the 20 members from the tract who are minority is two. More formally, the number of minority members follows a binomial distribution (\( n = 20, \pi = 0.1 \)) and one expects that \( n \pi = 2 \) are minority.

66 See Markey v. Tenneco Oil Co., 707 F. 2d 172 (5th Cir. 1983). The data and calculations are given in Gastwirth (1988, p. 69–70).

67 Recall that \( n \pi \) is the expected number of minority jurors in a sample of \( n \) from a large area with a minority fraction \( \pi \).
and the CD by

\[ \sum_{i=1}^{112} n_i \pi_i - 1. \]  

(5)

Notice that statistic (4) tests whether number of jurors from the tracts where most African American potential jurors reside are underrepresented on the venires compared with their expected number based on the county-wide population. It differs from the numerator of (2), \( x - n \pi \), in two ways: 1) the racial composition of the potential jurors had to be estimated from related information and 2) it uses information about the residence of the venire members. Indeed, the random components of (4) are the numbers of venire members coming from each tract. The mathematical properties of the statistic (4) are given in an Appendix, where its standard deviation is derived.

The data for the 6-month period from April through October 1993, which included the trial of the defendant and the data for the following 12-month period were examined separately by the expert. As the statistical methodology used was the same, the analysis for the earlier time period will be described in more detail.

Comment: The expert’s statistic (4) detects any discriminatory effect of the County’s jury selection process that is reflected by a shortfall in the number venire members coming from tracts with relatively high fraction of the age-eligible minorities. Within each tract, the expected number of African Americans chosen equals their expected number in a random sample of \( n_i \) age-eligible residents of the \( i \)th tract. Thus, any differential effect of the selection process within a tract, e.g. if the proportion of minority individuals whose summons’ are returned because they recently moved within the same tract is higher than that of whites, is not be reflected in the statistic used. Consequently, the statistical measure (4) is conservative in the sense that it is likely to overestimate the number of African Americans who actually were on the venires or somewhat underestimate their shortfall.\(^{68}\)

In the briefs of both parties and the lower court opinions, the expected numbers of African Americans obtained by defendant’s expert were rounded up to an integer that was taken as the “observed” number, \( x \), and all measures and tests of significance were calculated as usual.\(^{69}\) Since the expected value of a random variable and its observed value usually differ, this is not technically correct. However, to keep our discussion consistent with the previous decisions and court record, until Section 5.6 we will often do the same.

The defendant applied formula (4) to the data for the six terms of the first period. The data and summary calculations from the report of defendant’s statistical expert are reproduced in Table 1. The report observes that underrepresentation is expected in five of the six terms and that African Americans were overrepresented only in term 8. It notes that the CD measure of the expected shortfall in terms 9 and 12 exceeded 30%. For the entire period (Total), the expected shortfall of about 12 African Americans corresponds to an underrepresentation of 18%.\(^{70}\) Moreover, the second largest comparative shortfall (−34.8%) occurred in the defendant’s venire. Furthermore, no African Americans were among the 37 members of the original venire of 60, after individuals were removed.

\(^{68}\) Sometimes the measures were calculated from the estimated expected values, which are fractions, e.g. the expected number of African Americans among the 929 members of the venires in April 1993 to October 1993 was 55.4. Usually, this was rounded to 56.

\(^{70}\) Table 1 and the expert’s summary were obtained from the report, J.A. 168a–184a.
for cause or excused for other reasons. Finally, defendant’s statistician observed that in only one of the months were the estimated numbers of African American jurors at or above their expected number, based on their proportion of the age-eligible population. Formal statistical tests, however, were not applied to the data.

The report also conducted a statistical analysis based on the binomial model underlying the usual statistic (2) for each tract. The observed number, \( n_i \), of jurors from the tract was compared to its expected value, \( n f_i \), where \( n = 929 \) and \( f_i \) is the proportion of jury eligible population residing in the \( i \)th tract. Then, the tracts were classified as being in one of three categories: overrepresented, fairly represented or underrepresented according to the significance level or \( p \) value of the two-tailed test based on (2). Several tables summarizing the results are given; perhaps the one that most clearly illustrates the relationship between the African American proportion of the population of a tract and the underrepresentation of its residents in the jury pool classifies the tracts into three categories, very low fraction (<0.005) of African Americans or middle range (0.005–0.05) or high fraction (0.05–0.906). A reformulated version is reproduced in Table 2.

\[71\] Since the sample size in each tract is relatively small, the expert considered several critical or cut-off values. Here, we report the classifications using a 25% cut-off level rather than 5%. The representation of a tract on the venires is considered fair if the \( p \) value of the test is above 0.25. If the jury-eligible members of each tract were obtained from a random sample of the county, one would expect about 75% of them to be classified as fair. The tracts that were found to be statistically significantly over- or underrepresented, i.e. the \( p \) value of the test was less than 0.25 and if more (fewer) jurors than expected, they were classified as over (under).

---

### Table 1
Comparison of the expected number of African American jurors based on residence data to the expected number assuming a random sample of the county for the period April 1993 to October 1993

<table>
<thead>
<tr>
<th>Term</th>
<th>Tract-based Expectation</th>
<th>County-wide Expectation</th>
<th>Difference</th>
<th>CD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>7.7</td>
<td>10.6</td>
<td>−2.9</td>
<td>−27.4</td>
</tr>
<tr>
<td>8</td>
<td>12.3</td>
<td>11.4</td>
<td>0.9</td>
<td>7.9</td>
</tr>
<tr>
<td>9</td>
<td>6.3</td>
<td>10.9</td>
<td>−4.6</td>
<td>−42.2</td>
</tr>
<tr>
<td>10</td>
<td>10.7</td>
<td>11.2</td>
<td>−0.5</td>
<td>−4.4</td>
</tr>
<tr>
<td>11</td>
<td>10.9</td>
<td>12.0</td>
<td>−1.1</td>
<td>−9.2</td>
</tr>
<tr>
<td>12</td>
<td>7.5</td>
<td>11.5</td>
<td>−4.0</td>
<td>−34.8</td>
</tr>
<tr>
<td>Total</td>
<td>55.4</td>
<td>67.6</td>
<td>−12.2</td>
<td>−18.0</td>
</tr>
</tbody>
</table>

Source: Table 4 of the expert report of the defendant’s statistician (J.A.181a and Pet. Brief at 13).

### Table 2
The relationship between the representation of residents of census tracts on the April 1993 to October 1993 venires and their African American proportion

<table>
<thead>
<tr>
<th>Black proportion</th>
<th>Number of census tracts</th>
<th>Overrepresented</th>
<th>Fairly represented</th>
<th>Underrepresented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>34</td>
<td>12</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>Middle</td>
<td>43</td>
<td>7</td>
<td>27</td>
<td>9</td>
</tr>
<tr>
<td>High</td>
<td>35</td>
<td>1</td>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>112</td>
<td>20</td>
<td>71</td>
<td>20</td>
</tr>
</tbody>
</table>

Source: Table 3 of the expert report (J.A. 178a).
TABLE 3 Comparison of the expected number of African American jurors based on residence data to the expected number assuming a random sample of the county for the 1993–1994 venires

<table>
<thead>
<tr>
<th>Term</th>
<th>Tract-based Expectation</th>
<th>County-wide Expectation</th>
<th>Difference</th>
<th>CD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.2</td>
<td>16.0</td>
<td>4.2</td>
<td>26.3</td>
</tr>
<tr>
<td>2</td>
<td>15.5</td>
<td>16.2</td>
<td>−0.7</td>
<td>−4.3</td>
</tr>
<tr>
<td>3</td>
<td>10.3</td>
<td>17.5</td>
<td>−7.2</td>
<td>−41.1</td>
</tr>
<tr>
<td>4</td>
<td>10.1</td>
<td>12.4</td>
<td>−2.3</td>
<td>−24.9</td>
</tr>
<tr>
<td>5</td>
<td>8.7</td>
<td>15.4</td>
<td>−6.7</td>
<td>−43.5</td>
</tr>
<tr>
<td>6</td>
<td>11.1</td>
<td>13.7</td>
<td>−2.6</td>
<td>−19.0</td>
</tr>
<tr>
<td>7</td>
<td>13.0</td>
<td>14.2</td>
<td>−1.2</td>
<td>−8.5</td>
</tr>
<tr>
<td>8</td>
<td>12.1</td>
<td>13.7</td>
<td>−1.6</td>
<td>−11.7</td>
</tr>
<tr>
<td>9</td>
<td>11.1</td>
<td>13.8</td>
<td>−2.7</td>
<td>−19.6</td>
</tr>
<tr>
<td>10</td>
<td>14.0</td>
<td>14.1</td>
<td>−0.1</td>
<td>−0.7%</td>
</tr>
<tr>
<td>11</td>
<td>13.3</td>
<td>17.1</td>
<td>−3.8</td>
<td>−22.2</td>
</tr>
<tr>
<td>Total</td>
<td>139.4</td>
<td>164.1</td>
<td>−24.7</td>
<td>−15.1</td>
</tr>
</tbody>
</table>

Source: Table 4 of the expert report (J.A. 102a).

TABLE 4 The relationship between the representation of residents of census tracts on the November 1993 to October 1994 venires and their African American proportion

<table>
<thead>
<tr>
<th>Black proportion</th>
<th>Number of census tracts</th>
<th>Overrepresented</th>
<th>Fairly represented</th>
<th>Underrepresented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>34</td>
<td>13</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>Middle</td>
<td>43</td>
<td>5</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>High</td>
<td>35</td>
<td>0</td>
<td>23</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>112</td>
<td>18</td>
<td>69</td>
<td>25</td>
</tr>
</tbody>
</table>

Source: Table 3 of defendant’s expert report (J.A. at 99a).

The expert also analysed the data for the following year. Tables 3 and 4 present the summaries analogous to Tables 1 and 2 for the 11 terms in 1993–1994.

The defendant’s expert stressed that the results in Table 3 imply an underrepresentation of African Americans in 10 of the 11 terms. Indeed, an overrepresentation of this minority group only occurred in the first term. Two terms, the third and fifth, indicate an underrepresentation exceeding 40%. The defendant argued that over both periods in only 2 of 17 terms were African Americans expected to be overrepresented demonstrates that there was a consistent pattern of their systematic underrepresentation on the venires in the circuit court.

The relationship between the under- or overrepresentation of the census tracts and the minority proportion of residents is summarized in Table 4. Although the expert did not apply a formal statistical test to the data, the report notes that there 20 (21) tracts were over (under) represented and the data are consistent with too few (many) jurors being selected from census tracts with a high proportion of minority (white) residents.

72 J.A. at 103a.
73 Respondent’s Answer Opposing Petition for Certiori at 26.
74 Respondent’s Brief at 44.
4.3 Summary of the lower court decisions

The original state trial court rejected defendant’s claim that the Kent County’s jurors were not drawn from a fair cross section of the community. Although the court determined that African Americans were a distinct group and there was a constitutionally significant disparity, it found that the defendant had not shown that a systematic underrepresentation due to the sending of prospective jurors to their local district courts first. Rather, the economic and social factors described by the defendant’s expert demographer as creating the disparity were external to the jury selection process. While sympathetic to the defendant’s claim, the trial court noted that he had not submitted sufficient evidence showing that first sending potential African American circuit court jurors to the local district court in Grand Rapids, where 85% of them lived, created the disparity. Under the system, white residents of the other parts of the county were also sent to the local courts. The trial judge noted that the defendant had not submitted evidence showing that a higher percentage of minorities were sent to the local courts than whites.

In a 2-1 decision, the Michigan Court of Appeals reversed, finding that there was a significant underrepresentation of African Americans on the circuit court venires that resulted from the process of first sending members of the jury pool to the local courts. The decision credited the anecdotal testimony of Mr. Hillary, the public defender, who recalled that prior to 1993; there were perhaps two or three blacks on a panel of 150–175. It also noted the 61st District Court in Grand Rapids had six judges and only residents of that city can be jurors. The other district courts had fewer judges, so the effect of first sending potential jurors to the local courts would diminish the pool of minority members available to serve on the circuit Court. The majority opinion also states that minorities were underrepresented by 34.8% in the first period (April 1993 to October 1993) and although defendant’s statistical expert had not presented an analysis of the juries seated before that time, they were constrained to conclude that African Americans were significantly underrepresented on circuit court venires during the applicable time frame.

In a partial dissent, Judge Hoekstra agreed that the first two requirements of the Duren standard were met. He differed from the majority’s conclusion that the disparity arose from the system that first sent potential jurors to the local courts. He noted that defendant’s statistician had testified that the underrepresentation (1.28%) in the first period was “very consistent” with the underrepresentation (1.11%) in the 1993–1994 period, after that process changed. Thus, he concluded that a “statistically significant change” had not occurred after the system stopped siphoning off the largest concentration of African Americans from the circuit court by first selecting district court jurors. He observed that while systemic exclusion might exist, the defendant had not made the requisite showing based on the record.

The state appealed to the Michigan Supreme Court, which unanimously reversed the state Court of Appeals. The decision noted that courts have applied the three different methods based on the AD or CD measures or the standard deviation statistical test to assess whether minorities are

---

75 From Table 1, one sees that the venire from which defendant’s jurors were chosen had a CD of 34.8%. During the entire 6-month period, the CD was 18%. In Castaneda and Duren, the Court examined jury or venire composition data over a period of time rather than focusing on a particular venire.

76 All parties agreed on the first prong, African Americans are a distinct subgroup. The second is that the number of African Americans on the venires was not fair and reasonable in relation to the Census data.

reasonably represented.\textsuperscript{78} It also mentions the problems, described in Section 3 involved with each of the methods and notes that neither party submitted a standard deviation analysis.\textsuperscript{79} It concludes that while defendant presented some evidence of a disparity between the number of jury-eligible African Americans in the county and their number in the jury pools, the evidence failed to establish a legally significant disparity under either the AD or the CD measure.\textsuperscript{80}

The court gave the defendant the benefit of the doubt on underrepresentation and considered the third prong of the \textit{Duren} analysis, i.e. did the jury selection process systematically result in fewer than expected African Americans appearing on venires? The decision found that defendant’s claim that African Americans were underrepresented on circuit courts because they first were sent to serve on district courts was not supported as no comparative statistics on the composition of the jury pools of the two types of courts was offered. The court did not believe that the County needed to counteract social and economic factors that might diminish minority participation on juries.\textsuperscript{81} Finally, the decision noted that the duration, about 18 months, of underrepresentation was similar to that in \textit{Duren}, the magnitude of the disparities were “far short” of those in that case.\textsuperscript{82}

The case then was reviewed by the federal courts. The main statistical aspect of defendant’s petition was the claim that he was denied equal protection and due process since he was tried before an all-white jury in Kent County and there was systematic underrepresentation in the 1992–1993 juries. After reviewing the statistical findings of the previous decisions, the district court agreed that the disparities in the case were far more modest than those in Duren and that the petitioner had not offered evidence of a disparity between the minority proportions of venires in the local (district) courts and the circuit court.\textsuperscript{83}

On appeal, the Sixth Circuit reversed, finding that the Michigan Supreme Court’s application of the \textit{Duren} criteria was unreasonable. The opinion states that because of the relatively small size of the African American population of the County, none of the three statistical approaches can measure the underrepresentation that occurred on the venire panels.\textsuperscript{84} After noting that the AD of 1.28\% was less than the values accepted by other courts, the decision emphasized that when a distinctive group is small, the CD measure is more appropriate.\textsuperscript{85} It then found that the CDs of 18\% for the venires during the April to October 1993 period and the 34\% for the month of the trial were sufficient to
demonstrate that the representation of African Americans on the venires was unfair and unreasonable. 86 The court stated that the persistence of the disparity for 17 months combined with evidence that it was not random sufficed to show the jury selection process systematically underrepresented African Americans. 87 The decision cites U.S. v. Rogers88 for the proposition that further evidence of a systematic lack of African Americans on the venire pools is provided when chance or a random sample of the eligible population would have a low probability of yielding an underrepresentation of the magnitude seen in the data. While the defendant provided a statistical study indicating minority underrepresentation, unlike Rogers neither party submitted the results of a formal statistical test or a standard deviation analysis.

The Sixth Circuit’s opinion then notes that there were two nonrandom factors in the jury selection process. First, the hardship excuses, child care responsibilities and lack of transportation, tended to exclude a higher proportion of African Americans. Second, the fact that 85% of the African Americans in Kent County lived in Grand Rapids and residents of that city were first designated for panels in the local court diminished the number of African Americans available to serve on the county’s circuit courts. 89 The panel found that the state met its burden for justifying excusing prospective jurors citing hardships. However, it did not demonstrate a significant interest in assigning Grand Rapids residents to the 61st District Court in the city prior to their serving in the circuit courts. The opinion observed that the County had changed this priority assignment process on precisely the grounds highlighted by the petitioner. 90

4.4 The arguments presented to the U.S. Supreme Court

As the defendant’s statistical presentation has been described in detail, this section devotes more space to the state’s argument.

4.4.1 The brief of the petitioner. The state of Michigan, argued that the Sixth Circuit’s decision is improper under either the AEDPA 91 or fair representation clause. The statute limits the power of federal judges to grant habeas corpus relief unless the state court’s decision denying it was contrary to or an unreasonable application of clearly established federal law or based on an unreasonable determination of the facts given the evidence. Thus, the state asserted that the decision of the Michigan Supreme Court was not an unreasonable application of federal law, so the Sixth Circuit should not have reversed it.

The statistical discussion focuses on the 6-month time period, which included the trial of the defendant. It presents a summary table of the statistical expert’s report very similar to Table 1 above. 92 The brief mentions the data for the year following the trial only to refute the claim that the earlier system of assigning African Americans first to the district courts led to their underrepresentation on

---

86 The court, Ibid. at 338, noted that in United States v. Rogers, 73 F. 3d 774, 777 (8th Cir. 1996) the panel thought that a CD of over 30% met the underrepresentation requirement even when the AD was 1.87%.
TABLE 5 Petitioner’s summary of the statistical disparities accepted in fair representation cases in other federal circuits\textsuperscript{95}

<table>
<thead>
<tr>
<th>Circuit</th>
<th>Minimum % in population</th>
<th>Minimum % in jury pool</th>
<th>AD (%)</th>
<th>CD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>4.86</td>
<td>1.89</td>
<td>-2.97</td>
<td>-61</td>
</tr>
<tr>
<td>2d (AA)</td>
<td>7.08</td>
<td>5.0</td>
<td>-2.08</td>
<td>-29</td>
</tr>
<tr>
<td>(Latinos)</td>
<td>4.24</td>
<td>2.10</td>
<td>-2.14</td>
<td>-50</td>
</tr>
<tr>
<td>3d (AA)</td>
<td>3.07</td>
<td>1.84</td>
<td>-1.23</td>
<td>-40.01</td>
</tr>
<tr>
<td>(Latinos)</td>
<td>0.97</td>
<td>0.26</td>
<td>-0.71</td>
<td>-72.98</td>
</tr>
<tr>
<td>7th (AA)</td>
<td>3.0</td>
<td>0.0</td>
<td>-3.0</td>
<td>-100</td>
</tr>
<tr>
<td>8th (AA)</td>
<td>1.87</td>
<td>1.29</td>
<td>-0.579</td>
<td>-30.96</td>
</tr>
<tr>
<td>9\textsuperscript{th} (Hispanic ND)</td>
<td>3.87</td>
<td>1.82</td>
<td>-2.95</td>
<td>-52.9</td>
</tr>
<tr>
<td>Hispanic SD</td>
<td>5.59</td>
<td>2.79</td>
<td>2.8</td>
<td>-50.0</td>
</tr>
<tr>
<td>10\textsuperscript{th} (AA)</td>
<td>7.74</td>
<td>4.78</td>
<td>2.62</td>
<td>-35.41</td>
</tr>
<tr>
<td>(Native Am.)</td>
<td>4.21</td>
<td>2.66</td>
<td>1.55</td>
<td>-36.82</td>
</tr>
<tr>
<td>(Asians)</td>
<td>1.47</td>
<td>0.67</td>
<td>0.80</td>
<td>-54.41</td>
</tr>
<tr>
<td>(Latinos)</td>
<td>3.02</td>
<td>1.36</td>
<td>1.66</td>
<td>-54.97</td>
</tr>
</tbody>
</table>


circuit courts. Implicitly, the brief asserts that the reduction in the 18% CD in the first period to 15% in the second, when circuit courts were given priority, was minimal.\textsuperscript{93}

Later, the brief notes that the majority of federal circuits have relied on the AD, when the distinct group comprised a substantial proportion of the population.\textsuperscript{94} The state also argues that the AD measure should be used even when the minority proportion is small and presents a table of cases from seven circuits with both the AD and the CD measures in which the courts rejected disparities of the magnitude in Table 1 in this situation. The basic information is summarized in Table 5.

Statistical Comments on Table 5: While the results presented in Table 5 demonstrate that other courts have accepted numerical disparities greater than those in the current case, the source pool and jury pool data sets are not uniform. For example, the defendant in Ashley (7th Cir.) considered only the single venire from which his jury was selected. Thus, the number of potential jurors is likely to be much smaller than the 929 in the six terms most relevant here. It is difficult to imagine a court accepting a system that produced a venire of 929 with no members of a subgroup constituting 3% of the jurisdiction’s voting age population.\textsuperscript{96} In some cases, e.g. Weaver (3d Cir.), the race–ethnic composition of the master jury wheel containing several thousand individuals was compared to the Census data while in others, Rioux (2d Cir.) and Orange (10th Cir.) the qualified jury wheel was compared to the population. In the current case and Rogers (8th Cir.), the jury pools or venires, not

\textsuperscript{93} Ibid. at 14.
\textsuperscript{94} Ibid. at 30–33.
\textsuperscript{95} Organized by circuit, the cases are \textit{U.S. v. Royal}, 174 F. 3d (1st Cir. 1999); \textit{U.S. v. Rioux}, 97 F.3d 648 (2d Cir. 1996) and 930 F. Supp. 1558 (D. Conn. 1995); \textit{U.S. v. Weaver}, 267 F. 3d 231 (3d Cir. 2003); \textit{U.S.v. Ashley}, 54 F. 3d 311 (7th Cir. 1995); \textit{U.S. v. Rogers}, 73 F. 2d. 774 (8th Cir. 1996); \textit{U.S. v. Sanchez-Lopez}, 879 F. 2d (9th Cir. 1989) and \textit{U.S. v. Orange}, 447 F. 3d 472 (10th Cir. 2006), affirming 364 F. Supp. 2d 1288 (W.D. Okla. 2005).
\textsuperscript{96} Notice that one would expect 27.9 African Americans among the 929. Since 1 standard deviation equals 5.2, 0 out of 929 corresponds to a difference exceeding 5 standard deviations. Under random sampling from the relevant population, such an event has a probability less than one in a million of occurring.
the master wheels were compared to population data. Like the current case, most cases refined the Census data to include individuals over 18 or eligible to vote. In *Rogers* (8th Cir.) and *Sanchez-Lopez* (9th Cir.), however, the Census data used referred to the entire population, including children. Finally, in *Rogers* (8th Cir.), a probability calculation similar to the standard deviation analysis, indicated that the small number (70) of African Americans out of 5424 potential jurors would occur less than one in a thousand random samples from the population.97

The state’s brief points out that even the Sixth Circuit acknowledged that the Court had not clearly established the preferred method of measuring underrepresentation.98 Thus, the Michigan Supreme Court was reasonable when it considered the AD criteria followed by the majority of the courts. To buttress its case, the state criticizes several statistical aspects of the appellate opinion and discusses some others. The main points that involve statistical reasoning are

1) The Sixth Circuit emphasized the CD of −34% for the month of the trial rather than the −18% figure for the 6-month period. This contrasts with *Duren*, which did not isolate the month of the trial but considered the female fraction of venires for a period of time.99 Thus, the Sixth Circuit was incorrect when it isolated one month’s data to compare to disparities in other jurisdictions that were measured over a period of time.

2) The appellate opinion was in error by relying on *U.S. v. Rogers*100 which was the only case where a federal circuit court found a smaller CD justified a violation of *Duren*. That decision, however, rejected the Sixth Amendment claim because it felt constrained by a previous decision.101

3) One section argues that previous decisions emphasized that the purpose of requiring venires to have a fair cross section of the community is to ensure that the jury will not be ill disposed to a class of defendants or comprised only from a segment of the population that excludes a large distinctive group.102 Then, the brief observes even if one focuses on the particular venire of 158 individuals from which the jurors were chosen, there would be a shortfall of only four African Americans. Using the trial judge’s estimate that there were three minority members on the panel of 60 prospective jurors, there would only have been one or perhaps two more African Americans on the panel.103 Then, the brief observes that any one of the 60 panel members has only a 20% chance of being on the jury, so it was unlikely that an additional one or two African Americans on the panel would have affected the defendant’s petit jury.104

97 73 F. 2d 774 at 78.
99 Ibid. at 37, also citing *Duren*, 439 U.S. at 366 where the Court states that defendant must show a systematic disparity between the composition of the jury pool and the population. Thus, an occasional large disparity could occur in a generally fair system.
100 Supra n. 95.
101 Ibid. at 777, citing *United States v. Clifford*, 640 F.2d at 155.
103 Ibid. at 41. The data are given in Table 1 on the line for term 12. As the expected number of African Americans on a panel of 60 is 0.0728 × 60 = 4.37 and there were three, the shortfall is 1.37, i.e. one or two.
104 The brief cites *U.S. v. Jenkins*, 496 F.2d 57, 65 (2d Cir. 1974) for a similar argument based on the AI of a disparity. Problems with this approach are discussed in Section 3.3. Here, we add that the calculation that each member of a panel of 60 has a 1 in 5 probability of being on the jury assumes that the jury is a random sample of 12 from the panel. Whether the prospective jurors remaining after several have been removed for cause and both parties have exercised their peremptory challenges can be considered as a random sample from the original panel is problematic.
4) The brief emphasizes that other courts have criticized the CD measure because it exaggerates or distorts the "significance of the disparities for small groups." It asserts that if the Court relies on the CD measure in "fair representation" cases, then it will force the counties and states to effectively set up a quota system to ensure venires mirror the composition of the population. It also argues that while critics of the AD measure are properly concerned that its use would allow the complete exclusion of a distinct group comprising less than 10% of the population, these considerations are addressed by the "equal protection" clause.

5) While the brief argues for the adoption of a threshold requirement of a 10% AD, it points out that even if the Court decides that the CD is appropriate in some circumstances, the statistics from the cases summarized in Table 5 indicate that the disparities, whether measured in terms of the AD or CD, are smaller than those other courts have accepted as establishing a prima facie case of unfair representation.

The brief agrees with the Sixth Circuit’s finding that the system of excusals based on hardship was legitimate but argues that the Michigan Supreme Court was correct when it concluded that the defendant did not provide sufficient proof that African Americans were “siphoned” off to serve on the district courts caused their underrepresentation in the April 1993 to October 1993 period. It criticizes the Sixth Circuit for relying on anecdotal testimony even though there was only a small change in the African American percentage of the venires after the County discontinued the process. Furthermore, even if the CD of 18% of the earlier period was reduced to the 15% of the later period, this would have led to only two more African Americans out of the total of 929. The state concludes that such a small effect of the jury assignment process operating at the time of defendant’s trial is not constitutionally significant.

4.4.2 The brief of the respondent. The respondent’s brief reviews the main findings of its statistical expert and the demographic reasons that lead to minority underrepresentation. It notes that each measure has shortcomings and argues that the appropriate measure should be decided on a case by case basis.

The brief rebuts the state’s argument that courts should adopt a minimum AD of 10% as a threshold requirement by citing the equal protection case Norris v. Alabama that found discrimination.

---

106 Citing Foster v. Sparks 506 F. 2d 805 (5th Cir. 1975).
107 Pet Brief at 44.
108 Ibid at 45.
109 Ibid. at 58–61.
110 Ibid. at 61 at n. 86 and accompanying text. Under fair representation, the expected number of African American venire members is $0.0728 \times 929 = 67.63$, so another 3% increases their representation by $0.03 \times 67.63 = 2.02$.
111 Respondent’s brief, at 10, recalls their expert demographer’s testimony (J.A. 78a–85a) that in Kent County 64% of African American households have a single head in contrast to only 19% of whites; the incomes of 31.5% of African American households are below the poverty line in contrast to 6.7% of whites. Similar statistics relating to recently moved or not having a car are also cited.
112 Respondent’s brief at 26. On page 28–29, United States v. Chanthadara, 230 F. 3d 1237, 1256–1257 (10th Cir. 2000) (noting that ADs are of limited value when considering small populations) and several state courts, e.g. State v. Williams, 525 N.W. 2d 538, 543 (Min. 1995) (courts should rely on more than one statistical tool) and Lovell v. State, 702 A.2d 261,281 (Md. 1997) (courts may supplement the AD measure with CD figures when the distinctive group forms less than 10% of the total population).
113 294 U.S 587 (1935).
even though African Americans formed 7.2% of the population. In that case, no minority had served on a jury for years. The Court found discrimination even though the AD was less than 10% but the CD = 100%. It also provides data from the Census Bureau showing that many different minority populations could be excluded from juries if the Court required a minimum AD of 10% in an underrepresentation case.\textsuperscript{114} The defendant argues that the underrepresentation was systematic because of the consistent pattern seen in Tables 1 and 3, i.e. in 15 of the 17 months the estimated number of African American members of the venire were below their expected value.\textsuperscript{115} While the state claims that the underrepresentation was due to social and economic factors, the state did not offer evidence as to the effect these factors would have on the eligible jury pool. Thus, the respondent questions whether the state met its burden under \textit{Duren} of showing that the underrepresentation arose because of a significant state interest.\textsuperscript{116}

The brief makes an interesting point concerning the lack of proof of the effect of the excusal process as the county did not keep any records. This would allow a state or county that is arbitrary and inconsistent in its procedures to be insulated from court review of its jury selection process because it failed to preserve the information it had or easily could have obtained.\textsuperscript{117}

Finally, the respondent emphasizes that defendant’s jury was chosen when members of the jury pool were first assigned to the local courts and African Americans were underrepresented by 34%, measured by the CD, on this venire.\textsuperscript{118} Earlier the brief recalls that no African Americans were among the 37 who were actually considered to be on defendant’s jury and that his lawyer attempted to obtain a jury with at least one minority member.\textsuperscript{119}

4.5 \textit{The decision of the U.S. Supreme Court}

First, the decision summarizes the issues and evidence discussed in the previous decisions and states the \textit{Duren} criteria a defendant must satisfy to establish a \textit{prima facie} case. The opinion states that the data submitted in \textit{Duren} discussed in Section 3.2, clearly met the two statistical standards.

After noting that the Michigan Supreme Court correctly observed that neither \textit{Duren} nor any other decision of the Court specifies the method or test courts should use to assess data on the representation of distinctive groups in jury pools, the opinion observes that each test is imperfect.\textsuperscript{120} While the Michigan Court indicated that the results of all the tests could be considered, the Sixth Circuit declared that the CD is the appropriate measure when the distinctive group is small. The Court felt it did not need to choose between the various measures of underrepresentation in order

\textsuperscript{114} Respondent’s brief at 34–35 (e.g. Asians in Cook County, Illinois constitute 5.83% of the population and Hispanics in Fulton County, Georgia form 8.35% of the residents).
\textsuperscript{115} Ibid. at 38.
\textsuperscript{116} Ibid. at 40. The brief quotes \textit{Duren}, at 368. Before the state is required to justify an aspect of its jury selection system that leads to underrepresentation of a distinct group, the quoted section of \textit{Duren} states that the defendant had to establish a \textit{prima facie} case. The proper interpretation of the shifting evidentiary burdens under \textit{Duren} is clarified in the U.S. Supreme Court’s decision, \textit{supra} n. 2 at 15.
\textsuperscript{117} Ibid. at 50.
\textsuperscript{118} Ibid. at 51–52. See term 12 in Table 1 for the data.
\textsuperscript{119} Ibid. at 3–4.
\textsuperscript{120} 130 S. Ct. 1382 (2010) at 1393 (noting that both the AD and the CD measures may be misleading when the group in question forms a small percentage of the population and that no court had relied solely on a standard deviation analysis in a Sixth Amendment case).
to decide the case; like the Michigan Court it turned to the evidence concerning the third \emph{Duren} criterion.

Since the defendant asserted that assigning venire members to local district courts first led to their systematic exclusion from the county’s circuit courts, the decision turns to the evidence. While the record established that some officials and others believed the assignment order created racial disparities, this belief was not substantiated by Smith’s evidence. The opinion notes that evidence showing that the African American percentage of venires in the local courts was significantly higher than their percentage of circuit court venires, which was not submitted, would have supported the defendant’s claim. Two additional statistical points were made

1) The defendant had not demonstrated that Grand Rapids, where most of Kent County’s highest minority population resided, required more jurors per capita in the local courts than other districts.

2) Unlike in \emph{Duren}, the minority percentage of venires in the Circuit Court was not compared with that of Federal Court venires for the same region.

The Court stated that Smith’s best evidence of systematic exclusion was offered by his statistics expert, who demonstrated a decline in the CD from 18 to 15.1%, after the County changed the assignment system. This modest increase in African American representation on Circuit Court venires did not justify a conclusion that the district-court-first assignment process had a significant impact on minority representation.

In addition to the assignment process, the defendant claimed that a number of factors such as people requesting hardship excuses or failing to show up for jury service and the County’s not following up on “no shows” contributed to minority underrepresentation. The Court noted that no “clearly established” precedent allows a defendant to make out a \emph{prima facie} case by simply listing factors that might contribute to a group’s underrepresentation. Indeed, the precedential cases, \emph{Taylor} and \emph{Duren} envision that factors similar to those cited by Smith might well survive a fair cross-section challenge as they relate to an important state interest. Thus, the Michigan Supreme Court’s decision rejecting the fair cross section was consistent with \emph{Duren} and should not have been reversed by the Sixth Circuit.

5. \textbf{Alternative statistical analyses of the data and related issues}

Although the estimated numbers of African Americans on venires in the two time periods were compared to their expected numbers based on Census data in the expert report, the only standard deviation analysis in the record is a calculation using the data from the first period by a judge on the Michigan Supreme Court. Before conducting a statistical test, it is important to consider whether the sample size is sufficient to reach a sound conclusion. The concept of the power of a statistical

\begin{itemize}
  \item \textsuperscript{121} Ibid. at 12. In fn 4, the opinion also states that it need not discuss the appropriateness of the AD should exceed 10% as the state argued.
  \item \textsuperscript{122} Ibid. at 13.
  \item \textsuperscript{123} 439 U.S. at 367 n. 25 (the data are reported in Section 3.2, \emph{supra}).
  \item \textsuperscript{124} 130 S. Ct. 1382, 1394 (2010).
  \item \textsuperscript{125} Ibid. at 15–16.
  \item \textsuperscript{126} This calculation was made by Judge Cavanagh of the Michigan Supreme Court in his concurring opinion, 615 N.W. 2d, 1, 10 (2000). Unfortunately, he made a computational error. The correct analysis is given \emph{infra} n. 137.
\end{itemize}
test is helpful here, so Section 5.1 presents a power analysis for the usual statistical procedures based on the binomial model applied to the data in Berghuis. It will be seen that restricting the analysis to the 929 panel members in the first period does not enable one to detect some legally meaningful differences. The sample size of 2252 for the second period is adequate and it will be seen that the data for both periods are sufficiently similar that they can be pooled into a single sample. Then, one has a reasonably high probability of detecting an SR of 0.80. The next subsection carries out the usual tests on the data under the assumption, implicitly made in the calculations available to the Court, that the estimated number \( n \) of African Americans can be regarded as resulting from a random sample of size \( n \) from the age-eligible population. Section 5.3 applies appropriate statistical tests to the data in Tables 1–4. The results indicate that the defendant’s data provide more convincing evidence of minority underrepresentation than was demonstrated to the courts. The statistical significance of the consistency of the pattern of underrepresentation is shown in Section 5.4. The potential for the legitimate excusal of potential jurors for child care needs to explain a large part of the disparity is explored in Section 5.5.

The last subsection points out that the usual statistic (2) for the standard deviation analysis is not appropriate as the statistic (4) used by the defendant and described in Petitioner’s brief\(^ {127} \) is the difference of the number of minority jurors, estimated from their residence, and the number expected in a random sample from the county. A “standard deviation analysis” with the correct sampling error of the statistic (4) is used to reanalyse the data. The results indicate that the differences between the expected number of African Americans based on their percentage of the age-eligible population in each tract and the expected number based on the corresponding percentage in the entire County for all three data sets are statistically significant, i.e. have \( p \) values or probabilities of occurring under random sampling less than 0.05. The relevant mathematical derivations are given in the Appendix.

5.1 Which samples are large enough for the test to have adequate statistical power to detect a legally meaningful degree of underrepresentation?

The choice of the 2-standard deviation threshold for statistical significance arose because scientists did not wish to reject the theory generating the hypothesis being tested unless the data clearly contradicted it. To accomplish this, the error rate or probability of falsely rejecting a hypothesis when it is true is kept small, typically 5% or less. When courts accepted statistical decision theory and testing in discrimination cases, many decisions just looked at whether or not the appropriate statistical test applied to the data reaches statistical significance, usually at the 0.05 or 5% level. This oversimplification has been noted by well-respected judges. The role of statistical evidence in the legal system is somewhat different than in science, where one can wait for further studies before a decision is reached. Courts use the results of statistical analyses to shift the burden of producing evidence explaining the data. Furthermore, the implications of the statistical conclusions often can be understood in the context of the nonstatistical evidence. Focusing only on whether a statistical test reaches statistical significance at a fixed level, e.g. 5% fails to consider whether one has a reasonable chance detecting a legally meaningful underrepresentation or difference between the outcomes of two groups.

Since one also desires to reject hypotheses that are not true, some statisticians also consider the probability of the other error (type II), failing to reject a hypothesis which is not true. Consider an

TABLE 6 Approximate power of the binomial test statistic to detect underrepresentation of minorities when they form 7.28% of the age-eligible population: the three data sets in *Berghuis v. Smith*

<table>
<thead>
<tr>
<th>SR ((\psi))</th>
<th>Alternative ((\pi^*))</th>
<th>Sample of 929</th>
<th>Sample of 2252</th>
<th>Sample of 3181</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80</td>
<td>0.0591</td>
<td>0.376</td>
<td>0.719</td>
<td>0.877</td>
</tr>
<tr>
<td>0.70</td>
<td>0.0521</td>
<td>0.732</td>
<td>0.980</td>
<td>0.998</td>
</tr>
<tr>
<td>0.60</td>
<td>0.0450</td>
<td>0.951</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>0.50</td>
<td>0.0378</td>
<td>0.998</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

alternative that is sufficiently different from the hypothesis that is considered important to detect. In the discrimination context, it seems clear that courts would seriously question a jury system where the SR was 0.70 or less; implying that members of a jury-eligible protected group have no more than 70% the probability of whites of serving on juries. Indeed, in the disparate impact context, a job requirement with an SR of 0.80 needs to be shown to be related to one's ability to perform the job. The precise value of the SR that corresponds to a legally meaningful disparity that is important to detect is a legal issue and should be determined by the courts rather than statisticians. It may depend on the context of the case, e.g. whether the minority proportion of the jury eligible population was simply calculated from Census data or whether the Census data were refined to exclude individuals who would be legitimately excused from serving. Other factors courts might consider are: Did the County simply use voter registration list or did it use driver’s license lists, which tend to be more inclusive (Williams 1990, at 632) or did it use both? Did the County make further attempts to reach individuals in the original sample who were sent a summons with a questionnaire but failed to respond or the post office returned the letter as undeliverable?

The complement \(^{130}\) of the probability of a type II error is the power of a statistical test or the probability that the null hypothesis (proportion of the subgroup on the venire is consistent with the proportion in the jury-eligible population) is rejected when it is false; i.e. the alternative hypothesis that minorities are underrepresented is true. This probability will depend on the precise value of the measure of underrepresentation deemed important to detect and the sample size. Table 6 reports the power for alternative values of the \(\text{SR} = 0.5, 0.6, 0.7\) and 0.80 for the sample sizes of the data from *Berghuis*. Each of these alternatives corresponds to a system where the potential jurors are being sampled from a population where the minority fraction is \(\pi^*\), which is less than \(\pi\) (the value calculated from the Census data, i.e. in *Berghuis* \(\pi = 0.0728\)). The alternative models a system where the jury administrators are limiting the minority group to a smaller fraction of the venire than their proportion of the eligible population. A 2-standard deviation criterion for significance will be adopted. It implies that the number of African Americans on the panels in the first period will be statistically significantly less than the expected number (67.6 or 68) under random sampling.

\(^{128}\) It seems clear that one would have a more stringent standard in the second situation as the most likely explanation for an observed underrepresentation has been accounted for.

\(^{129}\) Using multiple source lists yields a more comprehensive master list; however, one needs to merge files carefully in order to eliminate duplication. See Fukurai et al. (1993) at 44–56 and Munsterman and Hannaford-Agor (2003) at 17–21, for a thorough discussion of these issues and relevant references.

\(^{130}\) The complement of an event is the “opposite” event, e.g. if drawing a red card from a deck of 52 is the event of interest, its complement is drawing a black card. The probability of the complement of an event occurring is one minus the probability of the original event.
from the age-eligible population if 52 or fewer were on the panel. Under the alternative that the jury selection process was consistent with an SR less than 1.0, the data would look like it came from a population where the fraction of African Americans was less than 0.0728. Indeed, for any value, $\pi$, of the minority fraction of the Census data and underrepresentation determined by the value, $\psi$, for SR, the data will follow a binomial distribution with a smaller fraction, $\pi^*$, where

$$\pi^* = \frac{\pi \psi}{1 - \pi + \pi \psi}.$$  

131 As one expects, the smaller the SR ($\psi$), the smaller the minority fraction ($\pi^*$) would be in a system with that degree of underrepresentation. Consequently, it will be easier to detect underrepresentation in systems with SRs substantially below 1.0. The results in Table 6 imply that when the SR is only 0.5, i.e. that the selection process is sampling from a binomial distribution where the African Americans form only 3.78% of the jurisdiction’s population in the age range eligible for jury service, instead of 7.28%, the statistical test on a sample of 929 panelists is almost sure (99.8%) to declare a statistically significant shortfall. If the SR is 0.6, the probability that there will be 52 or fewer African Americans in a sample of 929, is 0.951. This means that the sample size has adequate power to detect this degree of underrepresentation. On the other hand, if one desires to detect a system with an SR of 0.80, the power of the test is only 0.376, implying one has less than a 40% chance of detecting this magnitude of underrepresentation. Thus, if the Court deems that an SR of 80%, analogous to the 80% or “four-fifths” guideline in disparate impact cases, the number (929) of selections in the period from April to October 1993 clearly is inadequate. Indeed, one-fourth of the time a sample of 929 would fail to detect an SR of 0.70. If one restricts the analysis to the data for the first period and insists on testing at the 0.05 level, this implies that if the state’s system is fair, only 5% of the time will it be asked incorrectly to explain it. In contrast, if the true SR = 0.70, the defendant would fail to satisfy the second prong of Duren 25% of the time. This calculation could justify the defendant’s expert not performing a “standard deviation analysis.” 132 If possible, one should extend the time period in order to obtain a sample of sufficient size to have a reasonable chance of detecting this degree of underrepresentation.

In most applications, however, the type II error is allowed to be higher than the type I error because one does not want to erroneously reject a true hypothesis. Thus, many studies are designed to keep the type I error at 0.05 but type II error is set at 0.10 or even 0.20. The sample of 2252 individuals called for possible jury duty in the 11-month period following the trial is large enough to detect an SR of 0.70 but not 0.80. While neither party pooled the data for both periods, 133 the last column of Table 3 presents the power of the test if one had a sample of 3181. Now the sample size

131 The formula is equation (2.5) in Gastwirth and Greenhouse (1987), where a derivation is provided. To illustrate its use, suppose $\psi = 0.50$, i.e. an age-eligible African American has only one-half the probability of being on the relevant venires than a white. When $\pi = 0.0728$, $\pi^* = 0.03778$.

132 Indeed, the federal district court opinion, 2005 U.S. Dist. Lexis 40910 at 17 noted that the statistician did not test the data for the defendant’s venire for statistical significance. Since a sample of 929 is inadequate to detect an SR of 0.70 or 0.80 in this situation, one surely does not have adequate power to conclude that a nonsignificant finding is meaningful from an analysis of a venire of 158. Even a one-tailed test at the 0.05 level only has power or probability of 0.61 to detect an underrepresentation with an SR of 0.50. The power of a one-tailed test of detecting an SR of 0.70 in a sample of 158 is 27.9%, while a two-tailed test as in Castaneda would only have a power of 16.4%. These values imply that when the alternative hypothesis of unfair representation (SR < 0.70) is true, a statistical test on a sample of 158 would most likely fail to detect the minority shortfall. Thus, the expert was correct in not carrying out a test with low power.

133 Since the proportions of African Americans in the two periods were so close, 6.03% and 6.17%, respectively, it is surprising that the defendant did not combine both periods into one sample. The standard test for a difference between two proportions indicates that they were very similar as the $p$ value of a two-tailed test exceeds 0.80. This means that the two samples are homogenous, i.e. come from the same underlying population, and can be combined.
is sufficiently large to be nearly certain to detect an underrepresentation with a SR of 0.70 or less and will correctly identify as statistically significant a system with an SR of 0.80 almost 90% of the time.\textsuperscript{134} As noted in a Comment 1 in Section 3.1 when very large samples are available, statistical testing should be used in conjunction with a value of the measure, which corresponds to a legally meaningful or practical impact, e.g. $\text{SR} < 0.80$ or 0.85, of the disparity. This will avoid classifying a jury selection system that is reasonably representative as unfair.

5.2 Application of the usual statistical test to the data to assess underrepresentation of African Americans

The considerations discussed in the previous section indicate that it would be statistically preferable to combine the data for both time periods into a single sample because the two data sets were examined separately, those analyses will be described in this subsection. It should be pointed out that when several tests are carried out the probability of finding a significant result in at least one of the tests increases.\textsuperscript{135} There are techniques for adjusting for multiple testing but they will not be explored here.\textsuperscript{136}

The defendant’s statistician estimated there would be 55.4 African Americans on the panel of 929 in the April to October 1993 period. These numbers yield the values of the AD (1.28%) and CD (18%), which the Petitioner noted are less than those which many previous courts found inadequate to establish a \textit{prima facie} case. The SR measure is 81.7%, just exceeding the “80%” or “four-fifths” rule used to assess the potential disparate impact of an employment practice. This suggests that the issue of underrepresentation for the period is “statistically close.” In formula (2), the expected number of African Americans is 67.63, so the numerator equals −12.16. The denominator or standard deviation of the number of African Americans who would be in a random sample of 929 from the adult population of the County is 7.919.\textsuperscript{137} The ratio of the shortfall ($-12.16$) to the standard deviation (7.919) is $-1.54$, which does not meet the 2–3 standard deviation criterion used by the Court in \textit{Castaneda}.

For the following 11 months, the defendant’s expert estimated that there were 139.4 African Americans, while 163.9 would be expected. Assume that there were 139 African Americans among the venire of 2252. This will hardly change the previously reported values of AD ($-24.9$) or CD ($-15.1$). The SR equals 83.78%, again somewhat in excess of the 80% guideline used in disparate impact cases. The standard deviation analysis yields the value $-2.02$ for the statistic in (2). Thus, there was a statistically significant (at the 0.05 level) underrepresentation of African Americans during the second period.

\textsuperscript{134} This means that the type II error or probability a random sample of 3181 from a system with an $\text{SR} = 0.80$ will not indicate underrepresentation is slightly greater than 0.10 (0.123 to be precise).

\textsuperscript{135} \textit{Kaye and Freedman} (2000, at 127–128).

\textsuperscript{136} Since the preferred analysis would combine the data for both periods, only two tests are needed. The one that checks the homogeneity of the two samples, \emph{supra} n. 134 and the primary test of whether the minority proportion is consistent with the Census data. Routine calculation of the correlation between the test statistic for homogeneity and the main test statistic shows they are uncorrelated. Thus, the effect on the significance level of the main test is minimal and will be ignored here.

\textsuperscript{137} The calculations presented here differ from those in the record, as the judge made a minor arithmetic error. Apparently, instead of 0.0728, 0.728 was used in the denominator of (1). Thus, he obtained a standard deviation of 26 instead of the correct value 7.92.
As discussed previously, the minority proportions in both periods were quite similar (6.03% in the first and 6.17% in the second) so it is reasonable to pool the data for the entire 17 months. Thus, the estimated 195 African Americans out of 3181 should be compared with their expected number, 231.58. The \( AI = 36.58 \), the \( CD = -15.55\% \) and the \( SR = 83.17\% \). Because there were a much larger number of potential jurors in the second period than in the first, the values of the CD and SR for both periods combined are closer to those of the second period. The value of statistic (2), measuring the shortfall in standard deviation units is \(-2.496\) or \(-2.5\) standard deviations. The probability of such a shortfall, under the null hypothesis that African Americans are properly represented in the pool, is less than 0.01, which supports the defendant’s claim. To summarize, when the sample size is large enough to reliably detect an underrepresentation with an SR of 0.80 or less, African Americans were statistically significantly underrepresented on the circuit courts, assuming they formed 7.28\% of the eligible population. Thus, an important issue is how well does the age-eligible population in the Census data correspond to the pool of jury-eligible members of the population. An approach to this topic is given in Section 5.5.

5.3 **Analysis of Tables 2 and 4 indicates a strong relationship between a census tract’s being underrepresented and the minority percentage of its residents**

The data in Tables 2 and 4 indicate that in both time periods, proportionately more potential jurors came from the census tracts with a small proportion of African Americans, while the tracts with a high proportion were underrepresented. To formally test the hypothesis of no relationship between the underrepresentation of a tract and its minority proportion of residents, one uses the linear by linear test of association (Agresti 1996 p. 34–36), which is based on a type of correlation that measures whether the two characteristics are related. Applying it to the data in Table 2 for the April to October 1993 period yields a highly significant negative association \( (p \text{ value} = 0.0009) \). This is not surprising as 35.3\% of the tracts with a very low percentage of African Americans are overrepresented, while only 2.86\% of tracts with a high percentage of them are overrepresented.

For the subsequent period, the linear by linear statistical test of association applied to the data in Table 4 again yields a highly statistically significant result \( (p \text{ value} = 0.00001) \). This is consistent with the data as none of the 35 districts with a noticeable proportion (5–91\%) of African Americans are overrepresented, while 12 are underrepresented. In contrast, 13 of the 34 districts with a very low (less than one-half of 1\%) proportion of minority residents are overrepresented, while only 5 are underrepresented on the venires. Both \( p \) values correspond to probabilities that are smaller than the value, 0.0027, which demarcates statistical significance at the three standard deviation criteria, much less the 0.05 value required by the 2 standard deviation criteria.

5.4 **Analysis of Tables 1 and 3 may provide further support that the underrepresentation was fairly consistent during the combined period of 17 months**

While the expert noted that the data in Tables 1 and 3 showed that both periods, there were more terms where the number of African American venire members were expected to be underrepresented, i.e. in 5 of the 6 terms in the first period and 10 of 11 in the second, a formal statistical analysis was not presented. Rather, the expert observed that minorities were underrepresented in 15 of the 17 periods. Even if one questioned whether term 10 in Table 1, with an AD of half a juror and a CD of 4.4\% should really be classified as underrepresented, 14 of 17 also seems quite unbalanced.
Assuming that in a fair system, the probability of underrepresentation or overrepresentation of a minority group should be one-half, a sign test indicates a statistically significant imbalance. The probability of observing a deviation from the expected 8.5 underrepresented terms as great as the 14 of 17 is 0.013.\footnote{Considering term 10 as underrepresented implies that African Americans were underrepresented in 15 of the 17 months. Then, the sign test would have \( p \) value 0.00235, or such an imbalance would occur in less than one in 400 random samples from the population.} A statistical method that gives weight to the magnitude of the underrepresentation as well as the imbalance in the proportion of terms with underrepresentation is the Wilcoxon signed rank test. Because the number of venire members in a term was generally larger in the second period, it will be applied to the CDs. For all 17 terms in both periods, the \( p \) value of the test is 0.0038, which is less than the 0.01 level of significance.

5.5 Could the disparity be explained by the legitimate excusal of prospective jurors for hardship?

Both the analyses submitted by the defendant and the previous ones, assumed that African Americans formed 7.28\% of age-eligible potential jurors. The actual pool of venire eligible members of the community might differ from the age-eligible population after individuals are screened for legitimate qualifications and excused for legally appropriate reasons. Ideally, the minority percentage of members of the community eligible to serve on a jury should be compared to their representation on the venires. Alternatively, when eligibility is determined after individuals have answered a suitable questionnaire, the age-eligible population should be compared to the respondents before their qualifications and requests for excusal are evaluated. The appellate opinion found that the state justified the practice of excusing citizens from jury service if they had child care responsibility or were unable to obtain transportation or would suffer financial hardship.\footnote{Ibid. at 340–341.} Previously, that decision observed that in Kent County 64\% of African American households with children were headed by single parents, while only 19\% of white families were headed by single parents.\footnote{See the concurring opinion of Judge Cavanagh, 615 N.W. 1, 13–14.} Since the data analysed in the case refers to the racial composition of the actual venires, it is appropriate to adjust the 7.28\% figure to account for the residents of the county who would be legitimately excused from the potential jury pool. Unfortunately, the authorities did not collect data on the reasons people were excused, much less their race–gender and ethnic background.\footnote{It is possible that some individuals were excused for a different reason than the appellate court approved or there were other acceptable reasons, e.g. medical care. More information is needed on the excuses actually accepted and those the court believes are of valid state interest as well as the demographic composition of the individuals excused to make an accurate estimate of the venire-eligible pool.} Had they done so, one could estimate the fraction of the pool that would have been given a justifiable excuse.

Before suggesting an approximate adjustment of the 7.28\% minority share of the age-eligible residents of the county, it is useful to determine the smallest value that would be consistent with the defendant’s estimated number of actual African American venire members. This is accomplished by calculating the minimum value of \( \pi \) for which the number of African Americans on the venires would be consistent with a random sample from the age-eligible population. Using the 2-standard deviation or 0.05 level criteria, this value is the upper end of a 95\% confidence interval based on the
minority proportion of the venires being examined. The general formula is

\[ p + (1.96)\sqrt{\frac{p(1-p)}{n}}. \]

As noted previously, there are three potentially relevant data sets: the 1992–1993 venires, which had 56 African Americans out of 929; the 1993–1994 venires that included 139 African Americans out of 2252 and combining both periods. For the first data set, \( p = \frac{56}{929} = 0.0603 \) and formula (1) equals 0.0756. Thus, the 56 African Americans on the 1993–1994 venires would not be considered statistically significantly below their expected number provided that their share of the pool was not larger than 7.56%. Notice that the 7.28% figure is less than 7.56%, which is consistent with the previous finding that the shortfall in that time period was not statistically significant. For the second period, \( p = 0.0617 \) and formula (1) equals 0.0716 or 7.16%. This figure is less than 7.28%, consistent with the previous calculation showing a statistically significant underrepresentation in that period. Finally, consider the 195 African Americans out of 3181 in both periods. Now \( p = 0.0613 \) and formula (1) equals 0.0696, implying that if African Americans form more than 6.961% of the age-eligible population after the “legitimately excusable” individuals are removed, the data for the entire period, October 1992 to October 1994 would indicate a statistically significant underrepresentation.

The defendant’s expert demographer testified that a much higher proportion of African American households are headed by a single parent than white households, which might cause proportionately more of them to request to be excused as that single parents with a child under 5 years of age would have child care problems. This is most likely to be the case with poor families, who probably could not afford substitute child care. From the Census data on children in families below the poverty level, one finds that there were 1369 African American households headed by a single parent in Kent County with a child under 5 years of age out of a total of 3320 such households. Assuming that these heads of households would request to be excused for child care reasons, it is reasonable to subtract those households of all races from the age-eligible population in the report of the defendant’s expert. Doing this now yields a minority fraction, \( \pi = 0.0693. \) Notice that if this approximate adjustment is reasonable, then the observed African American percentages (about 6%) on the first two data sets would not reach statistically significance using the 2-standard deviation criteria of Castaneda but would be on the “borderline” for the pooled data. Because this argument

---

143 The formula for both ends of the confidence interval and illustrative examples are given in Gastwirth (1988, p. 119–126) and Finkelstein and Levin (2000, p. 162–126). The method was used in the sex discrimination case Capaci v. Besthoff, 711 F. 2d 647 (5th Cir. 1983) to demonstrate that testing the hiring data against the female proportion of any reasonable labour pool for managerial trainees would yield a statistically significant shortfall in their hires. In Section 5.6, it will be shown that the standard deviation of the binomial distribution used in (1) is an overestimate of the standard deviation of the estimator of \( p \) used by defendant’s expert.

144 Since the percentages of minorities on the venires in the two periods are quite close, one might wonder why one is statistically significantly below their expected value 7.28% and the other is not. The answer is that the sample sizes differ substantially and estimates from a larger sample are more accurate. Indeed, the sampling error is inversely proportional to the square root of the sample size.

145 The data are reported in Table P124B, Poverty Status in 1989 by Race of Householder and Family Type from the 1990 Census and are available from the Web site of the U.S. Census Bureau. The number of African American single parents with at least one child under 5 years of age was 1369 and the corresponding figure for whites was 1864 and there were 87 such families of other race–ethnic groups.

146 From Census data, the expert report states that there were 23476 African Americans out of a total of 322383 in the 18–69 age range living in Kent County. See J. A. at 168a. Subtracting the single parents with a child under 5 years of age, there were 22107 eligible African Americans out of a total of 319063. Thus, \( \pi = 0.0693. \)
assumes that all single heads of households in poverty would request to be excused and does not consider other hardship excuses that might have a greater impact on whites, e.g. medical reasons, it does not prove that the disparity would be reduced to statistical insignificance by the legitimate excusal of individuals claiming hardship. It does show that the state’s explanation is plausible, i.e. the underrepresentation of minorities on venires could well be a consequence of a higher proportion of them being excused for legitimate reasons. Had the defendant met the three Duren criteria, the state would have needed to submit hard data on the demographic composition of those excused for various reasons and showed how these legitimate exemptions caused the shortfall. The argument given here, however, is stronger than a mere suggestion or assertion as it shows that a plausible refinement of the Census data accounting for the major legitimate excuse might show that the minority representation on the venires was not statistically significantly different than their proportion of the jury-eligible population.

Comments: 1) It would be interesting to assess the impact of the above adjustment of the Census data to account for the legitimate excusal of single parents with young children on the statistical analysis of the defendant’s expert. The required data are not readily available at the individual tract level and it is not reasonable to assume that the effect of the adjustment would be the same in all 112 Census tracts.

2) As both parties discuss U.S. v. Rogers, it is interesting to show how formula (6) can be used to assess whether the statistical comparison in that case where the fraction \( \frac{70}{5424} = 0.0129 \) of African Americans was compared their fraction \( 0.0187 \) in the overall population would be rendered nonsignificant had the population data been restricted to individuals 18 years and older. Formula (6) yields 0.0159. Thus, as long as African Americans formed at least 1.6% of the age-eligible population in the jurisdiction in Iowa, there would be a statistically significant shortfall of minorities in the jury pool. The statewide data suggest that restricting the age range would nearly reduce the disparity to nonsignificance. African Americans formed 1.732% of the total population and 1.480% of those 18 and over. Applying the ratio, 0.8544 of the two statewide percentages to the 1.87% of the jurisdiction estimates that African Americans compose 1.598% of the 18 and over population. The Court should define the community or population that the venire should represent so that the Census Bureau could tabulate the relevant data for all jurisdictions in the nation.

5.6 Analysis of the venire composition data using the correct standard deviation of the statistic comparing the two expected numbers of African Americans

The usual method of analysis assumes that the venire of size \( n \) is a random sample in which every age-eligible resident of the County has the same probability of being included. One way of taking such a sample would to first randomly select a tract, where the probability of choosing a tract is its fraction of the age-eligible population of the County. Then, one randomly selects an age-eligible resident of the tract. Because the County did not keep racial information about the actual members of the venire, the defendant’s expert estimated the race of a juror from the racial proportions in the tract.

---

147 It will be seen in the next section that when the correct standard deviation of statistic (4) is used, the state’s explanation would need to justify reducing the African American fraction of the jury eligible population to a lower figure than derived here.

148 439 U.S. at 368–369.

149 The 1990 population data (available at censtats.census.gov) reports that the total population of Iowa was 276,775 of whom 48,090 were African American. The corresponding figures for individuals 18 and over were 205,785 and 30,451.
Then, he applied formula (4). Notice that the only random component of the formula is the numbers $n_i$ of jurors from the $i$th tract in the venire of size $n$. Thus, the only source of random variation in formula (4) comes from the first stage of a probability sample as the number of African Americans among the individuals from each tract is set equal to their expected number. In the Appendix, the appropriate standard deviation of statistic (4) is derived. Its normal form similar to the usual statistic (2) is

$$Z_T = \frac{\sum_{i=1}^{k} n_i \pi_i - n \pi}{\sqrt{n\left(\sum_{i=1}^{k} f_i \pi_i^2 - \pi^2\right)}}. \quad (7)$$

Our preferred analysis pools the data for both periods. Then, the numerator of (7) equals $-36.70$, the denominator $= 9.19$ and (7) equals $-3.99$ or a shortfall of virtually 4 standard deviations, which is highly significant. For the sample of 929 individuals on the venire in the first period, the value of the numerator of (7) is $-12.15$. The denominator of (7) becomes 4.967, so the statistic (7) becomes $-2.45$, which is significant at the 0.05 level. For the 2252 individuals on venires in the next year, the numerator of (7) equals $-24.54$, the denominator is 7.73 and statistic (7) becomes $-3.17$ standard deviations. Thus, the shortfall of African Americans on the venires in both periods or in the second period by itself is statistically significant at the 3-standard deviation level; the most stringent criteria suggested by the Court in Castaneda. In contrast to the calculations submitted to the Court, now the shortfall in the first period alone satisfies the 2-standard deviation criterion. Using the correct standard deviation in the analysis of the data indicates that it supports the defendant’s claim that African Americans were underrepresented on the venires.

Turning to the third Duren criterion, one needs to test whether the change in the assignment of jurors to the local courts first and then to the Circuit court during the first period to the reverse in the second that occurred after the first period created a statistically significant difference between the minority fraction of the Circuit court venires in the two periods. Since the CD, given by (5), is just $T$ divided by the constant, $n \pi$, the sampling distributions of one CD or the difference between the CDs of two independent samples are readily obtained from the result given in the Appendix. Because 1 standard deviation of the sampling distribution of the difference is 0.0873, while the difference between the two CDs is only 0.03, the normal form or number of standard deviations from expected is only 0.3441 (two-sided $p$ value $= 0.73$). This result is quite far from the 2-standard deviation criterion or $p$ value less than 0.05 and confirms the conclusion of the Supreme Court and the dissenting judge in the state court of appeals who noted the small difference in the proportion of minorities on the venires of both periods.

Comments: 1) The statistic $T$ given by (4) is shown in the Appendix to be equivalent to a statistic that measures how the deviation from the expected number of jurors from a tract is related to the difference between the minority fraction ($\pi_i$) of residents of the tract and their fraction ($\pi$) of the entire county. Thus, $T$ and consequently its normal form (7) already reflect the correlation between the under- or over-representation of a tract and the relative size of its minority population that is demonstrated in Tables 2 and 4. Using $T$ avoids the problem of defining how to classify each tract as being under, over or fairly represented.

2) It is important to distinguish between the hypotheses being tested by the usual statistic given by formula (2) and the variant (7). Since the race–ethnic–gender composition of venires is usually recorded, statistic (2) considers the data as a random sample from the age-eligible population and tests whether the percentage of the venire from the distinct subgroup is consistent with their
percentage in the population. Because Kent County did not record this information on members of the venire, the expert compared the expected number of African Americans on the venire based on the minority proportion in the census tracts where the venire members resided with the expected number in the population, statistic (4). This statistic tests whether the tracts where minorities reside are fairly represented on the venires. While the two hypotheses are clearly related to each other, they are not identical; the second can be regarded as an approximation to the first. In particular, the second test assumes that in the selections within each tract, minority members are selected in accordance with their share of the tract’s population in the appropriate age range. Thus, deviations from random selection within each tract are not captured by the test given by (7), which needed to be used in the case.

3) Using the correct standard deviation in the method discussed in Section 5.6 for assessing the effect of the legitimate exclusions would enable a court to better assess whether they fully explained the disparity. Unfortunately, given the available information, one cannot reliably carry out such a calculation.

6. Conclusions

Until now many courts have required an AD of at least 10% between a minority group’s percentage on venires and in the relevant population. As many commentators have observed and seen again in the data from Alston in Section 3.3, this allows small minority groups to be totally excluded from jury service. In Berghuis, the minority group in question formed 7.28% of the age-eligible population, so requiring a 10% difference between it and the percentage on venires would clearly be inappropriate. The Court did not need to address this issue in order to resolve the case. The data in Berghuis, however, suggest that the case is statistically “close” with respect to the issue of underrepresentation. While the AD is only about 1.1%, the CD is in the range 15–18% and the SR is 82%, just exceeding the 80% threshold used in disparate impact equal employment cases.

If one applies the usual standard deviation analysis by assuming the estimated number of African Americans is the result of a random sample from the age-eligible population of the county, the hypothesis test showed that the underrepresentation on venires of African Americans is not statistically significant for the 6-month period that included the trial. As shown in Section 5.1, the sample of 929 is inadequate for the statistical test to have high power to detect a meaningful shortfall, e.g. SR = 0.70. Thus, one should consider larger samples. Because the data for both periods were homogenous, they should be pooled for analysis. Then, the usual statistic yields a shortfall of about 2.5 standard deviations. More importantly, using the correct standard deviation of the distribution of the statistic actually computed by the defendant’s expert, one concludes that the tracts in which African Americans reside were statistically significantly underrepresented in the venires for all three time frames; meeting the 2-standard deviation level for the 6-month period, exceeding the 3 standard deviation criterion for the year following defendant’s trial and nearly reaching 4 standard deviations for both years combined. Thus, calculation of the statistical test applied by the expert, incorporating its correct standard deviation supports for the Sixth Circuit’s finding that the statistics satisfied the underrepresentation criterion in Duren. The statistical test of the equality of the CDs in the two periods, which used different assignment systems, however, indicates that the difference of about 3%

150 Even though the source lists for potential jurors did not contain racial identification, the County could have recorded this information for the individuals who were on the venire when they came to the court.
between the two CDs was not even “close” to statistical significance. Thus, the underrepresentation could not be ascribed to the assignment system used at the time of defendant’s trial as the data in the record did not meet the third criterion.

It is worth noting that the county could have recorded the race of members of the venires, which would have enabled the usual analysis based on statistic (2) to be carried out. Because the DMV data did not identify the race of potential jurors and the County did not preserve the race of potential jurors, the racial mix of the venire had to be estimated from the residence data. It is likely that this method overestimated the number of African Americans on the jury panels as any differential effect of the selection process between African Americans and whites within each tract is ignored. Thus, the degree of underrepresentation of minorities is probably slightly underestimated.

Several other important statistical issues were not addressed in the prior proceedings. In determining that 7.28% of the age-eligible population in the community should be African Americans, the defendant’s expert assumed that all individuals over 70 years of age would request an exemption. While most people take advantage of such exemptions, it is unclear that all will do so. Given the age structure of the population, this would likely lead to a slight overestimate\(^\text{151}\) of the African American percentage of age eligible. Neither party offered to adjust the percentage to account for other legitimate factors that a state could consider. In particular, the Sixth Circuit found that excusing individuals who have child care responsibilities was justified. If the state could justify excusing all single household heads in poverty with a child of five or less from the pool of potential jurors, the approximation to the usual “standard deviation” analysis on all three date sets became nonsignificant, although the analysis of the combined data would almost reach significance. Moreover, the AD would be reduced to nine-tenths of 1%, the CD to 12.99% and the SR to 86.17%. The authors have not come across a case finding that data with this amount of underrepresentation established a\(^\text{prima facie}\) case of either systematic underrepresentation or a violation of the equal protection clause. Before reaching the conclusion that this one factor would “explain” the shortfall, not only should the data be analysed with the correct standard deviation but the impact of other reasons for excusing individuals, such as medical or job-related reasons should be explored. Hopefully, the Court will require the states and local jurisdictions to preserve information about the race and gender and reason individuals are excused from jury duty in the future. Then, one can compare the demographic composition of the potential jury pool to the census data before individuals are excused from service. This will enable the impact of the exemption and excusal practices to be examined and ensure that while jury service is not an undue burden on individuals, they do not lead to a legally meaningful underrepresentation of important subgroups of the population.\(^\text{152}\)

Note: Although several amicus briefs were filed in the case, only one submitted by a group of well-known law professors, social scientists and statisticians focuses on statistical issues. They agree with the authors that the Court should not require defendants to demonstrate an AD of 10% and that a relative measure is preferable. The brief goes further and recommends that a CD of 15% be required for subgroups forming at least 10% of the age-eligible population. For smaller subgroups, they would

\(^{151}\) According to the 1990 Census data for Kent County, 6.64% of African Americans over 18 were at least 70 years of age. In contrast, 10.89% of the white population 18 years or older were at least 70.

\(^{152}\) The Court indicated in \textit{Duren}, n. 5 at 370 that reasonable exemptions would lead to some over or under inclusiveness and reiterated the view expressed in \textit{Taylor}, 419 U.S. at 534 that this should not result in an underrepresentation sufficiently severe as to constitute a violation of the constitution.
require a CD of 25% or more. The brief does not describe a reanalysis along the lines described here nor does it consider the correct standard deviation analysis for the statistical comparison used by the defendant’s expert.

Acknowledgements

The authors requested information from both parties and we thank Mr. James Lawrence, the lawyer for the defendant, for sending us the expert report and related materials. We also thank Mr. Bradley Hall for providing us with the expert reports from two subsequent jury discrimination cases in Kent County and for several helpful comments. Finally, we thank the referee and associate editor for many helpful suggestions.

REFERENCES


153 Brief for Social Scientists, Statisticians and Law Professors at 33. As Detre (1994 at 1922 n. 42) observed changing the criterion at an arbitrary line leads to inequitable results in borderline cases. For example, under the proposed criterion, a minority group forming 10.5% of the jury-eligible population will be classified as underrepresented if their percentage of the jury pools is less than 8.925%, corresponding to an AD = 1.575%. A minority group composing 9.5% of the jury-eligible population would be classified as underrepresented if its percentage on the venires was less than 7.125%, corresponding to an AD of 2.1375%. In the first case, an SR just below 0.835 is classified as underrepresented but in the second case, an SR less than 0.731 is required.

Appendix. Derivation of the variance of the statistic $T$ comparing the two expected numbers of minority members on the venires

Assume a jurisdiction contains $k$ tracts with a total of $N_i$ individuals, including $B_i$ African Americans in the $i$th tract. The minority fraction, $f_i$ of the total population residing in the $i$th tract is $N_i/N$, where $N$ is the total population of all the tracts $(\sum_{i=1}^{k} N_i)$ and the minority fraction of the $i$th tract is $\pi_i = B_i/N_i$. Let $B$ denote the total minority population in all the tracts and $\pi = B/N$ is the minority fraction of the population of the jurisdiction. Assume one takes a random sample of size $n$ from the population and denote by $n_i$ the number of individuals in the sample who come from the $i$th tract. Then, the statistic

$$T = \sum_{i=1}^{k} n_i \pi_i - n\pi \quad (8)$$

has expectation 0 and variance given by

$$V(T) = n\left(\sum_{i=1}^{k} f_i \pi_i^2 - \pi^2\right) \quad (9)$$

Furthermore, as the sample size $n$ increases the standardized form, $Z_T$, given by (7) of $T$ approaches a normal distribution with mean 0 and variance 1.

Since the expected value of each $n_i$ equals $nf_i$ and $\pi = \sum_{i=1}^{k} f_i \pi_i$, $E(T) = 0$. Since the random $\{n_i\}$ have a multinomial distribution with parameters $n$ and $(f_1, f_2, \ldots, f_k)$, so the vector of cell proportions $\{n_i/n\}$ has mean $\{f_i\}$. $\text{Var}(n_i/n) = f_i(1 - f_i)/n$ and for each distinct pair $i, j \text{ Cov}(n_i/n, n_j/n) = -f_i f_j/n$ and the joint distribution can be approximated by the multivariate normal with the same moments (Severini 2005, p. 41, 120 and 377). Since the random component, the first term, of $T$ is a linear combination of the cell counts; the limiting normality of $Z_T$ follows from the fact that linear combinations of normal variables also follow a normal distribution.
Because of the large number, the accuracy of the normal approximation to the distribution of the $T$ statistic for three sample sizes (929, 2252, and 3181) in the case was checked by simulation. $T$ was calculated on 500 samples from multinomial data with the parameters specified by null hypothesis. For the two largest sample sizes, the normal approximation was accurate. The means were close to 0 ($-0.6$ and $-0.4$) and the standard deviations of 7.89 and 9.58 were close to the values 7.73 and 9.19 obtained from formula (9). For the sample of 929, the Shapiro–Wilk test rejected the normal fit to the distribution of $T$ ($p$ value $<0.0001$). The test statistic $T$ calculated from the actual data ($-12.16$) fell below the first percentile of the simulated null distribution, and the two-sided $p$ value was less than 0.02, again leading to rejection of the null hypothesis.

Notes: 1) The difference in the variances of $T$ and the variance $n\pi(1-\pi)$ of the usual test based on the difference between $X$, the number of minority venire members, and its expected value $n\pi$, is $\sum_{i=1}^{k} n_i\pi_i(1-\pi_i)$. Each term $n_i\pi_i(1-\pi_i)$ is the variance of the number of minority members in a random sample of residents from the $i$th tract. Because $T$ estimates the number of minority individuals from the $i$th tract from the residence pattern that component of sampling variation, which enters into the usual statistic (2), does not contribute to the variability of $T$. A more sophisticated derivation uses the fact that the statistic $T$ is the conditional expectation of $X$, given the $n_i$ and the overall $V(X)$ is the sum of the variance of its conditional expectation plus the expectation of its conditional variance given the $n_i$. This is why $V(T)$ is less than $V(X)$, which leads to more significant findings.

2) Because the CD (5) equal $T/n\pi$ its variance is $V(T)/(n\pi)^2$. This is used in the test of whether the difference of 3% in the CDs of the two periods was statistically significantly different from zero reported in Section 5.6.

3) To examine the relationship between the under- or over-representation of a tract on the venires and the minority proportion of the population of the tract, the defendant’s expert ordered the tracts by the minority proportion. Then, the tracts were classified into three categories. Whether the residents of the tract were under, fairly or over represented on the venires was determined from the $p$ value of the usual statistic as the number of individuals from the tract follows a binomial distribution with parameters $n$ and $f_i$. The use of a 0.25 level of significance is somewhat arbitrary as is the use of three categories, rather than five, say. An alternative approach measures the association between the deviation of the number ($n_i$) of residents of the tract from its expected value ($nf_i$) and the deviation between the minority proportion ($\pi_i$) of the tract and $\pi$. This statistic is

$$R = \sum_{i=1}^{k} (n_i - nf_i)(\pi_i - \pi).$$

Under random sampling, the expected value of $R$ is 0 as the expected value of $n_i = nf_i$. When tracts with large minority populations are underrepresented, the first expression will be negative but the second will be positive. Thus, $R$ will tend to be negative if tracts with relatively large minority populations are underrepresented on the venires. Routine algebra shows that $R$ reduces to $T$. Hence, the statistic $T$ used by the defendant’s statistician measures the relationship between the shortfall of venire members from a census tract and the minority fraction of the population of the tract. Thus, the simpler $T$ statistic reflects the same association seen in Tables 2 and 4 and avoids the problem of choosing the cut-off values used to classify a tract as under, fairly or over represented. As seen in Section 5.6, when the correct variance of $T$ is used, a statistically significant negative association
is found between the minority fraction of the age-eligible population of the tract and the number of residents from the tract serving on the venires.

4) The formulas derived here may be useful in the resolution of two other cases concerning underrepresentation of African Americans on Kent County venires. Apparently, a computer problem led to fewer minorities being called for service in 2001–2002. These cases relied on population data by ZIP code rather than tracts and their experts reported the African American fraction of the age-eligible population in each tract and the number of venire members from each ZIP code during the first three months of 2002. There were 3898 individuals called for possible jury service during the period and African Americans formed 7.84% of the age-eligible population in the County. Thus, one expects 305.6 of them to be African American. Using the ZIP code data, 165.1 African Americans were estimated to have been on the venires, i.e. they formed 4.24% of the venires. Although the AD is 3.60%, the CD is −46% and the SR = 0.52. This implies that an age eligible African American residing in Kent County had just over one-half the chance of being in a venire as a similar white. The normal form (7) of the test based on $T$ is $Z = \frac{-22.16}{\sqrt{2}}$, with a corresponding $p$ value of less than 1 in a million. Thus, the data in these new cases are statistically more convincing on the issue of underrepresentation than the data in Berghuis. If the jury composition data for the years before the computer problem arose was similar to the data in during both periods in Berghuis then comparing the change in the CD of −15.85% to −45.98% indicates that the additional shortfall resulted from the computer problem. A formal test of the significance of the difference (−30.13%) of the two CDs yields a $Z$ of −6.73, corresponding to a $p$ value less than 1 in a million. This highly significant result shows that the degree of underrepresentation of African Americans in the period when there was a computer problem in the jury selection process was noticeably greater than in the earlier period.

154 Powell v. Howes Case No. 05-71345 (ED. MI) and People v. Ramone Bryant (Kent Docket No. 01-08625-FJ).
155 A Demographic Report of Kent County’s Minority Population and its relation to Jury Pool selection by of Drs Lyke Thompson and Mr Douglas Towns, December 2007 at Table 5.
156 A Statistical Analysis regarding the Potential Systematic Exclusion of Black or African Americans in the Jury Pools for the Kent County Circuit Court in January 2002, by Dr Paul L. Stephenson at page 9, Table A.