

Testimony of

William O. Jenkins

June 17, 2008

Statement of William O. Jenkins, Jr., Director

Homeland Security and Justice

June 17, 2008

## FEDERAL JUDGESHIPS

### General Accuracy of District and Appellate Judgeship Case-Related Workload Measures

Twice a year, the Judicial Conference, the federal judiciary's principal policymaking body, assesses the need for additional judges. The assessment is based on a variety of factors, but begins with quantitative case-related workload measures. This testimony focuses on (1) whether the judiciary's quantitative case-related workload measures from 1993 were reasonably accurate; and (2) the reasonableness of any proposed methodologies to update the 1993 workload measures. The comments in this testimony are based on a report GAO issued in May 2003.

### What GAO Recommends

In 2003, GAO recommended that the Judicial Conference, among other things, develop a methodology for measuring the case-related workload of courts of appeals judges by using methodologies that support objective, statistically reliable means of calculating the accuracy of the weights and workload measures, respectively. The Conference disagreed and stated that, among other things, GAO's report did not reflect the sophisticated methodology of the study and that the workloads of the courts of appeals entail important factors that have defied measurement. GAO believes the importance and costs of creating new judgeships requires the best possible case-related workload data to support the assessment of the need for more judges.

In 2003, GAO reported that the 1993 district court case weights were reasonably accurate measures of the average time demands that a specific number and mix of cases filed in a district court could be expected to place on the district judges in that district. At the time of GAO's 2003 report, the Judicial Conference was using case weights approved in 1993 to assess the need for additional district court

judgeships. The weights were based on data judges recorded about the actual in-court and out-of-court time spent on specific cases from filing to disposition. This methodology permitted the calculation of objective, statistical measures of the accuracy of the final case weights.

In 2003, GAO reviewed the research design the Judicial Conference's Subcommittee on Judicial Statistics had approved for updating the 1993 district court case weights, and had two concerns about the design. First, the design assumed that the judicial time spent on a case could be accurately estimated by viewing the case as a set of individual tasks or events in the case. Information about event frequencies and, where available, time spent on the events would be extracted from existing databases and used to develop estimates of the judge-time spent on different types of cases. However, for event data, the research design proposed using data from two data bases that had yet to be integrated to obtain and analyze the data. Second, unlike the methodology used to develop the 1993 case weights, the design for updating the case weights included limited data on the time judges actually spent on specific types of cases. Specifically, the proposed design included data from judicial databases on the in-court time judges spent on different types of cases, but did not include collecting actual data on the noncourtroom time that judges spend on different types of cases. Instead, estimates of judges' noncourtroom time were derived from the structured, guided discussions of about 100 experienced judges meeting in 12 separate groups (one for each geographic circuit). Noncourtroom time was likely to represent the majority of judge time used to develop the revised case weights. The accuracy of case weights developed on such consensus data cannot be assessed using standard statistical methods, such as the calculation of standard errors. Thus, it would not be possible to objectively, statistically assess how accurate the new case weights are--weights on whose reasonable accuracy the Judicial Conference relies in assessing judgeship needs.

The case-related workload measure for courts of appeals judges is adjusted case filings in which all cases are considered to take an equal amount of judge time except for pro se cases--those in which one or more of the parties is not represented by an attorney--which are discounted. In our 2003 review, we found no empirical basis on which to assess the accuracy of this workload measure. Although a number of alternatives to the adjusted filings measure have been considered, the Judicial Conference has been unable to agree on a different approach that could be applied to all courts of appeal.

Mr. Chairman and Members of the Committee:

I appreciate the opportunity to be here today to discuss our work on case-related workload measures for district court and courts of appeals judges. My statement today is based on work completed and reported in 2003 and is focused exclusively on these workload measures.<sup>1</sup> We have no views on the Judicial Conference's pending request for additional judgeships.

Biennially, the Judicial Conference, the federal judiciary's principal policymaking body, assesses the judiciary's needs for additional judgeships.<sup>2</sup> If the Conference determines that additional judgeships are needed, it transmits a request to Congress identifying the number, type, (courts of appeals, district court), and location of the judgeships it is requesting.

In assessing the need for additional judgeships, the Judicial Conference considers a variety of information, including responses to its biennial survey of individual courts, temporary increases or decreases in case filings and other factors specific to an individual court. However, the Judicial Conference's analysis begins with the quantitative case-related workload measures it has adopted for the district courts and courts of appeals--weighted case filings and adjusted case filings, respectively. These two measures recognize, to different degrees, that the time demands on judges are largely a function of both the number and complexity of the cases on their dockets. Some types of cases may demand relatively little time and others may require many hours of work. Generally, each case filed in a district court is assigned a weight representing the average amount of judge time the case is expected to require. The weights are relative to one another; the higher the case weight, the greater the time the case would be expected to require. For example, on average a case with a relative weight of 2.0 would be expected to require twice as much judge time as a case with a weight of 1.0. In the courts of appeals, all case filings are weighted equally at 1.0, except for pro se case filings--those in which one or both parties are not represented by an attorney--which are discounted.

Using these measures, individual courts whose past case-related workload meets the threshold established by the Judicial Conference may be considered for additional judgeships. These thresholds are 430 weighted case filings per authorized judgeship for district courts and 500 adjusted case filings per three-judge panel of authorized judgeships for courts of appeals (courts of appeals judges generally hear cases in rotating panels of three judges each).<sup>3</sup> Authorized judgeships are the total number of judgeships authorized by statute for each district court and court of appeals.

The Judicial Conference relies on these quantitative workload measures to be reasonably accurate measures of judges' case-related workload. Whether these measures are reasonably accurate rests in turn on the soundness of the methodology used to develop them. This statement provides information on two of the objectives in our 2003 report: (1) whether the judiciary's quantitative case-related workload measures were reasonably accurate measures of district judge and courts of appeals judges' case-related workload; and (2) the reasonableness of any proposed methodologies to update the workload measures. In this statement, we discuss those two objectives first for district courts then for courts of appeals.

Our 2003 report was based on the results of our review of documentation provided by the Federal Judicial Center (FJC) and the Administrative Office of the U.S. Courts (AOUSC) on the history and development of the case-related workload measures and interviews with officials in each organization. The scope of our work did not include how the Judicial Conference used these case-related workload

measures to develop any specific request for additional district and courts of appeals judgeships. We conducted our performance audit in April and May 2003 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

## Summary

District Courts. In 2003, we reported that the methodology used to develop the 1993 district court case weights resulted in reasonably accurate measures of the average time demands that a specific number and mix of cases filed in a district court could be expected to place on the district judges in that district. At the time of our 2003 report, the Judicial Conference was using case weights approved in 1993 to assess the need for additional district court judgeships. The weights were based on data judges recorded about the actual in-court and out-of-court time spent on specific cases from filing to disposition. This methodology permitted the calculation of objective, statistical measures of the accuracy of the final case weights (e.g., standard errors).

In 2003 we reviewed the research design the Judicial Conference's Subcommittee on Judicial Statistics had approved for updating the 1993 district court case weights, and had two principal concerns about the design. First was the challenge of collecting reliable, comparable data for the analysis on in-court events from two different automated data systems, one of which had not been implemented in all district courts. The FJC established a technical advisory group to work through this issue. Second, unlike the methodology used to develop the 1993 case weights, the design for updating these case weights included limited data on the time judges actually spent on specific types of cases. Specifically, the proposed design included data from judicial databases on the in-court time judges spent on different types of cases, but did not include collecting actual data on the noncourtroom time that judges spend on different types of cases. Instead, estimates of noncourtroom time would be based on estimates derived from the structured, guided discussions of about 100 experienced judges meeting in 12 separate groups (one for each geographic circuit). Noncourtroom time was likely to represent the majority of judge time used to develop the revised case weights. The accuracy of case weights developed on such consensus data cannot be assessed using standard statistical methods, such as the calculation of standard errors. As the Federal Judicial Center acknowledged in commenting on our 2003 report, it is not possible to objectively, statistically assess how accurate the new case weights are.<sup>4</sup>

## Courts of Appeals.

Adjusted case filings, used to measure the case-related workload of courts of appeals judges, are based on available data from standard statistical reports from the courts of appeals. Unlike the case weights

used to measure district judge case-related workload, adjusted case filings are not based on any empirical data regarding the time that different types of cases required of courts of appeals judges. The adjusted filings workload measure basically assumes that all cases have an equal effect on judges' workload with the exception of pro se cases--those in which one or both parties are not represented by an attorney--which are weighted at 0.33, or one-third as much as all other cases, which are weighted at 1.0. On the basis of the documentation we reviewed, there is no empirical basis on which to base that assumption or on which to assess the accuracy of adjusted filings as a measure of case-related workload for courts of appeals judges. Although a number of alternatives to the adjusted filings measure have been considered, the Judicial Conference has not been able to agree on a different approach that could be applied to all courts of appeals.

#### Case Weights Are Intended to Measure Judicial Time Required to Handle Their Caseloads

The demands on judges' time are largely a function of both the number and complexity of the cases on their dockets. To measure the case-related workload of district court judges, the Judicial Conference has adopted weighted case filings. The purpose of the district court case weights was to create a measure of the average judge time that a specific number and mix of case filed in a district court would require. Importantly, the weights were designed to be descriptive not prescriptive--that is, the weights were designed to develop a measure of the national average amount of time that judges actually spent on specific cases, not to develop a measure of how much time judges should spend on various types of cases. Moreover, the weights were designed to measure only case-related workload. Judges have noncase-related duties and responsibilities, such as administrative tasks, that are not reflected in the case weights.

With few exceptions, such as cases that are remanded to a district court from the court of appeals, each civil and criminal case filed in a district court is assigned a case weight. For example, in the 2004 case weights, drug possession cases are weighted at 0.86 while civil copyright and trademark cases are weighted at 2.12. The total annual weighted filings for a district are determined by summing the case weight associated with all the cases filed in the district during the year. A weighted case filings per authorized judgeship is the total annual weighted filings divided by the total number of authorized judgeships. For example, if a district had total weighted filings of 4,600 and 10 authorized judgeships, its weighted filings per authorized judgeships would be 460. The Judicial Conference uses weighted filings of 430 or more per authorized judgeship as an indication that a district may need additional judgeships. Thus, a district with 460 weighted filings per authorized judgeship could be considered for an additional judgeship. However, the Judicial Conference does not consider a district for additional judgeships, regardless of its weighted case filings, if the district does not request any additional judgeships.

1993 Case Weights Reasonably Accurate, But Accuracy of 2004 Case Weights Cannot Be Statistically Determined

In our 2003 report, we found the district court case weights approved in 1993 to be a reasonably accurate measure of the average time demands a specific number and mix of cases filed in a district court could be expected to place on the district judges in that court. The methodology used to develop the weights used a valid sampling procedure, developed weights based on actual case-related time recorded by judges from case filings to disposition, and included a measure (standard errors) of the statistical confidence in the final weight for each weighted case type. Without such a measure, it is not possible to objectively assess the accuracy of the final case weights.

At the time of our 2003 report, the Subcommittee on Judicial Statistics of the Judicial Conference's Judicial Resources Committee had approved the research design for revising the 1993 case weights, with a goal of having new weights submitted to the Resources Committee for review in the summer of 2004. The design for the new case weights relied on three sources of data for specific types of cases: (1) data from automated databases identifying the docketed events associated with the cases; (2) data from automated sources on the time associated with courtroom events for cases, such as trials or hearings; and (3) consensus of estimated time data from structured, guided discussion among experienced judges on the time associated with noncourtroom events for cases, such as reading briefs or writing opinions.

According to the FJC, the Subcommittee wanted a study that could produce case weights in a relatively short period of time without imposing a substantial record-keeping burden on district judges. The FJC staff provided the Subcommittee with information about various approaches to case weighting, and the Subcommittee chose an event-based method--that is, a method that used data on the number of and types of events, such as trials and other evidentiary hearings, in a case. The design did not involve the type of time study that was used to develop the 1993 case weights. Although the proposed methodology appeared to offer the benefit of reduced judicial burden (no time study data collection), potential cost savings, and reduced calendar time to develop the new weights, we had two areas of concern--the challenge of obtaining reliable, comparable data from two different data systems for the analysis and the limited collection of actual data on the time judges spend on cases.

First, the design assumed that judicial time spent on a given case could be accurately estimated by viewing the case as a set of individual tasks or events in the case. Information about event frequencies and, where available, time spent on the events would be extracted from existing administrative data bases and report and used to develop estimates of the judge-time spent on different types of cases. For event data, the research design proposed using data from two data bases (one of which was new and had not been implemented in all district courts) that would have to be integrated to obtain and analyze the event data. The FJC proposed creating a technical advisory group to address this issue.

Second, the research design did not require judges to record time spent on individual cases. Actual time data would be limited to that available from existing data bases and reports on the time associated with

courtroom events and proceedings for different types of cases. However, a majority of district judges' time is spent on case-related work outside the courtroom. The time required for noncourtroom events would be derived from structured, guided discussion of groups of 8 to 13 experienced district court judges in each of the 12 geographic circuits (about 100 judges in all). The judges would develop estimates of the time required for different events in different types of cases within each circuit using FJC-developed "default values" as the reference point for developing their estimates. These default values would be based in part on the existing case weights and in part on other types of analyses. Following the meetings of the judges in each circuit, a national group of 24 judges (2 from each circuit) would consider the data from the 12 circuit groups and develop the new weights.

The accuracy of judges' time estimates is dependent upon the experience and knowledge of the participating judges and the accuracy and reliability of the judges' recall about the average time required for different events in different types of cases--about 150 if all the case types in the 1993 case weights were used. These consensus data could not be used to calculate statistical measures of the accuracy of the resulting case weights. Thus, the planned methodology did not make it possible to objectively, statistically assess how accurate the new case weights are--weights whose accuracy the Judicial Conference relies upon in assessing judgeship needs.

We noted that a time study conducted concurrently with the proposed research methodology would be advisable to identify potential shortcomings of the event-based methodology and to assess the relative accuracy of the case weights produced using that methodology. In the absence of a concurrent time study, there would be no objective statistical way to determine the accuracy of the case weights produced by the proposed event-based methodology--a major difference with the methodology used to develop the 1993 case weights.

#### Accuracy of Courts of Appeals Case-Related Workload Measure Cannot Be Assessed

The principal quantitative measure the Judicial Conference uses to assess the need for additional courts of appeals judgeships is adjusted case filings. The measure is based on data available from standard statistical reports for the courts of appeals. The adjusted filings workload measure is not based on any empirical data regarding the time that different types of cases required of appellate judges.

The Judicial Conference's policy is that courts of appeals with adjusted case filings of 500 or more per three-judge panel may be considered for one or more additional judgeships. Courts of appeals generally decide cases using constantly rotating three-judge panels. Thus, if a court had 12 authorized judgeships, those judges could be assigned to four panels of three judges each. In assessing judgeship needs for the courts of appeals, the Conference may also consider factors other than adjusted filings, such as the geography of the circuit or the median time from case filings to disposition.<sup>5</sup>

Essentially, the adjusted case filings workload measure counts all case filings equally, with two exceptions. First, cases refiled and approved for reinstatement are excluded from total case filings.<sup>6</sup> Second, pro se cases--defined by the Administrative Office of the U.S. Courts as cases in which one or both of the parties are not represented by an attorney--are weighted at 0.33, or one-third as much as other cases, which are weighted at 1.0. For example, a court with 600 total pro se case filings in a year would be credited with 198 adjusted pro se case filings ( $600 \times 0.33$ ). Thus, a court of appeals with 1,600 filings (excluding reinstatements)--600 pro se cases and 1,000 non-pro se cases--would be credited with 1,198 adjusted case filings (198 discounted pro se cases plus 1,000 non-pro se cases). If this court had 6 judges (allowing two panels of 3 judges each), it would have 599 adjusted case filings per 3-judge panel, and, thus, under Judicial Conference policy, could be considered for an additional judgeship.

The current court of appeals workload measure represents an effort to improve the previous measure. In our 1993 report on judgeship needs assessment, we noted that the restraint of individual courts of appeals, not the workload standards, seemed to have determined the actual number of appellate judgeships the Judicial Conference requested.<sup>7</sup> At the time the current measure was developed and approved, using the new benchmark of 500 adjusted case filings resulted in judgeship numbers that closely approximated the judgeship needs of the majority of the courts of appeals, as the judges of each court perceived them. The current courts of appeals case-related workload measure principally reflects a policy decision using historical data on filings and terminations. It is not based on empirical data regarding the judge time that different types of cases may require. On the basis of the documentation we reviewed for our 2003 report, we determined that there is no empirical basis for assessing the potential accuracy of adjusted case filings as a measure of case-related judge workload.

#### Various Proposals Have Been Considered for Changing the Court of Appeals Workload Measure

In the past decade the Judicial Conference has considered a number of proposals for developing a revised case-related workload measure for the courts of appeals judges, but has been unable to reach a consensus on any approach. As part of its assistance to the Conference in this effort, the FJC in 2001 compiled a document that reviewed previous proposals to develop some type of case weighting measure for the courts of appeals. Table 1 outlines some of these proposals and their advantages and disadvantages, as identified by the FJC. Generally, methods that rely principally on empirical data on actual case characteristics and judge behavior (e.g., time spent on cases) are more appropriate than those that rely principally on qualitative data because statistical methods can be used to estimate the accuracy of the resulting workload measure.

1 GAO, Federal Judgeships: The General Accuracy of the Case-Related Workload Measures Used to Assess the Need for Additional district Court and Courts of Appeals Judgeships, GAO-03-788R (Washington, D.C., May 30, 2003).



2 The Chief Justice of the United States presides over the Conference, which consists of the chief judges of the 13 courts of appeals, a district judge from each of the 12 geographic circuits, and the chief judge of the Court of International Trade. The Conference meets twice a year.

3 In the documentation accompanying its 2007 request for additional judgeship, the Judicial Conference notes that in 2004 it adopted a starting point of more than 430 weighted case filings per authorized judgeship with an additional judgeship.

4 We have not reviewed in detail the materials the FJC has posted on its Web site with regard to the methodology actually used to develop the revised case weights approved in 2004. However, those materials indicate that the FJC essentially followed the design we reviewed and that standard errors were not computed for the final weights.

5 At the time of our 2003 report, the FJC had suggested that adjusted case filings may not be an appropriate measure for the D.C. Circuit Court of Appeals, given the distinctive characteristics of the administrative agency appeals that were a major source of that court's caseload. Details on the FJC analysis for the D.C. Circuit can be found in our 2003 report: GAO, Federal Judgeships: The General Accuracy of the Case-Related Workload Measures Used to Assess the Need for Additional district Court and Courts of Appeals Judgeships, GAO-03-788R (Washington, D.C., May 30, 2003).

6 Such cases were dismissed for procedural defaults when originally filed, but "reinstated" to the court's calendar when the case was later refilled. The number of such cases, as a proportion of total case, is generally small.

7 GAO, Federal Judiciary: How the Judicial Conference Assesses the Need for More Judges, GAO/GGD-93-31 (Washington, D.C., Jan. 29, 1993).