



GAO

Accountability * Integrity * Reliability

United States General Accounting Office
Washington, DC 20548

May 30, 2003

The Honorable Lamar Smith
Chairman
Subcommittee on Courts, the Internet,
and Intellectual Property
Committee on the Judiciary
House of Representatives

Subject: *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*

Dear Mr. Chairman:

Biennially, the Judicial Conference, the federal judiciary's principal policymaking body, assesses the judiciary's needs for additional judgeships.¹ If the Conference determines that additional judgeships are needed, it transmits a request to Congress identifying the number, type (courts of appeals, district, or bankruptcy), and location of the judgeships it is requesting. In 2003, the Judicial Conference sent to Congress requests for 93 new judgeships—11 for the courts of appeals, 46 for the district courts, and 36 for the bankruptcy courts.²

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

¹The 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.

²This report covers the methodology used to develop the case-related workload measures for district court and courts of appeals judges. We recently testified on the methodology used to develop the case-related workload measure for bankruptcy judges. (See *Federal Bankruptcy Judges: Weighted Case Filings as a Measure of Judges' Case-Related Workload*, GAO-03-789T (Washington, D.C.: May 22, 2003)).

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. A case with a weight of 3.0, for example, would be expected to take twice as much time as a case with a weight of 1.5. In the courts of appeals, pro se case filings—those in which one or both parties are not represented by an attorney—are weighted at 0.33 and all other case filings at 1.0.

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 the courts of appeals (courts of appeals judges generally hear cases in rotating panels of three judges each). 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. As agreed with your office, our objectives were to (1) determine whether the methods the Judicial Conference uses to quantitatively measure the case-related workload of district court and court of appeals judges results in a reasonably accurate measure of judges' case-related workload, (2) assess the reasonableness of any proposed methodologies to update the workload measures, and (3) obtain information from the Administrative Office of the U.S. Courts (AOUSC) on the steps the Judiciary takes to ensure that the case filing data required for these workload measures are accurate. The information for the last objective is presented in enclosure I. The scope of our work specifically excluded any analysis of how the Judicial Conference used the case-related workload measures to develop its current judgeship request.

Results in Brief

The district court weighted case filings, as approved in 1993, appear to be reasonably accurate and are based on a reasonable methodology. However, they are about 10 years old, and we have concerns about the research design approved to update them.

Overall, the weighted case filings, as approved in 1993, appear to be a reasonably accurate measure 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 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 filing 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 assess the accuracy of the final case weights. However, the case weights are about 10 years old, and the data on which the weights were based are as much as 15 years old. Changes since 1993, such as the characteristics of cases filed in federal district courts and changes in case management practices, may have affected whether the 1993 weights continue to be a reasonably accurate measure of the average time burden on district court judges resulting from a specific volume and mix of cases. Some of these changes may have increased time demands;

others may have reduced time demands. To the extent that the current case weights understate or overstate the total case-related time demands on district judges, the weights could potentially result in the Judicial Conference understating or overstating the need for new district court judgeships.

The Judicial Conference's Subcommittee on Judicial Statistics has approved a research design for updating the current case weights, and we have some concerns about that design. The design would include limited data on the time judges actually spend on specific types of cases. Much of the time data used would be based on consensus estimates from groups of experienced judges. Such data cannot be used to develop an objective, statistical measure of the accuracy of the final case weights. Without such a measure, it is not possible to determine whether the case weights are in fact a reasonably accurate measure of case-related judge workload. In assessing the need for judgeships in specific courts, the Judicial Conference relies on the case weights to be a reasonably accurate measure of judges' case-related workload.

Unlike the district court case weights, the adjusted filings workload measure for appellate judges is 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 a lawyer—which are weighted at 0.33, or one-third as much as all other cases. In the documentation we reviewed, we found no empirical data to support that assumption. The current court of appeals case-related workload measure, adopted in 1996, reflects an effort to improve the previous measure, which may have tended to overstate judgeship needs. 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 of appeals perceived them. However, on the basis of the documentation we reviewed, there is no empirical basis on which to assess the accuracy of adjusted filings as a measure of case-related workload for courts of appeals judges.

Weighted Case Filings: District Judge Case-Related Workload Measure Is Reasonably Accurate, but 10 Years Old, and the Plan to Update It Raises Some Concerns

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 cases 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 types of cases, not to develop a measure of how much time judges should spend on various types of cases. Finally, the weights were designed to measure only case-related judge workload. Judges have noncase-related duties and responsibilities, such as administrative tasks, that are not reflected in the case weights.

Case Weights Measure Average Judicial Time Demands

With a few exceptions, such as cases that are remanded to a district court from the courts of appeals, each civil and criminal case filed in a district court is assigned a case weight that varies from 0.031 (for cases involving defaulted student loans or veterans benefit overpayments) to 5.99 (for death penalty habeas corpus cases) based on the subject matter of the case.³ The weight of the overall average case is 1.0. All other case weights were established relative to this national average case.⁴ Thus, a case with a weight of 0.5 would be expected to require on average about half as much judicial time as the national average case. Conversely, a case with a weight of 2.0 would be expected to take twice as much time as the national average case. Case weights for criminal felony cases are applied on a per defendant basis.⁵ For example, the case weight for heroin/cocaine distribution is 2.27. A heroin/cocaine distribution case with two defendants would be weighted at 4.54—two times the assigned weight of 2.27. The actual amount of time a judge may spend on individual cases of any specific type may be more or less than the national average for that type of case.

The total annual weighted filings for a district are determined by summing the case weights associated with all the cases filed in the district during the year. 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 judgeship 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.

In assessing judgeship needs, the weighted case filings are calculated using authorized judgeships (a number which includes any vacancies). This is a measure of the average workload per judge in a district court if all the court's authorized judgeships were filled. Calculating the weighted case filings per active judge—that is, on the basis of the number of authorized judgeships filled—would show the burden of existing vacancies on active judges, but not necessarily the need for more judgeship positions.

³Weights are assigned to each civil case counted as an original filing, removal from state courts, or interdistrict transfer (transfers from one district to another). Weights are also assigned to each felony defendant counted as an original filing, reopened filing, or interdistrict transfer. Generally, felonies are those crimes that carry a term of imprisonment of more than 1 year. Weights are not assigned to civil cases remanded to the district courts from the courts of appeals, reopened cases, or multidistrict litigation transfers—cases transferred to a single district from a number of districts for disposition, such as asbestos or breast implant litigation.

⁴Some types of civil cases were weighted differently if they involved the United States as a party or were removed from state court to federal court.

⁵The weights do not include nonfelony criminal cases, which are generally the responsibility of magistrate, not district, judges.

Case Weights Calculated in 1993 Using Time Data Recorded by Judges

The Judicial Conference approved the use of the current district court case weights in 1993. The weights are based on a “case-tracking time study,” conducted between 1987 and 1993, in which judges recorded the amount of time they spent on each of their cases included in the time study.⁶ The study included about 8,100 civil cases and about 4,200 criminal cases that were generally “tracked” from filing to disposition.⁷ All judges who worked on each case were supposed to record the time they worked on the case.⁸

Data collection for the time study began in November 1987. Districts were brought into the study over a 2-year period, with the last district entering the study in January 1990. When a district was brought into the study, a 2-week period was designated for sampling, during which all cases filed were included in the time study sample.

At the conclusion of the study, sample cases were grouped into civil and criminal cases, with individual subclassifications (case types) for each, such as "Contract: Insurance" and "Bank Robbery." Each sample case had a value associated with it, which was the total number of minutes reported by the district judge(s) who worked on it. The number of sample cases in the subclassifications ranged from 18 to 1,563. Within each subclassification, a simple average and a standard error were computed. The averages and standard errors were converted into relative values as the final step in creating the case weights—that is, all the weights were calculated relative to the time required for the average case in the study.

Methodology Used to Develop Case Weights Was Reasonable

Overall, the weighted case filings, as approved in 1993, are a reasonably accurate method of measuring the average judge time that a specific number and mix of cases filed in a district court could require. The methodology used to develop the weights is reasonable. It used a valid sampling procedure, developed weights based on actual case-related time recorded by judges from case filing to disposition, and included a measure (standard errors) of the statistical confidence in the final weight for each weighted case type.

⁶The time study for bankruptcy courts was a “diary study” in which judges recorded the time spent on case-related and noncase-related work during a 10-week period. Although each method has different strengths and limitations, each method can produce useful, reasonably accurate results. Enclosure II includes a comparison of these two methodologies.

⁷Not all cases were completed by the end of the study; some were still pending.

⁸This included district judges, senior judges, magistrate judges, and visiting judges. District judges—nonsenior and senior—exercise the full judicial authority vested in the district courts. Nonsenior district judges are those who hold a designated judgeship position and generally carry a full caseload. Senior district judges are judges who have retired from regular, full-time active service but remain on the bench and perform such judicial duties as they are willing and able. Magistrate judges, appointed for a fixed term of years, exercise the judicial duties permissible by statute and the Constitution that the district courts delegate to them. Visiting judges are those visiting from their “home court” to assist in addressing the workload of the court they are visiting. Visiting judges may or may not be senior judges. Time reported by magistrate judges was not included in the final computations of the case weights.

The sampling method was appropriately designed to ensure that all district judges and all case types could potentially be included in the sample. The staggered entries of districts into the study ensured the selection of case samples were taken throughout the year, reducing or eliminating bias due to seasonal variation in case filings. Every district court judge could potentially have been a participant in the study (depending on when the 2-week window was designated at a given district and case assignments during that period).

The method of recording the time spent on each case was designed to capture all judge time spent on a sample case. Although it was not possible to determine if all reportable judge time was in fact recorded and reported, validity checks on the reported time were made where possible. For example, judge-reported courtroom time in each sample case was compared with the time reported for the same case in the judiciary's database on courtroom proceedings.

The empirical data on hours expended on each case in the sample were used to develop the case weights. The case weights for specific types of cases were basically determined by dividing the total amount of time judges reported for that type of case by the number of such cases in the study. For example, if judges reported a total of 2,000 hours for 200 cases of a specific type in the study, this would translate into 10 hours per case. Sampling variability in the estimates based on the time study data was quantified and provided with the weights. The standard error that is associated with each weight provides an indicator of variability due to the weight being produced via a sample, rather than data from the universe of cases during the study period. The standard errors can be used to display the statistical reliability of the weighted case filings estimate for each district. Without some measure of statistical reliability, it is not possible to objectively assess how accurate the case weights are.

The case weights are relative weights. That is, each case weight was calculated relative to the average case as determined in the study, which was assigned a value of 1.0. For example, a case type with a weight of 2.0 would be expected to require twice as much judge time as the "average" case. Relative weights were determined by dividing the absolute weight of each type of case by the weight or value of the average case. The Federal Judicial Center (FJC) converted absolute weights to relative weights by dividing the absolute weight values by 2.132. This value was chosen after FJC conducted research to determine how to produce a new set of relative weights that they considered to be comparable to the previous set of relative weights. As described by FJC officials, this approach was reasonable.

For the purposes of applying the national weights to individual districts, the methodology assumed two things: (1) that the district's judges were typical of district judges as a whole and (2) that the district's cases of any given type were typical of that case type as a whole. This may or may not have been true, but these are reasonable assumptions given the purpose of the study—to develop weights based on national averages, not to develop weights for individual districts or judges.

Research Design for Updating the District Court Case Weights Raises Concerns

The case weights are almost 10 years old, and the time data on which they were based are as much as 15 years old. Changes since the case weights were finalized in

1993, such as changes in the characteristics of cases filed in federal district courts and in case management practices, may affect how accurately the weights continue to reflect the time burden on district court judges today. For example, since 1993, new civil causes of action (such as telemarketing issues) and criminal offenses (new terrorism offenses) needed to be accommodated within the existing case-weight structure. According to FJC officials, where the new cause of action or criminal offense is similar to an existing case-weight type, the weight for the closest case type is assigned. Where the new cause of action or criminal offense is clearly different from any existing case weight category, the weight assigned is that for either “all other civil” for civil cases or “all other criminal” for criminal cases.

The Subcommittee on Judicial Statistics of the Judicial Conference’s Judicial Resources Committee has approved the research design for revising the current case weights, with a goal of having new weights submitted to the Resources Committee for review in the summer of 2004. The research would be led by FJC, who developed the research design. Although the methodology for updating the case weights appears 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 have some concerns about the basic research design.

Our principal concerns are two: the challenge of obtaining reliable, comparable data from two different automated data systems for the analysis and the limited collection of actual data on the time judges spent on cases. Essentially, the design for the new case weights relies on three sources of data for specific types of cases: (1) data from automated databases identifying the docketed events associated with cases; (2) data from automated sources on the time associated with courtroom events for cases; and (3) consensus estimates from structured, FJC-guided discussions among experienced judges on the judge-time required for noncourtroom events in the cases, such as reading briefs or writing opinions. The design assumes that judicial time spent on a given case can 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 administrative databases and reports, and then used to develop estimates of the judge-time spent on different types of cases. For event data, the research design proposes using new technology (the Case Management/Electronic Case Filing system) that is currently being introduced into the court system for recording case management information. However, not all courts have implemented the new system, and data from the existing and new systems will have to be integrated in the study. Successfully integrating the data from these two databases will be a challenge. FJC recognizes this and has developed a strategy for addressing the issues, which includes forming a technical advisory group from FJC, AOUSC, and individual courts to develop a method of reliably extracting and integrating data from the two case management systems for analysis.

Second, the design for developing the new weights does not require judges to record time spent on individual cases. A significant limitation of the time data to be used is that the time data available from existing databases and reports are limited to time associated with courtroom events and proceedings, while a majority of district judges’ time is spent on case-related work outside the courtroom. The time required for noncourtroom events, such as reviewing briefs, will be based on the consensus of groups of experienced judges. Groups of 8 to 13 district judges in each of the 12

circuits (about 100 in all) will meet in a series of structured discussions to 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 this series of meetings, a national group of 24 judges (2 from each circuit), using structured procedures, will consider the data from the 12 circuit groups and develop consensus time estimates for use in developing the weights. These consensus time estimates are likely to represent a majority of the judge time used to develop the new weights. These consensus data are 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 case types in the current case weights were used. The greater the number of events and types of cases for which judges are asked to make estimates, the greater the demands on judges to recall accurately the judge time associated with specific events and types of cases. These consensus data cannot be used to calculate statistical measures of the accuracy of the resulting case weights. Thus, it will not be possible to objectively, statistically assess how accurate the new case weights are—weights on whose reasonable accuracy the Judicial Conference will rely in assessing judgeship needs in the future.

A concurrent time study using "case tracking" or "diary" methods would be advisable to identify potential shortcomings of the event-based procedure and to assess the relative accuracy of the case weights that are produced using that procedure. 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.

Adjusted Case Filings: Courts of Appeals Judge Workload Measure Lacks Empirical Basis for Assessing Its Potential Accuracy

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

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 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 judgeships each. The Conference may also consider factors other than adjusted case filings, such as the geography of the circuit or the median time from case filings to disposition. For 11 of the 12 courts of appeals, the Judicial Conference counts all case filings equally, with two exceptions. (There is no specific workload measure established for the D.C. circuit, as discussed later.) First, cases

refiled and approved for reinstatement are excluded from total case filings.⁹ Second, two-thirds of pro se cases—defined by AOUSC as cases in which one or both of the parties are not represented by legal counsel—are deducted from total case filings (that is, they are effectively weighted at 0.33). For example, a court with 600 total pro se case filings in fiscal year 2001 would be credited with 198 adjusted pro se case filings (600 x 0.33). The remaining nonpro se cases would be weighted at 1.0 each. Thus, a court of appeals with 1,600 case filings (excluding reinstatements)—600 pro se cases and 1,000 nonpro se cases—would be credited with 1,198 “adjusted” case filings (198 discounted pro se cases plus 1,000 nonpro 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 the Judicial Conference’s policy, could be considered for additional judgeships.

The current case-related workload measure for courts of appeals judges, adopted in 1996, is similar in concept to the measure we reviewed in 1993.¹⁰ Table 1 illustrates the similarities and differences in the two measures. Although the current workload measure is expressed in terms of appellate case filings, both the 1986 and 1996 case-related workload measures are based on assumptions about the judge workload associated with merit dispositions. Merit dispositions are cases that are decided on the legal rights of the parties to the case rather than on technical issues, such as lack of federal jurisdiction.

The workload measure we reviewed in 1993 was based on 5-year averages of merit dispositions in each circuit separately, and the result was not necessarily comparable among circuits because of the different methods that each circuit used to decide its cases. The current measure uses a single national standard for all circuits. Using national data on merit dispositions as a percentage of case filings in 1994, the current workload measure was based on the assumption that nationally about 55 percent of all appellate case filings—except for pro se filings and reinstated filings—result in merit dispositions. Thus, 500 adjusted case filings would represent 275 merit dispositions—or 20 more than the 255 used in the 1986 measure. The increase from 255 to 275 was basically a matter of establishing equity between the district courts and courts of appeals workload thresholds. To be considered for additional district court judgeships, the Judicial Conference had raised the threshold from 400 to 430 weighted case filings per judgeship (a 7.5-percent increase). The new merit dispositions standard raised the threshold for courts of appeals from 255 to 275 merit dispositions (a 7.8-percent increase).

⁹Such cases were dismissed for procedural defaults when originally filed but “reinstated” to the court’s calendar when the case was later refiled. The number of such cases, as a proportion of total cases, is generally small.

¹⁰U.S. General Accounting Office, *Federal Judiciary: How the Judicial Conference Assesses the Need for More Judges*, GAO/GGD-93-31 (Washington, D.C.: Jan. 29, 1993).

Table 1: A Comparison of the 1986 and 1996 Methods of Measuring Case-Related Workload for Courts of Appeals Judges

1986	1996
The benchmark for considering additional judgeships in a court of appeals is 255 merit dispositions per 3-judge panel.	The benchmark for considering additional judgeships in a court of appeals is 500 adjusted case filings per 3-judge panel.
Prisoner petition cases (a subset of pro se cases) are counted as one-half of an appellate case filing.	Pro se cases (which include prisoner petitions) are counted as one-third of an appellate case filing.
Uses a 5-year average merit termination rate for each individual circuit.	Uses single standard of 500 adjusted filings for all courts of appeals.
No other adjustments.	Does not count number of appeals reinstated after procedural default as part of adjusted filings to prevent double counting of appeals.
Calculations are based on actual 5-year average merit terminations rate for each court of appeals.	Calculations apply to each circuit's appellate case filings.

Source: FJC documentation and interviews.

The current court of appeals case-related workload measure represents an effort to improve the previous measure. As we noted in our 1993 report, using the previous measure the courts of appeals' own restraint, not the workload standard, seemed to have determined the actual number of appellate judgeships the Judicial Conference requested. 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 court of appeals case-related workload measure principally reflects a policy decision using historical data on filings and terminations. In 1995, the Subcommittee on Judicial Statistics of the Judicial Conference's Judicial Resources Committee sent a survey to the chief judge of each circuit court of appeals. In the responses, there was no agreement that either the 500 adjusted filings standard or a weight of 0.33 for pro se cases were the appropriate standards. Unlike the district court case weights, the adjusted filings workload measure is not based on empirical data regarding the judge time that different types of case may require. On the basis of the documentation we reviewed, we determined there is no empirical basis for assessing the potential accuracy of adjusted filings as a measure of case-related judge workload.

The D.C. Circuit—Adjusted Case Filings Not Applicable to Its Unusual Caseload

In a report to a Judicial Conference subcommittee,¹¹ FJC discussed some of the distinctive features of the Court of Appeals for the D.C. Circuit. The report noted that approximately 30 percent of the circuit's filings in fiscal years 1996-1997 were administrative agency appeals that occur almost exclusively in the D.C. circuit and were more burdensome than other cases in several aspects. On average, these cases

- had more independently represented participants per case;
- were more likely to have participants with multiple objectives, involve complex or statutory law, and require the mastery of technical or scientific information;
- had more briefs filed per case;
- had a higher proportion of cases that were terminated; and

¹¹Federal Judicial Center, *Assessment of Caseload Burden in the U.S. Court of Appeals for the D.C. Circuit*, Report to the Subcommittee on Judicial Statistics of the Committee on Judicial Resources of the Judicial Conference of the United States (Washington, D.C.: 1999).

- had a higher rate of case consolidation (where two or more cases are combined for decision).

The report concluded that the need for additional judgeships in the D.C. circuit should not be measured using the general workload threshold of 500 adjusted case filings per 3-judge panel. However, because no information was available on judges' actual time expenditures, there was no empirical basis for suggesting a specific alternative formula for assessing the D.C. circuit's judgeship needs. The report also concluded that the D.C. circuit's remaining caseload—that is, all cases other than administrative agency appeals—was generally not distinguishable from the caseloads of the other circuits. The report suggested several possible ways to integrate the D.C. circuit into the existing adjusted weighted filings system, such as giving greater weight to federal agency appeals or lowering the general threshold of 500 adjusted filings per 3-judge panel for the D.C. circuit. The Judicial Conference has not yet adopted any specific workload measure for the D.C. circuit. However, the Judicial Conference requested no additional judgeships for the D.C. circuit in 2003.

No Judicial Conference Consensus on How to Revise Adjusted Filings Workload Measure

In 1993, we recommended that the Judicial Conference improve its workload measure for the courts of appeals.¹² In the last decade, the Judicial Conference has considered a number of proposals for developing a revised case-related workload measure for courts of appeals judges, but the Conference has been unable to reach a consensus on any approach. As part of its assistance to the Conference in this effort, FJC in 2001 compiled a document that reviewed previous proposals to develop some type of case weighting measure for the courts of appeals.¹³ Table 2 outlines some of these proposals and their advantages and disadvantages, as identified by FJC.

¹²U.S. General Accounting Office, *Federal Judiciary: How the Judicial Conference Assesses the Need for More Judges*, GAO/GGD-93-31 (Washington, D.C.: Jan. 29, 1993).

¹³ Federal Judicial Center, *Review of Previous Appellate Case Weighting Proposals*, (Washington, D.C.: Aug. 22, 2001).

Table 2: Past Proposals to Revise the Case-Related Workload Measure for Courts of Appeals Judges

Proposal	Advantages	Disadvantages
1. Estimation of case burden based on actual time required to process the case.	<ul style="list-style-type: none"> The quantitative approach would be very thorough. Empirically based data. 	<ul style="list-style-type: none"> Judges may not be amenable to the time-consuming task of recording the hours spent on individual cases. Time spent gathering data could be used elsewhere.
2. Estimate of case burden based on the assessment of burden of only "certain characteristics" from an already-existing database of "factors."	<ul style="list-style-type: none"> Would not be very time-consuming for judges. Would assess the frequencies of certain "factors." Analysis of an existing database would save time. Can use a "wealth" of factors to get a big picture of the caseload burden. 	<ul style="list-style-type: none"> Difficult to agree on which factors to use. Difficult to decide if presence and absence of factors is enough information. Database and survey accuracy may be compromised.
3. Normative assessment of cases to look qualitatively at the cases as a whole.	<ul style="list-style-type: none"> Convenient to extract information from surveys or group discussions. 	<ul style="list-style-type: none"> Difficult to decide which factors to use. Dependent upon accuracy of judges recall about the case. Lack of empirically based data.
4. Using multiple regression to use information about the proportional mix of cases with different defined characteristics in the different circuits to account for the differences in case termination level.	<ul style="list-style-type: none"> Quantitative approach to determine factors to use. 	<ul style="list-style-type: none"> Use of a potentially incomplete model. Inherent statistical limits. Cannot assess appellate burdens on a national level.
5. Using district court weights for the appellate system.	<ul style="list-style-type: none"> Already available data. Save time by using existing data. 	<ul style="list-style-type: none"> Little consistency between the two court systems. Sacrifice accuracy.
6. Tallying court opinions (published and unpublished).	<ul style="list-style-type: none"> Most appellate judge work leads to production of appellate opinions in chambers. 	<ul style="list-style-type: none"> Necessary information cannot be obtained consistently.
7. Sampling cases for approximately 3 months for a case-based study (Nov. 8, 1993).	<ul style="list-style-type: none"> Can project the results of 3 months of cases, to the rest of the year. 	<ul style="list-style-type: none"> There is no way to anticipate possible sample sizes, so cannot make a statistical prediction.

Source: FJC documentation.

Additionally, there are more proposals that are variations of the above or combinations of the above. Some of these possibilities have more potential than others. Generally, methods that rely principally on empirical data on actual case characteristics and judge behavior (e.g., time expended 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.

Conclusions

Overall, the methodology used to develop the district court case weights is reasonable, and the resulting case weights are a reasonably accurate measure of district court judge case-related workload. However, the weights are about 10 years old, and the time data on which they are based are as much as 15 years old. Consequently, it is uncertain whether the case weights continue to be a reasonably accurate measure of the average district judge time burden resulting from a specific volume and mix of cases. The Judicial Conference's Subcommittee on Judicial Statistics has approved a research design for updating the current case weights, about which we have two concerns. The design would rely in large part on data from

two different case management data systems and it will be a challenge to reliably and usefully integrate the data from these two systems for analysis. FJC recognizes this and is developing a strategy for addressing the issue. Second, the design includes limited actual data on the time district judges spend on different types of cases. All the data on noncourtroom time will be based on estimates developed by 13 groups of experienced judges (about 124 in all) using structured, guided discussions. These data cannot be used to calculate statistical measures of the accuracy of the resulting case weights. Thus, it will not be possible to objectively, statistically assess how accurate the new case weights are—weights on whose reasonable accuracy the Judicial Conference will rely in assessing judgeship needs in the future.

The adjusted case filings workload measure used for the courts of appeals is not based on actual data about the time that courts of appeals judges expend on different types of cases. Rather, it represents a policy judgment of the appropriate workload benchmark for considering new judgeships that is based on an analysis of past trends in case filings and merit dispositions. Because of the lack of empirical data on the time demands on courts of appeals judges, neither we nor the judiciary can assess whether adjusted filings is a reasonably accurate measure of the workload of courts of appeals judges. Any methodology to revise the current workload measure that relies solely on qualitative data is unlikely to provide reasonably reliable and verifiable estimates of judges' workload. In 1993, we recommended that the Judicial Conference develop a better measure of the workload of courts of appeals judges. Although the Conference has studied many potential methods of improving its workload measure, it has been unable to agree on any methodology for doing so.

We recognize that a methodology that provides greater empirical assurance of a workload measure's accuracy will require judges to document how they spend their time on a cases for at least some period of weeks. We believe that, given the importance and cost of federal judgeships, this would be a good investment to ensure that the workload measures that are used to support judgeship requests are reasonably accurate and based on the best data available using sound research methods.

Recommendations

We recommend that the Judicial Conference of the United States

- update the district court case weights using a methodology that supports an objective, statistically reliable means of calculating the accuracy of the resulting weights; and
- develop a methodology for measuring the case-related workload of courts of appeals judges that supports an objective, statistically reliable means of calculating the accuracy of the resulting workload measures and that addresses the special case characteristics of the Court of Appeals for the D.C. Circuit.

Agency Comments and Our Response

We provided the Director of the Administrative Office of the United State Courts and the Director of the Federal Judicial Center with a draft of this report for comment. Both provided technical comments, which were incorporated into the report as appropriate. In a May 27, 2003 letter, the Chair of the Committee on Judicial Resources of the Judicial Conference of the United States provided comments (see enc. III) that offered four major observations: (1) the case-related workload in each court district court and court of appeals for which the Judicial Conference has requested one or more judgeships considerably exceed the minimum thresholds the Conference has established for considering additional judgeships in district courts and courts of appeals; (2) we did not provide the full context in which the Judicial Conference uses the district court case weights in assessing district court judgeship needs; (3) the workload of the courts of appeals entail important factors that have defied measurement, including significant differences in case processing techniques; and (4) we did not fully and accurately describe the full context of the new district court case weighting study.

With regard to the first two observations, the scope of our work was limited to an assessment of the relative accuracy of the weighted case filings and adjusted case filings measures of district court judge and courts of appeals judge workload, respectively. Our report clearly states that the workload measures we reviewed are one of many factors the Judicial Conference considers in assessing judgeship needs, although the assessment begins with these workload measures. With regard to the courts of appeals, we recognize that there are significant methodological challenges in developing a more precise workload measure for the courts of appeals. However, using the data available, neither we nor the Judicial Conference can assess the accuracy of adjusted case filings as a measure of the case-related workload of courts of appeals judges. We believe it is premature to conclude that it is not possible to develop a case-related workload measure for courts of appeals judges whose accuracy can be reasonably determined.

The Deputy Director of FJC provided comments in a May 27, 2003 letter (see enc. IV). Both the FJC Deputy Director and the Chair of the Judicial Conference's Committee on Judicial Resources said that we did not fully describe the proposed methodology for updating the district court case weights and why this methodology could produce case weights whose accuracy could be reasonably assessed. We have added language to the report that provides more detail on the iterative Delphi technique that would be used to develop the consensus estimates of the judge time required for noncourtroom events in many different types of cases. FJC agrees that the Delphi methodology would not support the calculation of standard errors for the new case weights, but said that it would allow FJC to assess the integrity of the resulting case weight system. We do not believe that the proposed methodology can be used to assess the accuracy of weights based in large part on consensus data. The Delphi technique of guided, structured discussions inherently relies for its accuracy and reliability on the experience and knowledge of the participating judges and the accuracy and reliability of judges' recall about the average time required for different events in many different types of cases—about 150 if all case types in the current weights were used. The greater the number of events and types of cases for which judges are asked to make estimates, the greater the demands on judges to recall

accurately the judge time required by those events and types of cases. Generally, the Delphi technique is most appropriate when more precise analytical techniques are not feasible and the issue could benefit from subjective judgments on a collective basis. However, more precise analytical techniques are available and were used to develop the current district court case weights. We believe that any methodology used should support the calculation of standard errors. Such statistical measures are essential for assessing the potential error of the weighted case filings for any specific district that has requested additional judgeship(s).

We believe that the importance and cost of creating new federal judgeships requires the best possible case-related workload data to support the assessment of the need for more judgeships. The methodology approved for the revision of the bankruptcy case weights offers an approach that could be usefully adopted for the revision of the district court case weights. The bankruptcy court methodology would use a two-phased approach. First, new case weights would be developed based on the time data recorded by bankruptcy judges for a period of weeks—a methodology very similar to that used to develop the current bankruptcy case weights. The accuracy of the new case weights could be assessed using standard errors. The second part represents experimental research to determine if it is possible to make future revisions to the weights without conducting a time study. The data from the time study can be used to validate the feasibility of this approach. If the research determines this is possible, the case weights could be updated more frequently with less cost than required by a time study. We believe this methodology would provide (1) more accurate weighted case filings than the design proposed for revising the district court case weights and (2) a sounder method of developing and testing the accuracy of case weights that were developed without a time study.

Objectives, Scope, and Methodology

As agreed with your office, our objectives were to (1) determine whether the methods the Judicial Conference uses to quantitatively measure the case-related workload of district court and court of appeals judges results in a reasonably accurate measure of judges' case-related workload, (2) assess the reasonableness of any proposed methodologies to update the workload measures, and (3) obtain information from the AOUSC on the steps the Judiciary takes to ensure that the case filing data required for these workload measures are accurate. To do this, we obtained and reviewed documentation on the methodology used to develop the existing workload measures and proposals to revise those measures from AOUSC and FJC and interviewed officials at both agencies. We based our assessments on our experience with and knowledge of sound research design and generally accepted statistical analysis methods. We also obtained information on the methods the judiciary uses to ensure the accuracy of the case filings data on which the workload measured rely. Although the Judicial Conference considers a number of factors in assessing judgeship needs for the district courts and courts of appeals, our work focused only on the relative accuracy of the weighted case filings and adjusted case filings measures. We did our work in Washington, D.C., in April and May 2003.

- - - -

We will send copies of this report to interested congressional committees, the Director, Administrative Office of the U.S. Courts; Director, Federal Judicial Center; and the Chair, Committee on Judicial Resources, Judicial Conference of the United States. We will make copies available to others on request. In addition, this report will be available at no charge on GAO's Web site at <http://www.gao.gov>.

If you have any questions about this report, please contact me at (202) 512-8777. The key contributors to this report were David Alexander, Kriti Bhandari, Rochelle Burns, and Chris Moriarity.

Sincerely yours,

A handwritten signature in black ink, reading "William O. Jenkins, Jr." with a large, stylized flourish at the end.

William O. Jenkins, Jr.
Director, Homeland Security and Justice Issues

Enclosures - 4

Enclosure I

Quality Assurance Steps the Judiciary Takes to Ensure the Accuracy of Case Filing Data for Weighted Filings

Whether the district court case weights are a reasonably accurate measure of district judge case-related workload is dependent upon two variables: (1) the accuracy of the case weights themselves and (2) the accuracy of classifying cases filed in district courts by the case type used for the case weights. If case filings are inaccurately identified by case type, then the weights are inaccurately calculated. Because there are fewer categories used in the courts of appeals workload measure, there is greater margin for error. The database for the courts of appeals should accurately identify (1) pro se cases (2) reinstated cases, and (3) all cases not in the first two categories.

All current records related to civil and criminal filings that are reported to the Administrative Office of the U.S. Courts (AOUSC) and used for the district court case weights are generated by the automated case management systems in the district courts. Filings records are generated monthly and transmitted to AOUSC for inclusion in its national database. On a quarterly basis, AOUSC summarizes and compiles the records into published tables, and for given periods these tables serve as the basis for the weighted caseload determinations.

In responses to written questions, AOUSC described numerous steps taken to ensure the accuracy and completeness of the filings data, including the following:

- Built-in, automated quality control edits are done when data are entered electronically at the court level. The edits are intended to ensure that obvious errors are not entered into a local court's database. Examples of the types of errors screened for are the district office in which the case was filed, the U.S. Code title and section of the filing, and the judge code. Most district courts have staff responsible for data quality control.
- A second set of automated quality control edits are used by AOUSC when transferring data from the court level to its national database. These edits screen for missing or invalid codes that are not screened for at the court level, such as dates of case events, the type of proceeding, and the type of case. Records that fail one or more checks are not added to the national database and are returned electronically to the originating court for correction and resubmission.
- Monthly listings of all records added to the national database are sent electronically to the involved courts for verification.
- Courts' monthly and quarterly case filings are monitored regularly to identify and verify significant increases or decreases from the normal monthly or annual totals.
- Tables on case filings are published on the Judiciary's intranet for review by the courts.
- Detailed and extensive statistical reporting guidance is provided to courts for reporting civil and criminal statistics. This guidance includes information on

general reporting requirements, data entry procedures, and data processing and reporting programs.

- Periodic training sessions are conducted for district court staff on measures and techniques associated with data quality control procedures.

AOUSC did not identify any audits to test the accuracy of district court case filings or any other efforts to verify the accuracy of its electronic data by comparing the electronic data to “hard copy” case records for district courts. Within the limited time for our review, AOUSC was unable to obtain information from individual courts to include in its responses. We have no information on how effective the procedures AOUSC described may be in ensuring that the data in the automated databases were accurate and reliable means of assigning weights to district court case filings.

Enclosure II

Measuring Judicial Workload Using the Collection of Time Study Data

The current bankruptcy court and district court workload measures were developed using data collected from time studies. The district court time study took place between 1987 and 1993, and the bankruptcy court time study took place between 1988 and 1989.

Different procedures were used in these two time studies. The bankruptcy court time study protocol is an example of a "*diary*" study, where judges recorded time and activity details for all of their official business over a 10 week period. The district court time study protocol is an example of a "*case-tracking*" study, where a sample of cases were selected, and all judges who worked on a given sample case recorded the amount of time they spent on the case. Time studies, in general, have the substantial benefit of providing quantitative information that can be used to create objective and defensible measures of judicial workload, along with the capability to provide estimates of the uncertainty in the measures.

Estimating Judge Time in Diary and Case-Tracking Studies

At the conclusion of a case-tracking study, total time spent on each sample case closed during the study period is readily available by summing the recorded times spent on the case by each judge who worked on the case. For a given case type, the summed recorded times can be averaged to obtain an estimate of the average judicial time per case for that case type.

For a diary study, however, it is necessary to make estimates of judicial workload for all cases that were not both opened and closed during the data collection period. This estimation step requires information from the caseload database, and thus the accuracy of estimates depends in part on the accuracy of the caseload data. Two kinds of information are required from the caseload database: case type and length of time the case has been open. Using these data and the time data judges have recorded for specific cases, estimates can be made of the overall time required for cases that were not opened and closed during the calendar period covered by the diary study.

Comparing Case-Tracking Studies and Diary Studies

Each study type has advantages and disadvantages. The following outlines the similarities and differences in terms of burden, timeliness of data collection, post-data collection steps, accuracy, and comprehensiveness.

Burden on Participants

Each study type places burden on judicial personnel during data collection. It is not clear that one study type is less burdensome than the other. The diary study procedure requires more concentrated effort, but data are collected for a shorter period of time.

Timeliness of Data Collection

Data collection for a diary study can be completed more quickly than for a case-tracking study.

Post Data Collection Steps

More effort is needed to convert diary study data to judicial workload estimates than case-tracking study data. Also, the accuracy of estimates from diary study data depends in part on the accuracy and objectivity of the information in the caseload database.

Data Accuracy

It is not clear that one study type collects more accurate data than the other study type. Some of the bankruptcy court case-related time study data could not be linked to a specific case type due to misreporting errors and/or errors in the caseload database. Some error of this type likely is unavoidable because of the requirement to record all time rather than record time for specific cases only. However, it is plausible that a diary study collects higher quality data, on average, because all official time is to be recorded during the study period; judicial personnel become accustomed to recording their time. In contrast, the data quality for a case-tracking study could decline over the study's length; for example, after a substantial proportion of the sample cases are closed, judicial personnel could become less accustomed to recording time on the remaining open cases.

Comprehensiveness and Efficiency

In theory, a case-tracking study collects more comprehensive information about judicial effort on a given case than a diary study, because data for a sampled case almost always are collected over the duration of the case. (Data collection may be terminated for a few cases that remain open, or are reopened, many years after initial filing.)

With the diary approach, the total judicial time that is required for lengthy case types is estimated by combining "snap shots" of the time required by such cases of different ages. Thus, in theory, producing accurate weights for lengthy case types is not problematic. In practice, however, difficulties may be encountered. For example, in the 1988-1989 bankruptcy time study, the asset and liability information for cases older than 22 months was inadequate and appropriate adjustments had to be made. In addition, difficulties may arise if only a small number of cases of the lengthy type are in the system. This is an issue FJC said it is considering as it finalizes how to assess the judicial work associated with mega cases in the upcoming bankruptcy case-weighting study.

Enclosure III

**Comments from the Chair of the Judicial Resources Committee,
Judicial Conference of the United States**

**UNITED STATES COURT OF APPEALS
SECOND CIRCUIT**

CHAMBERS OF
DENNIS JACOBS
CIRCUIT JUDGE
UNITED STATES COURTHOUSE
FOLEY SQUARE
NEW YORK, NEW YORK 10007

May 27, 2003

William O. Jenkins, Jr.
Director, Homeland Security and Justice Issues
General Accounting Office
441 G Street, NW
Washington, DC 20548

RE: Draft correspondence on “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)

Dear Mr. Jenkins:

Thank you providing a draft copy of “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” for comment. The time constraint is such that my comments are less comprehensive than I would wish. But I do want to convey a few salient observations.

The General Accounting Office (GAO) finds that the current case weights used in formulating the judiciary’s judgeship requests are “reasonably accurate” and based on a “reasonable” methodology. The focus of the critique is on the techniques for fine-tuning those numbers over time. I have three observations.

1. All of the current judgeship requests pending before Congress are to augment courts in which the workloads far outstrip the threshold standards reflected by the formulas. Thus, although the guideline for district courts is a weighted caseload of 430, all of the requests for additional district judgeships are for courts in which the weighted caseload per judgeship is 491 and higher. Similarly, although the guideline for circuit courts is an adjusted caseload of 500, all of the requests for additional circuit judgeships are for courts in which adjusted filings per panel are 583 and higher. These workloads transcend any deviations that superior fine-tuning could correct.

2. The GAO report does not explain the role of the case weights in the full context of the process through which the Judicial Conference develops its judgeship recommendations. While important to the Judicial Conference's process for evaluating judgeship needs in the district courts and courts of appeals, the weighted caseload in an individual court is only the starting point. A recommendation to create a new judgeship is not strictly formulaic in any sense, and is based on numerous other factors considered after a court has satisfied the guideline of 430 weighted filings per district judgeship or 500 adjusted filings per circuit panel. These factors include: a) careful review of the court's use of senior, visiting, and magistrate judges to assist in handling the caseload; b) consideration of whether the court's workload situation may be transient (and would not therefore justify recommending an additional judgeship even if weighted filings are in excess of the guideline); c) review of the court's use of alternative dispute resolution or other methods to better manage caseload; and d) consideration of a temporary (rather than permanent) judgeship if the caseload level justifying an additional judgeship arose only in the most recent years, or when an additional judgeship would reduce the court's per judgeship or panel caseload close to the guideline.

Moreover, the development of a judgeship recommendation requires several levels of review involving the collective judgment of the individual court, the circuit judicial council, the Subcommittee on Judicial Statistics, the Judicial Resources Committee, and the Judicial Conference. It is a reasoned, conservative approach that results in requests for far fewer judgeships than would be called for by a mechanical application of caseload guidelines. This is true for circuit judgeships as well as district judgeships.

As to the circuit courts, a simpler methodology is used to calculate the adjusted filings. This methodology has served the judiciary's needs for developing the baseline for considering requests for additional appellate judgeships. The judiciary has considered various other methodologies over the last several years, but the workloads in the courts of appeals entail important factors that have defied measurement, including the significant differences in the courts' case processing techniques.

3. I am disappointed that the GAO report does not fully or accurately depict the context of the new district court case weighting study. This study, which involves no changes to the role of weighted filings in the decision-making process, is intended to serve as a bridge from the current set of weights to an updated set that takes into account the legislative and case-management changes occurring in the past ten years. The Subcommittee on Judicial Statistics, which is a subcommittee of the Judicial Resources Committee, does not intend to start from scratch; it will use substantial data already collected and the current case weights will serve as "anchors" in the development of

William O. Jenkins, Jr.
Page 3

updated case weights. Moreover, the report does not adequately describe the sophisticated methodology designed by the Federal Judicial Center (FJC) for collecting input from judges. The Committee was informed by the FJC that the Delphi technique is commonly used and is a scientifically valid way of developing this kind of information from subject matter experts, such as from judges about the work of judges. I am advised that numerous state court systems apply this technique for developing case weights and judgeship formulas. While the GAO observes that other techniques may be even more statistically accurate, the GAO does not consider the substantial increased costs and time delays involved in attaining what we believe is likely to offer little or no added value for the investment.

The Judicial Resources Committee and its Subcommittee on Judicial Statistics take very seriously our responsibilities for proposing effective case weight formulas, for identifying other factors that affect judgeship needs, for assisting courts in weighing and coping with their needs, for scrutinizing every request, and for ensuring that our judgeship requests are based on sound and rigorous requirements.

Sincerely,



Dennis Jacobs
Chair, Committee on Judicial Resources

Comments from the Federal Judicial Center

THE FEDERAL JUDICIAL CENTER
THURGOOD MARSHALL FEDERAL JUDICIARY BUILDING
ONE COLUMBUS CIRCLE, N.E.
WASHINGTON, DC 20002-8003

RUSSELL R. WHEELER
DEPUTY DIRECTOR

TEL: 202-502-4164
FAX: 202-502-4099
email: rwheeler@fjc.gov

May 27, 2003

William O. Jenkins, Jr.
Director, Homeland Security and
Justice Issues
General Accounting Office
Washington, D.C.
jenkinswo@gao.gov

RE: Draft correspondence on "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)

Dear Mr. Jenkins:

This responds, on behalf of the Federal Judicial Center, to your May 22 letter inviting agency comment on the draft correspondence referenced above, prepared by your office at the request of the Chairman of the House Judiciary Subcommittee on Courts, the Internet, and Intellectual Property. We have provided you technical comments separately, as you requested.

The relatively brief time you were able to provide us to review the draft correspondence, although understandable in light of the project schedule, obliges us to limit our comments to pp. 9-11 ("Research Design for Updating District Court Case Weights Raises Concern"). The research design prepared by the Center for the Judicial Resources Committee's Statistics Subcommittee, and approved with some modification by the Subcommittee, differs in several ways from the methods used to calculate the current weights. We anticipate that this design will yield reliable case weights and provide the additional benefits that you describe at p. 9—reduced burden on judges (and, we note, on staff both in chambers and in the clerks offices), potential cost savings, and more rapid development of the weights for use in the judgeship estimation process.

Your description of the research design (p. 10) is basically accurate but, as a brief summary, omits some necessary elements that, we believe, should allay the two concerns that you express about the design.

You note first the challenge that we will face in obtaining necessary data from the different automated systems that the district courts will be using during the study. We recognize this challenge, as you note, and have taken several steps to meet it. One is the formation of the FJC-AO-courts technical advisory group that you describe at p. 10. In

addition, we will survey court operations managers about docketing practices and systems to ensure that procedures for extracting data account for important differences among courts. We have also identified collateral data sources that we will analyze for event information (e.g., JS-10 data and PACER docket sheets) if extractions from individual courts prove problematic.

Second, you are concerned that the design will not require judges to record contemporaneously the time they spend on cases. Case weights developed from the new study will rely on two categories of time expenditure measures: actual and estimated time. We agree that the method used to obtain estimated time expenditures will not permit the calculation of a standard error term around a resulting time estimate and therefore a case weight, and thus, in that sense, the resulting case weights might be considered somewhat less objective than would case weights using data derived from a properly executed diary or case-tracking study.

We disagree, however, that the absence of standard error calculations will preclude us from assessing the integrity of the resulting case weights system, because we will rely on the alternative procedures described below. Specifically, we will use a variation of the Delphi method to obtain estimates of the time required to complete various case activities. The Delphi method, widely used in state courts to develop weighted caseload measures, is an iterative process designed to produce consensus decisions among participating experts. In this study, district judges are the experts and their task is to decide on appropriate estimates of the average time district judges expend on specific case-related events.

The first step in this process will be to convene representative groups of eight to thirteen district judges, meeting within each of the 12 regional circuits (approximately 100 judges in all) to obtain time estimates that represent the cases and practices within the circuit. Each work session will take place over one-and-a half to two days; Center staff will facilitate each session, using a standardized protocol.

The activities that the groups of judges will assess in these individual work sessions will include such events as conducting a conference, conducting a motion hearing, preparing an order on a dispositive motion, preparing an order on a discovery motion, and preparing to begin trial. We will provide the judges with a default value for each event, in order to anchor their initial time judgments. Accurate default values can be expected to enhance the reliability of the judgment process. About half of the default values are derived from actual time-reports that were either collected from judges during the 1993 case weights study or were reported on JS-10 forms.¹ Judges will be encouraged to depart from the

¹ The default values derived from JS-10 reports will be obtained by regression analysis. Independent variables are frequency counts for specific nontrial proceedings and the dependent variable is time expenditure (in half-hour increments). The values so derived represent averages that make no distinction among case types. These default values may represent sound estimates for most case types, but circuit-based judge groups may find them inappropriate for some case types. If the judge groups conclude that a default value does not apply to a given case type, they are likely to depart from the JS-10 defaults, despite the defaults having been derived from objective data.

default values only if their experiences, revealed in the focused interchanges in the circuit meeting, suggest that a departure will improve the accuracy of the estimate. We expect such departures to result in case type-specific event estimates that are both above and below the default value.

Following this series of meetings, a national group of 24 judges (2 from each circuit, some of whom may have participated in the circuit-based meetings) will consider the information provided by the circuits, and using structured procedures, will arrive at consensus estimates for case activities that best represent the national practice.

Although this method does not permit a statistical assessment of standard error for consensus-based estimates, the method can be evaluated on the basis of adherence to the procedures, summarized above, that we will use to gather the data and promote their reliability.

The Federal Judicial Center appreciates the opportunity to comment on the draft correspondence. Please contact me or members of the project team if you need additional information.

Cordially,



Russell Wheeler

(440195)