



Kingdom of the Netherlands

Public Disclosure Authorized

Public Disclosure Authorized

Public Disclosure Authorized

Public Disclosure Authorized



Case-Weighting Analyses as a Tool to Promote Judicial Efficiency: Lessons, Substitutes and Guidance

December 2017

Case-Weighting Analyses as a
Tool to Promote Judicial Efficiency:
Lessons, Substitutes and Guidance

Contents

Acknowledgements.....	1
Acronyms	3
Executive Summary.....	5
1. Introduction	10
2. CWA Basics	11
a. What issues does a CWA address?	11
b. What makes CWA different from other tools?.....	12
c. Typical uses of a CWA	13
d. Principal CWA design options	15
3. Guidance on Selecting between Options for CWA Design.....	17
a. Why consolidate case and event types?.....	17
b. Are event-based case weights preferable to calculating case times directly?.....	19
c. Is timekeeping more expensive and time-consuming than panels?	20
d. Is timekeeping more accurate than panels?.....	22
e. Are there advantages in combining time logs and panels?.....	24
f. Should support staff be included in a CWA?	26
g. For time logs, is sampling more effective than total participation?.....	26
4. Are There Quicker and More Effective Substitutes for a CWA?	28
a. CMIS analysis to improve allocative and technical efficiency	29
b. Using a CMIS to update an earlier CWA	30
c. Advanced benchmarking techniques.....	30
i. Data Envelopment Analysis (DEA).....	30
ii. Stochastic Frontier Analysis (SFA)	32
iii. DEA and SFA limitations	33
d. Can a CMIS be used to calculate case weights?	34
5. Four Non-Methodological Issues for Analysis Design	35
a. Extent of stakeholder engagement in CWA implementation.....	35
b. Effects of systemic inefficiencies on findings	37
c. Reforms' impacts on case weights and recommendations	38
d. Consequences for court users	41
6. Lessons on the Uses of CWA and Alternative Approaches	43
a. For all approaches, a CMIS, preferably automated, is needed	43
b. Conducting CWAs during or before major reforms is unadvisable	43
c. There are design considerations beyond time and cost.....	43
d. Efficiency analyses are not one-time undertakings.....	44
e. Information from a CWA is underutilized.....	44
f. Faster, less costly alternatives are available, but carry some risks also.....	45
g. Concerns about accuracy may be exaggerated	45
h. Understanding technical inefficiency requires additional methods.....	45
i. Neither the CWA, nor its alternatives, address possibly more important problems ...	46
ANNEX 1: GUIDE TO USING CWA AND ALTERNATIVES.....	48
ANNEX 2: ANNOTATED BIBLIOGRAPHY ON THE CWA AND RELATED APPROACHES TO IMPROVING SECTOR EFFICIENCY	50
REFERENCE LIST	62

Acknowledgements

This report examines various approaches to conducting a case weighting analysis, and offers some good practices and several lessons. The report discusses the limitations of the CWA technique and offers some possible substitute approaches. The report aims to help policy makers and international partners to decide whether and when to undertake a CWA or one of several alternatives and to guide them through the various design options.

The report was prepared by Dr Linn Hammergren (Lead Author and Judicial Performance Expert), Ms. Georgia Harley (Senior Governance Specialist) and Ms. Svetozara Petkova (Justice Reform Expert) from the Governance Global Practice of the World Bank.

The team benefited from input from Mr. Adrian Fozzard (Practice Manager) and Dr. Marina Matic Boskovic (Justice Reform Expert) from the World Bank and from Jan Petry, Directorate General for Justice and Consumers, European Commission. The team also thanks Ms. Patricia Carley and Mr. Nenad Milic (Consultants) for their editorial support.

Acronyms

AWOP	Affirmation without [Written] Opinion (U.S.)
CMIS	Case Management Information System
CWA	Case-Weighting Analysis
DEA	Data Envelopment Analysis
EU	European Union
EWMI	East-West Management Institute
IPEA	Institute for Applied Economic Research (Instituto de Pesquisa Econômica Aplicada, Portugal)
IT	Information Technology
NCSC	National Center for State Courts (U.S.)
PRIS	Judicial Information System (CMIS) (Montenegro)
SFA	Stochastic Frontier Analysis
USAID	United States Agency for International Development

Executive Summary

A case-weighting analysis (CWA) is a technique developed in the United States in the 1970s to help courts estimate their personnel needs, adjust staff distribution, and support requests for more human resources. Although often referred to as a case-weighting *study*, the term “analysis” is used here as a more accurate description of the process involved—an analysis of data on court activity. The practice has expanded to other countries in recent years, which in turn has broadened its uses.

This report examines various approaches to conducting a CWA, and offers some good practices and several lessons. The report discusses the limitations of the CWA technique and offers some possible substitute approaches. The report aims to help policy makers and international partners to decide whether and when to undertake a CWA or one of several alternatives and to guide them through the various design options.

CWA BASICS

Issues addressed by a CWA. The CWA was developed to address output insufficiencies created by inadequate staffing, caseload distribution, work units, and human resources relative to the actual demand for judicial services. Recent applications have also attempted to address other problems, such as how to deal with significant variations in individual productivity, a tendency of judges to prioritize easier cases over more complex ones, and the impact of both on accumulating backlog.

What makes a CWA different? The CWA innovation is the recognition that not all cases require the same amount of effort from judges, court staff, or their equivalents in other organizations. This implies that allocating or evaluating personnel according only to input or output numbers is insufficient. The CWA provides a means to define the level of effort invested in handling each case type. Converted to average case weights, the results can be used to determine reasonable caseloads and distribution of staff.

Typical uses of a CWA. As first designed, the CWA was used to document the need for additional staff based on calculations of maximum feasible caseloads. More recent uses include reallocating staff or cases between work units, setting productivity quotas and evaluation standards, and planning the merger or reduction of work units.

Principal CWA design options. The CWA has evolved over the past 40 years, and designers of these analyses now have to make decisions on such details as:

- Whether to use time logs¹ to record real-time inputs or to opt for estimates of the level of effort (provided through judges’ self-evaluation or by expert panels);
- Whether to calculate time for total cases or for events within them;
- For time logs, whether to use manual or online recording of time;
- How the optimal duration of time log exercises shall be determined;
- Whether to require time logs of all judges/units or only of a portion;
- How to select participants if timekeeping is not universal;
- Whether to use time logs alone or in conjunction with an expert panel;
- For expert panels, whether to use real-time estimates or relative values, and;

¹ A time log is a timekeeping exercise in which judges (or other staff) record the time spent on a daily basis on each case or events within it.

- How much to aggregate case and event types.

SELECTING BETWEEN THE VARIOUS OPTIONS

Why consolidate case and event types? For practical reasons (ease of use), both case and event types should be consolidated into a lesser number of categories. A good case management information system (CMIS) can provide guidance, although sometimes there are problems in determining what is counted as a case (or event). The definition of case and event types is a good place to involve working groups and/or panels of experts, both in order to utilize their expertise and to enhance their understanding of the process and of how the CWA will be used.

Are event-based case weights preferable to calculating case times directly? For both time logs and panels, event processing is faster and more accurate now that advanced CMIS can facilitate the calculation of event and disposition frequencies. Not all cases go through all events, and a significant number may be disposed early in the process, thereby reducing the overall average time invested in each case type.

Is timekeeping via time logs more expensive and time-consuming than the provision of estimates by panels? The cost and duration of timekeeping have decreased with the advent of online logs and a shift to event tracking. Moreover, the greatest investment of time and cost for either method occurs in the preparatory and analytic stages. There may be other reasons to choose between the two approaches, but cost and time are no longer the most important.

Is timekeeping via logs more accurate than panels? Accuracy is always relative, and both methods tend to lead to an overestimation of levels of effort. Panels are sometimes used because they are perceived to be slightly more objective, since their members are not accounting for their own time. The choice often comes down to stakeholder preference; stakeholders may have greater trust in time logs and prefer this method because it requires input from a large proportion of judges.

Are there advantages to combining time logs and panels? This is often done in the United States for a follow-up CWA, but typically by using data from a prior CWA in combination with CMIS statistics, panels, and occasionally, more limited timekeeping. The purpose is to reduce the duration and cost of iterative exercises. When conducting a CWA for the first time, logs and panels are sometimes combined to enhance accuracy, increasing both cost and time overall. This is not generally recommended, however, because the inherent complexity can generate confusion and eventually resistance to any recommendations. If panels are added, they are most effective in the design stage and in making adjustments to the time log data.

Should support staff be included in a CWA? Although staff is not usually included in the United States, in other countries the decisions vary. Staff has a considerable effect on level of effort by the lead professionals, but its inclusion can be costly and complicated. Moreover, staff input (and even numbers) often varies within a single country, which may be difficult to capture even with time logs. Some analyses (e.g., Moldova) use a fixed (average) staff-to-judge ratio. This may be sufficient over the short run.

For time logs, is sampling more effective than total participation? Forty years ago, sampling was considered necessary to avoid burdening all judges with time logs. In those jurisdictions where online logs are available and the timekeeping periods are shorter, this argument no longer holds, and some U.S. state courts require that all judges participate. Sampling can still

be done, but sample design is another highly technical undertaking that can cause controversy once the analysis results are in. A more popular, if less scientific, approach in some transitional countries has been to require logs of all judges in a selection of courts.

ARE THERE QUICKER AND MORE EFFECTIVE SUBSTITUTES FOR A CWA?

Analysis of CMIS contents. For rebalancing staff or caseload, a simple analysis of CMIS statistics may be sufficient and completed far more quickly. A CMIS can also be used to set productivity targets and evaluation standards. If required, it can include some simple estimates of case values, but without timekeeping or a full panel exercise. As many have observed, when similar courts are compared, the differences in caseload mix are not that great; however, outputs (technical efficiency) can differ depending on where individual judges direct their efforts.

Using a CMIS to update a CWA. This is a widespread practice in countries that have already carried out one or more full CWA. It cuts time and costs, and absent major changes in caseload size and composition, is considered adequate. In the United States, CWA popularity means that periodic updating is expected, but a full follow-up CWA would be implemented only after a decade or more.

Advanced benchmarking techniques. These approaches were developed for industrial analysis where a “*productivity frontier*” is used to compare the relative efficiency of work units, and resource distribution is planned accordingly. Their application to judicial organizations poses some problems, since they typically do not distinguish between product types (by level of effort or any other means) but rather focus on overall output or revenue. They also are mathematically complex and thus not easily understood by non-experts. When used to compare and reallocate resources between similar types of courts, these shortcomings are less important. However, like all efficiency approaches, they appear to assume that simple resource redistribution will resolve more problems than it can.

A CMIS to calculate case weights. It seems that these systems cannot yet do this, but it is hoped that the capability will become available in time. With it, the reliance on panels and time logs would likely become redundant. However, neither content nor methods currently permit this, and it is not evident that anyone is pushing to develop a means to allow it.

FOUR NON-METHODOLOGICAL ISSUES

Level of stakeholder involvement. Except for academic studies, all approaches use some sort of working group to initiate the process, whether a first-time effort or the latest iteration of a series of periodic reviews. However, although this is an unstated principle in donor-funded efforts, there are signs that it may not be enough, even when complemented by additional training of and outreach to all judicial personnel. This especially affects the post-CWA approval stage and the move to adopt and implement recommendations, when displeasure with the conclusions may lead to criticism of the poorly understood methods and especially the analysis.

Effects of inefficient practices on findings. Both logs and panel estimates are based on efforts invested by actors under existing rules and conditions. Similarly, the alternative calculations of relative efficiency are limited to what the present system permits. In many countries, current practices encourage the inefficient use of time, especially—but not only—in courts with relatively little work. Where the inefficiencies are not acknowledged, the CWA risks legitimizing and hardwiring them for years to come. This can lead to demands for more

judges, prosecutors, or public defenders, when other kinds of change are really needed, such as procedural reforms, standardization of practices, training in court administration, and process simplification and reengineering.

Impacts of reforms on case weights and recommendations. Changes in procedural rules, work force organization, traditional practices, overall workload and its composition, and the behavior and incentives of external actors (such as attorneys, parties to a dispute, or other sector institutions) can significantly alter the amount of work required. Rushing ahead with a CWA when such changes are anticipated or have recently occurred is not wise, as this often results in an incomplete analysis or a recommendation that it be shelved.

Consequences for court users and justice service delivery. Because of their origins, many CWA efforts seem structured to produce recommendations on staff/caseload redistribution. This is one of the more surprising aspects of recent CWA applications: “success” is often defined merely as completion of the analysis and, in some cases, the adoption of recommendations. However, there is rarely any indication of the proposed or real impact of a CWA on court users, or any investigation of whether or how the CWA produced any tangible improvements in justice service delivery.

LESSONS ON THE USES OF CWAs AND ALTERNATIVES

A CMIS, preferably automated, is needed. If one does not exist, a CMIS should be developed before a CWA is initiated. Working with incomplete or paper records is possible, but the compilation of an analyzable database is time-consuming and may require, as in Moldova, additional checks on data accuracy.

Conducting an analysis during or before a major reform effort is not advisable. Reforms change the values derived from a CWA, and for this reason, its recommendations are often shelved. The wiser course is to hold off until the reform has been fully implemented and reviewed for its effectiveness. Alternatively, policy makers who wish to proceed during reform periods would be wise to use less complex procedures, the results of which can be utilized to track reform results or as a prelude to a second analysis, once the time is right.

There are design considerations beyond time and cost. The multiple design options offer ways to cut the cost and time of any CWA method. However, the selection of methods and design options also depends on local preferences, immediate goals, and the desired level of stakeholder engagement in the implementation of the CWA.

CWA are not one-time undertakings. Donors requesting and/or financing a CWA are often in a hurry to do it all at once. Their haste overlooks the lengthier evolution of the approaches in other countries and the likelihood that pending reforms will require early follow-ups. Giving stakeholders more time to engage with the process, simplifying objectives, defining service improvements, and thinking ahead to the next stage are thus advisable.

Information from CWA is underutilized. Given the effort (and funds) invested in identifying the levels of effort required by different case types, it is surprising that the information is often used only to establish average case weights as the basis for staff (or caseload) redistribution. This may be all that is requested, but further analysis of the data (deviations from the mean or average, comparisons with results from other countries) should be considered as a way to understand the causes of problems, which most likely extend beyond insufficient or excessive staffing. Since case weights may quickly be outdated, this is an opportunity lost if not seized at once.

Faster, less costly alternative approaches have drawbacks. First, none of them have been used to track level of effort (case weights) and there is no indication that they can. Where case weights are required, analysts inevitably add information from a CWA (usually using a panel) approach. Second, the alternatives, and especially the benchmarking approaches, usually require experienced consultants and may not be easily understood by local stakeholders. Third, because they are less intuitive, these approaches can suffer credibility problems as can the use of expert panels, especially with relative case weights.

Concerns about accuracy may be exaggerated. Where the results of different methods are dissimilar, there are no independent means of evaluating their relative accuracy. Even if they coincide, accuracy can be questioned. The importance of greater accuracy in estimating level of effort hinges on uses; where these are limited to staff or workload redistribution, simpler methods (even using relative weights) should be adequate.

Understanding technical inefficiency requires additional methods. For all the approaches reviewed here, efficiency is always relative, defined as the optimization of output under present conditions. Over the short run, this is sufficient to correct significant within-system imbalances in inputs (staff and/or caseload). Nevertheless, the assumption of an effective production function (a term used by the benchmarking approaches) is inherently risky and diverts attention from other factors impeding higher productivity and other service improvements.

Neither the CWA nor the alternatives address additional, possibly more important problems. Aside from allocative efficiency or the improved distribution of personnel relative to caseloads, these problems include limited access to justice for citizens and businesses, the lack of standardization of procedures and practices between work units, significant variations in capacity and professionalism among judges (or their equivalents in other sector organizations) and staff, a legacy of bureaucratic and inherently inefficient procedures, political pressures, corruption, and other types of external interference. In most justice institutions, these issues are more of a concern than allocative efficiency. Nonetheless, if these additional issues are not salient or simply cannot be addressed, a review of allocative and relative (within-system) technical efficiency may still be worthwhile, as long as it does not lull anyone into a sense that all else is well or that a CWA can help to solve these other, often more compelling, performance challenges.

1. Introduction

The case-weighting analysis (CWA) is a technique developed in the United States in the 1970s to help courts estimate their personnel needs, readjust staff distribution, and support requests for more human resources. Although often referred to as a case-weighting *study*, the term “analysis” is used here as a more accurate description of the process involved—an analysis of data on court activity. In recent years, the CWA has been adopted in other countries and regions and used for several additional purposes. In transitional² country contexts, donors have often recommended conducting a CWA as a way to increase judicial efficiency, even though not all donors and national policy makers have had a detailed understanding of what this analysis actually can and cannot deliver. This review of the CWA’s value relative to the challenges commonly facing justice systems in both transitional and developing nations was prepared as a response to this knowledge gap.

Based on a survey of the literature and complemented by interviews with participants and direct observation in a number of countries, this review aims to help policy makers and donors decide whether and when to undertake a CWA and to guide them through the various design options. As further elaborated below, a CWA is best at improving allocative efficiency or ensuring that the amount and distribution of human and other resources match the demand for services. Used alone, its ability to advance “technical efficiency” (i.e., getting the most out of the resources in place) is limited. Moreover, it does not address a series of additional problems with potentially greater impacts on the quality of judicial services. Thus, like all reform tools, its effectiveness hinges on applying it to the right problems, as it is not necessarily the best option in every setting or circumstance. If a system is plagued by issues outside the scope of a CWA, its application should be reconsidered or, at the very least, timed and tailored to encourage simultaneous or subsequent attention to potentially more significant concerns.

² The conventional distinction between developing and transitional systems is maintained, with the latter term referring principally to Eastern European nations.

2. CWA Basics

a. What issues does a CWA address?

As judicial caseloads increase and available budgets face limitations, courts and other justice sector institutions have become concerned about how to resolve several related problems that, in conjunction, affect the quality and quantity of their outputs:

- Unequal allocation of workloads between work units and the staff within them
- Territorial distribution of work units and staff out of sync with actual and projected demand
- Possibly excessive or insufficient number of work units and staff because of traditional expectations or legal impediments to readjustments (immovability, tenure, or laws stipulating where courts and other units will be placed³)
- Accumulation of backlogs and longer delays in disposing cases

Two additional problems, rarely mentioned in the United States but possibly important in other countries, are:

- Sizable differences in outputs between judges working in the same or similar courts
- The tendency of personnel to focus on “easier” cases, leaving the more complex ones behind to feed a growing accumulation of “pending” caseload

These issues are only a few among many factors that can negatively impact court output. What they share is their focus on increasing judicial efficiency, largely allocative but to some extent technical, and the perception that any improvements require changes (usually increases) in funding for staff and/or greater flexibility in its use. To advance either remedy, those allocating court budgets (the executive and legislature) and setting other rules must be convinced of this need. This is where the CWA comes in, as a means of documenting the problems, helping courts resolve them on their own, and with luck, convincing the budgetary authorities that more staff or fewer constraints on their distribution would produce better results.

As sector institutions develop better case management information system (CMIS) methods, some of these problems can be addressed through an analysis of the data the CMIS provides. However, 40 years ago, first-generation systems were not ready for the challenge. Thus, for its proponents, a CWA was and often remains the mechanism of choice for addressing the problems. CWA efforts are time-consuming and, depending on their design, costly, but advocates argue that they are the best means of determining how work is done and where more (or less) staff is needed to do it. Although first developed for and most often applied to courts, the CWA method can also be used for prosecutors and public defenders (lawyers

³ Expectations and the resulting political pressures are universal; removing a court from a small district can face political opposition from local politicians and/or their constituents. The weight of legal impediments varies by region. In Latin America, tenure and immovability have less of an impact than in Eastern Europe, but laws stipulating work unit placement are more frequent, with Colombia as a prominent case, guaranteeing a judge in every *municipio* (county).

assigned to defend indigent clients),⁴ and even occasionally police investigators, as these organizations and organizational actors often share the same problems.

Box 1. The Difference between Case Duration and Time Invested in its Resolution

It bears emphasizing that the time actors put into processing a case has little or no relationship to how long it remains in the system or to any legally defined deadlines. A case that takes six months or six years from filing to disposition may only require a few hours of work from a judge, her assistants, and other support staff.

One reason for the enormous difference is that courts and other sector institutions process multiple cases simultaneously, giving each one only a part of their attention. The other explanations concern the role of the parties, who need time to file additional documents and motions, review evidence offered by opposing parties, and prepare their own evidence and arguments. Where hearings are held, scheduling problems or difficulties in convening all parties may contribute to delays; parties may also add their own dilatory maneuvers to “buy time.”

b. What makes CWA different from other tools?

The introduction of CWA represented a great step forward in recognizing that not all cases place the same work burden on those charged with handling them. This fact highlights the inadequacy of calculating staffing needs based only on the number of cases filed and/or disposed. Essentially, a CWA differentiates between cases according to the level of effort required to handle each type. In its simplest form, it looks only at the time invested by the major professionals (judge, prosecutor, or defender) involved in resolving the types of cases covered. A CWA may also track input by support staff and/or do its analysis by tracking the time devoted to each step involved in processing each case. A classic CWA uses time logs⁵ to register time invested in each case type and/or event and then weights the cases accordingly. Recent variations or, according to some authors, alternatives use estimates by expert panels (usually applying the Delphi method) or some combination of the two approaches.

Box 2: The Delphi Method

The Delphi method (with its name inspired by the Oracle of Delphi) was first developed by the Rand Corporation on assignment by the United States Air Force in the early 1950s. It was designed as a technique aimed to obtain the most reliable consensus of opinion from a group of experts by replacing direct confrontation and debate with a series of sequential individual interviews usually conducted by questionnaires. Respondents are

⁴ See APRI (2002) and Gramckow (2011) for prosecutors; Pace et al. (2011) and Washington State (2014) for public defenders. It is important to note here that while case weighting may be appropriate for the organizational needs of Public Defense Offices in the United States, the setup of many legal aid systems (e.g. in Europe) is based on completely different principles and case weighting might not be appropriate or needed.

⁵ A time log is a timekeeping exercise in which judges (or other staff) record the time spent on a daily basis on each case or events within it.

asked to give reasons for their opinions and the interviews are usually interspersed with feedback derived from other respondents or from computed consensus from previous stages of the process. The technique entails avoiding face-to-face debate and providing anonymity of opinion and of arguments advanced in defense of those opinions. This approach is intended to overcome the deficiencies of the classical brainstorming sessions or roundtable discussion formats, where the outcome may be unduly influenced by psychological factors such as the presence of a dominant, persuasive personality, approval-seeking or unwillingness to change one's opinions once they have been expressed publicly.⁶

If the purpose of the method is the estimation of a numerical value (e.g. the purpose of the first Delphi exercise was to estimate the number of atomic bombs that a Soviet strategic planner would use against an optimal U.S. industrial target system), then it is quite probable that the final responses of experts would not fully coincide and justifiable corrections would need to be applied to the ultimate answer.⁷

For best results, the Delphi method shall be administered by a skilled facilitator whose individual approach and experience can greatly impact the results.

c. Typical uses of a CWA

Once a CWA has been completed, its subsequent uses vary. In the United States, a CWA's principal use has been to document needs for additional staff and thus to argue for budgetary increases with the federal or state legislature and executive. Case weights, based on average times invested by each actor, are used to calculate the maximum workload each type of professional can handle and therefore the number of professionals (or staff) required to process a given number of cases.⁸ U.S. analyses rarely include non-judicial staff, as the budgetary battles focus on the numbers of key professionals. However, both in the United States and elsewhere, observers note the importance of support staff and the types of work they do in enhancing the productivity of judges, prosecutors, and public defenders. This factor may be still more critical in developing and transitional countries, where the work and quality of support staff are less standardized. In the United States, a CWA occasionally feeds into other projects, such as changes in jurisdictions, merger of work units, and even staff reductions.⁹ Although the CWA process calculates maximum workloads by case type for individual actors, in the United States, the results are not used to evaluate judges' performance.¹⁰

⁶ See Bernice B. Brown (1968)

⁷ See Dalkey and Helmer (1963).

⁸ The terms "caseload" and "workload" are used interchangeably here, but frequently each term (not necessarily the same one) is used to designate either incoming cases or the total of incoming and pending cases. Time is also allocated to administrative tasks, holidays, vacations, and sick leave so that the time available to each staff member for processing cases is calculated as a proportion (usually roughly 70–80 percent) of the annual working days.

⁹ Where the number of judges is reduced, this is typically through attrition, not immediate cut-backs.

¹⁰ This is because the U.S. system of non-career and/or elected judges does not accommodate formal evaluations. Informal assessments by outside groups do occur but usually focus on issues other than quantity of output. Other U.S. organizations may use case-weight-derived workloads for evaluations of career staff.

The adoption of the CWA approach has gradually expanded, although not to developing regions, where cultural factors, inadequate technology, and a host of higher-priority issues may explain its absence.¹¹ It has had more success in other developed common law nations and, still more recently, in the European civil law countries (Langbroek and Kleiman 2016). The spread of the use of CWA to transitional states in Eastern Europe follows its endorsement by the European Union (EU) for candidate countries and its financing by several donors. Throughout continental Europe, the intended uses are quite broad, including not only for promoting plans to transfer cases or staff from one work unit to another, but also for setting targets for individual or work unit productivity, occasionally linked to salaries or budgets;¹² justifying reductions as well as increases in staffing levels and work units; creating incentives for reducing backlogs; factoring case weights into the information technology (IT) system for random assignment of cases between individual judges within a court; evaluating judges' performance; and so on.

Many of these individual purposes could be met through simpler mechanisms, including analysis of statistics from a more sophisticated CMIS. Production targets can be calculated from the average outputs of similar courts, and unequal distributions of caseloads or resources of similar types of work units are evident even in simple statistical systems. However, when calculations must include the level of effort required, the CWA remains the mechanism of choice, what Flango and Ostrom (1996) call the "gold standard." Its inherent complexity lends an image of technical and scientific superiority while the greater opportunity (relative to proposed alternatives) for stakeholder participation in its design and data collection promote user ownership. These characteristics can be useful in combating intra-institutional resistance to reforms. Such resistance is most likely when applications extend to setting productivity (or production) quotas or evaluation criteria or simply to reducing staff numbers in some jurisdictions.

¹¹ Langbroek and Kleiman (2016) do cite a case from the West Bank, suggesting that donors working in the Middle East may promote CWA adoption by financing these efforts. Gramckow (2011) also cites a study done in Mongolia in 2003. The CWA's absence in Africa may be explained by technological shortcomings (and thus relatively little data on caseloads), a generalized shortage of sector staff (making talk of reallocation a pipedream), and more concern with issues such as judicial integrity and independence. Latin America is a puzzle despite the region's adoption of increasingly sophisticated CMIS structures. Lack of donor interest and financing is one explanation; resistance from sector institutions to measurement of any type is another. Measurement resistance is less frequent in countries of the former Soviet Bloc, where judicial output had traditionally been used to monitor performance (Solomon 2012).

¹² A few Western European countries link salaries to weighted productivity (unsuccessfully in Spain, whose judges rebelled, but successfully in Austria). In the Netherlands, court budgets are linked to anticipated caseloads and dispositions. Although Holland was a forerunner in calculating processing times, its budgetary formula uses case weights indirectly to estimate costs per case.

Box 3: A Short Note on Productivity Targets or Quotas

Productivity quotas or targets, one potential CWA product, are increasingly common management tools and just as increasingly controversial.

The first issue raised by such targets is conceptual and forms part of the classical dichotomy between judicial accountability and judicial independence. While it has long been recognized that judicial accountability is an indispensable element of the checks and balances inherent to every democratic system, the extent and forms of that accountability have continuously been under scrutiny lest they encroach on judicial independence. This tension has taken on a new life with the advent of the concept of the so-called managerial accountability whose methods aim to promote efficiency and cost control and link results to resources.¹³ Productivity targets are one such method and should therefore be used with caution. Productivity targets can be particularly sensitive when introduced by the executive or by the legislature.¹⁴ Similarly, even when introduced by judicial councils, these productivity measurements can spur criticism, especially if individual judges' results in meeting such quotas automatically result in sanctions or affect their salaries.¹⁵

Another concern related to productivity targets is their potential encouragement of "cherry picking," or the tendency of those subject to the quotas or targets to focus on easier cases to increase their output levels. A CWA can discourage this tactic as long as its differential case weights are incorporated into the target setting.

A third concern, where a CWA is of less help, is that a quota or target may come to represent a ceiling, beyond which no further improvement is required. Should this appear to be occurring, the solution is to continue raising the target, treat it as a minimum requirement, or provide recognition for those who exceed it.

In short, using a CWA to develop productivity quotas should be done with caution, taking into account the various shortcomings of this tool.

d. Principal CWA design options

Some 40 years after its introduction, the CWA has developed numerous variations, presented here as options. Although some authors separate a timekeeping-based CWA from an analysis using panel estimates to calculate case weights, this is not universal and thus both are treated here. This is partly because of disagreements on the separation between the two methods; it is also because recent analyses often combine the two. The most recent views and findings on the relative merits of the approaches (only time logs, expert panels, or both) are outlined below.

¹³ See Contini and Mohr (2007).

¹⁴ Thus, commencing from 1995 the Ministry of Justice of Finland has collaborated with courts to introduce a "management by objectives" system by setting productivity indicators. Even though these indicators were not used automatically, but solely as guidance in the allocation of resources, some representatives of the judiciary have criticized the approach as contrary to judicial independence. See Contini and Mohr (2007).

¹⁵ For example, the Spanish Judicial Council has been under strong criticism for tying judges "output measures" to their remuneration. See Contini and Mohr (2007).

Much of the existing literature on the CWA (see Annex 2 for an annotated bibliography) details how the evaluation was applied in each court or national system. The resulting recipes vary, especially as involves:

- Use of actual time logs completed by judges, prosecutors, and/or their assistants or estimates of level of effort provided by expert panels. As discussed below, some early conclusions as to the costs and accuracy of each approach may no longer hold.
- Whether time logs or expert panel estimates focus directly on total time to process a case or begin with the time taken for specified events within each case. An event-based calculation is now more common; in the end, it translates to average weights (or time invested) by case type, but as elaborated below, its results may differ from a direct focus on cases.
- For time logs, manual or online recording of time. Recent variants increasingly use the latter to facilitate the process and avoid errors made in transferring manual logs into an electronic database that will be used for analysis.
- Where logs are used, for how long a period. Duration varies in the articles referenced from four weeks to six months or more. Event-based timekeeping seems to require less time, though this may be a function of its incorporation of CMIS data and online timekeeping to accelerate the process.
- Whether time logs are required of all judges/units or only of a portion.
- If timekeeping is not universal, how participants are chosen. Is there an effort to construct a representative or random sample or are there other criteria for selection?
- Whether time logs are used alone or in conjunction with an expert panel to make further adjustments to the results.
- For expert panels, use of real-time estimates or relative values. In the latter, experts are asked to assign relative weights to cases/events rather than provide real-time estimates for an entire case or steps within it.
- Level of disaggregation of case types. Both for panels and for real-time logs, respondents are rarely asked to treat all case types individually, but rather to work with grouped types. The number of groupings can range from a few dozen to over a hundred. This may add a second level of disaggregation based on substantive or procedural “complexity.”
- Disaggregation of events covered. Some systems divide events into three or four categories (for criminal cases, pre-charge, pre-trial, and trial periods), whereas others list them separately. To facilitate recording, a lesser number of categories may list typical events within them but not require that these events be noted separately.

A few descriptions of real analyses do discuss the alternatives considered and the reasons for the choices eventually made. However, among those reviewed, none question the design choices already made, as a means of developing a cost-benefit analysis that might help future CWA efforts. Generally, funds (estimated costs and available budgets) and time available had the most impact on the choices made. Analyses using both time logs and expert panels often detect minimal differences in the outcomes, but this finding is not universal and does not appear to have influenced subsequent CWA projects, either in the same country or in other nations.

3. Guidance on Selecting between Options for CWA Design

Designing a CWA now requires a number of choices that were not present four decades ago. Moreover, some initial assumptions about the relative strengths and weaknesses of different options may no longer hold. Discussion here addresses the principal changes or challenges to earlier views. Annex 1 provides a table that summarizes and compares the observations outlined below and in Section 4.

It bears mentioning that the major obstacle to assessing the options is the virtual absence of studies comparing their results, either in terms of differences in their findings and recommendations or ultimately, their impact on service quality. It is considered that event tracking is faster and provides different and seemingly more accurate weights than logging total case times. A few studies suggest that the results of timekeeping and of panels do not differ much, but there are examples that indicate the contrary.¹⁶ Otherwise, the conclusions that can be reached are largely limited to relative cost, duration of the process, extent of stakeholder engagement, and credibility issues. If, as seems likely, CWAs continue to be used, it would be worth doing some comparative studies. Stakeholder preferences are often the deciding factor, but stakeholders could make their choices more wisely if the outcomes of each of the various options were assessed systematically.

a. Why consolidate case and event types?

The only CWA reviewed that incorporated all case types (284, including some that occurred only once in the five-year period covered) was done by the Rand Corporation (Pace et al. 2011) for the U.S. Federal Office of Defender Services. However, its ability to do so hinged on the availability of timekeeping data on all cases handled by the network of defenders working in the federal system. Timekeeping for public defenders is a common practice, despite some doubts about its accuracy.¹⁷ The absence of this practice among prosecutors, judges, and support staff makes it advisable to consolidate an often vast number of case types into a simpler set of case categories for all CWA variations.

A well-designed CMIS may already provide a structure for categorizing case types, using a sort of decision tree, and with some further simplification (to combine less frequent case types considered comparable in level of effort), can serve as the basis for any of the case-weighting alternatives. Still, despite the increasing sophistication of CMIS design, observers have noted a series of problems with potential impacts on both timekeeping and panel exercises. Several issues involve what even a well-designed CMIS does not include, such as data (e.g., number of complaints, parties, or witnesses) required for case complexity assessments. Others are more basic, concerning how cases are defined and counted.

¹⁶ In both Serbia and Montenegro, the panels and logs provided significantly different estimates of level of effort. Although Serbia used the time log estimates, in Montenegro, the panel estimates prevailed. Serbia's choice was not explained in the report; in Montenegro, panel estimates were preferred, as they incorporated initial levels of complexity.

¹⁷ Skepticism about "billable hours" runs rampant in the United States, although less so with regard to over-loaded public defenders. The issue has been raised for court-appointed attorneys in some other countries, but its treatment goes beyond the scope of the present document.

One such problem is mentioned by the studies in Romania and Moldova (EWMI 2011, 2013; Hriptievski, Gribincea, and Wittrup 2014) and by observers in other countries within and outside Eastern Europe. It is a tendency for many transitional and developing countries to register some events as case types within their CMIS, thereby substantially inflating the number of incoming and disposed cases.¹⁸ Another issue raised by Langbroek and Kleiman (2016) is the tendency in some countries to count multiple charges against a single defendant or multiple defendants as separate cases. This is another example of caseload inflation, which complicates time log registries and leads to overestimating levels of effort. Since the authors do not specify the countries where this occurs, its frequency and potential or actual remedy are not addressed.¹⁹

These practices should be corrected, preferably before a CWA is attempted. Where case inflation is used consistently across the jurisdiction, it can be controlled and need not impact the CWA findings. If the CMIS counts the annulment of a misdemeanor as separate from a misdemeanor case, as long as the latter does not include annulments or the practice is not universal within the country, any of the case-weighting methods or alternatives can work with it. However, experience shows that case inflation is often used inconsistently by different courts, judges, and staff. In that case, it is virtually impossible for the CWA to truly compare cases and workloads. The tools needed to address that problem may include guidance and training to standardize case categorization and data entry, supported by data integrity audits and managerial oversight.

The selection of events for logs and panel estimates follows the same logic and presents few additional issues. As with case types, it is practical to condense events into a smaller number of categories. This is usually done by dividing case processing into a few stages, with the events within each stage clearly defined so the respondent knows where to count them. A problem can be posed when a CMIS does not include disposition types, which even for judges can make a difference. In the United States, in part because many CMIS schemes track the times involved, distinctions can be made between bench and jury trials, as the latter require more time from all actors.

Weighting the efforts of prosecutors and public defenders poses some additional problems. Prosecutors, for example, often have a variety of means to terminate cases. These include mediation, temporary or permanent dismissal, deferred prosecution, plea bargaining, and so on, all of which typically require less time than a full-blown trial. It also bears noting that both prosecutors and public defenders conduct a good deal of their activities outside of their offices and outside of the courts. All of this should also be counted when logging or estimating the level of effort required for each case type.

¹⁸ This is often less a problem of CMIS classifications than of data entry, although CMIS design (a long list of alternative categories) and inadequate training in and vetting of data entry exacerbate the problem.

¹⁹ The phenomenon and a potential solution have been documented in Bulgaria (Bulgaria 2017). For example, where Bulgarian investigative authorities file requests with the court to authorize access to electronic traffic data on a single user from several mobile phone operators, some courts log this as one case while others register a separate case for each operator.

b. Are event-based case weights preferable to calculating case times directly?

Event-based logs and estimates are currently so widely accepted that this question hardly merits discussion. It is addressed for the benefit of anyone considering a return to tracking total case times directly. The adoption of event-based methods has several explanations. First is the ability to use improved CMIS techniques to calculate event frequency by and across case types and to provide times for such events as initial hearings and trials (where it can be assumed that the length of the event represents the time invested by all participants). As noted in Florida Legislature (1998) (see box 5 below), both cost and time of CWA were increased by the need to calculate frequencies manually. This should no longer be necessary in judiciaries using a sophisticated CMIS.

Second, experience shows that cases of the same type have different trajectories depending on how and when they are disposed, meaning that any effort to go directly to average times can be misleading. The type of disposition may be as important as when it occurs. A study (Doerner, Douglas, and Tallarico 2010) on Oregon's Appellate Court noted, for example, that an affirmation without opinion (AWOP) required far less effort (time) than an authored opinion, although both disposition types occurred at the same stage in the process. Moldova's recent waiver of an obligatory written opinion for first instance civil cases should likewise shorten judges' input except in cases where the parties request one. As another example, in several Latin American countries (e.g., Guatemala), a prosecutor's decision to dismiss a case or defer prosecution sometimes requires judicial approval. Where a judge must rule, the prosecutor's level of effort (and time spent in hearings) is increased compared to what is required when only his/her immediate superior decides.

Third, tracking cases from filing to final disposition requires more time to capture those that go through the entire trajectory rather than some form of early disposition (e.g., early dismissal, parties' decision to terminate the claim, or fast track disposal as in the use of plea bargaining in the United States).

Box 4. An Example of Calculation of Case Weights Based on Events

Event-based calculations are now used in most countries, but APRI (2002) provides the most concise example of their outcomes. In the end, event-based data translate into average times by case type. However, the average is calculated by considering the frequency and average level of effort required for the events tracked, as well as the frequency of different disposition types. APRI makes this logic especially clear. Its calculation of the feasible caseload for U.S. state prosecutors for each type of case tracked assigns different weights according to when in the process the disposition occurs. Using three case stages (pre-charge, pre-trial, and trial), it calculates the average time taken by the events in each stage as well as the portion of cases disposed there. It then uses the resulting partial values to calculate the total time required by a prosecutor to dispose the number of cases covered. Finally, the results are compared to the prosecutorial time available to determine the number of cases one prosecutor can work in a year. APRI's illustrative example thus concludes that a prosecutor can handle 19 homicide cases

annually, if a majority are disposed in the pre-charge and pre-trial phases, with only one going to full trial.

Similar calculations can be done for judges and public defenders as well as for their staff. Given the lack of detail on the direct total-time-per-case calculations (e.g., Ketterle 2013 on Germany), it is not clear whether or how they deal with early dispositions. The same may be true of event-based estimates, which, unless truncated trajectories are considered, may overcount the average times required. Presumably, as it becomes possible to use a CMIS and other organizational data to track disposition types and frequency of events, this issue is in the process of being resolved.

c. Is timekeeping more expensive and time-consuming than panels?

Several reports note that a decision to use relative values or time estimates provided by an expert panel hinged on the anticipated greater expense and duration of a log-based timekeeping CWA. Although the differences may initially have been significant, the situation has changed over the past two decades. Many of the changes involve the evolution of court automation (and especially CMIS techniques), timekeeping requirements already in place (especially for public defenders and court-appointed attorneys), and the development of software to allow online timekeeping and facilitate data analysis. Especially if an event-based process is used, judges or other actors may only be asked to log their activities over a few weeks and can enter data directly online. However, even in countries without these IT innovations, the greatest investment of time, if not of funds (software can be expensive), for a CWA occurs in the preparatory period. This is when working groups presiding over the process are organized and meet, case and event types are selected, decisions are taken on who will be surveyed or logged, questionnaires and time logs are developed, and training for participants is designed and implemented.

Further costs and time are required for the subsequent analysis, review, and possible modification of the results. If additional questionnaires or expert panels are appended to a time-logging exercise, these also increase costs and add to the overall time required. As a result, the notion that timekeeping is inherently more costly and time-consuming than panels may be somewhat exaggerated. This is because for all methods, much of the cost and time invested occur before and after the collection of the basic data. Even as estimated for the Florida state courts in 1998 (see box 5), the differences detected were largely related to the preparation and analysis of data collection instruments for timekeeping exercises. Moreover, many additional costs would be incurred by manual tasks that can be eliminated with a more advanced CMIS.

Box 5. An Illustrative Comparison and Costing Out of the Four Major CWA Alternatives

In preparation for conducting a new CWA in 1998, the Florida state legislature and the courts' administrative office prepared a briefing paper on the four principal alternatives, their input requirements, and their budgets (Florida Legislature 1998). The costs (in 1998 U.S. dollars) are outdated, but this is the sole example found that documented the differences. A timekeeping CWA, either by case type or event, was judged most expensive at a maximum cost of US\$267,300 and US\$344,275, respectively. The use of Delphi panels, either for case or event estimates, came to US\$145,000 and US\$221,975, respectively. Interestingly, the study's authors concluded that Delphi estimates were no less reliable than those derived through timekeeping. Unfortunately, the reasons for this conclusion are not included in the briefing paper.

Since the budgets were not completely disaggregated, the difference of more than US\$120,000 between the timekeeping and Delphi versions is not entirely explained. However, much of it derives from costs associated with supervising the timekeeping exercises and the manual collection of additional data. The total time required ranged from eight months for a Delphi case-based analysis to 13–14 months for an event-based timekeeping procedure. In the latter case, much of the extra time and cost were for hiring and training support staff and collecting and analyzing data from the time logs. In both event-based variations, expenditures were increased by the need for manual sampling of case files to determine event frequency, something usually now available from a CMIS. The design of time logs and software added costs to timekeeping. If they can be obtained from analyses done elsewhere (with minor modifications), those costs can be reduced.

In terms of opportunity costs, a classic timekeeping CWA is more onerous because it requires that judges and staff spend time filling in logs rather than doing their substantive work. Still, if timekeeping exercises are kept short, this effect can be reduced. Most analyses done in the United States, based on event times, are conducted over four–six weeks. They often include all judges rather than a sample, but there was no indication that judges resented the imposition. Efforts to log times directly by case types (as in Germany and apparently in Switzerland (Ketterle 2013; Leinhard et al. 2015) usually take longer, as they must track cases from beginning to final disposition. Apparently for this reason, the initial German analysis focused only on a few case types, with later iterations including additional variants (Ketterle 2013).

U.S. analyses never mention an issue that is sometimes raised in transitional countries, which is that timekeeping must extend over longer periods because case input fluctuates from month to month. Fluctuating input is a given everywhere, but courts' stock of pending cases guarantees sufficient work to keep them busy even when inputs temporarily fall. If, in a period of high inputs, staff places more attention on the initial processing of new cases, there is no reason to believe this will alter the overall time invested in processing the new or existing stock. Only in countries where extended official judicial vacations remove most staff for a month would one want to avoid those periods. However, this is not for lack of work but rather because there would be fewer people remaining to fill out the logs.

d. Is timekeeping more accurate than panels?

Although a timekeeping CWA may add costs and time, if not as much as initially believed, there is still the question of whether its presumed greater accuracy (i.e., the “gold standard”) is sufficient to justify the additional expenses as well as the burden on respondents’ time. As mentioned above, several analyses found that values produced by logs and expert panels were not that different and were thus equally useful for calculating resource needs. Still, there are exceptions, especially when conducting a CWA for the first time. Unfortunately, identified differences or similarities have not led to any concrete conclusions; both methods have been criticized for potential biases.

Despite its gold standard status, timekeeping does not escape criticism. First, it includes less efficient practices (or courts) and second, it is highly unlikely that anyone filling out a log will leave any time slot blank. Hence, time logs tend to overestimate the time needed to process an event or case, although time log design (see box 6) may also encourage downward estimates. Expert panels may also overestimate times because they too reflect current practices and tend to be skewed towards a focus on outlier cases.²⁰ However, panel members do not face the additional challenge of having to account for eight hours of their own work and thus the temptation to exaggerate inputs.

Box 6. Effects of Time Log Design on Accuracy

Accuracy can be affected by the ways time logs are designed. Where respondents must fill in the activities corresponding to specific time slots (for example, in 15-minute intervals), they are more likely to be “inventive.” However, when they are asked to assign times worked to specific cases, they may under-report efforts unless reminded that the sum should cover a full day’s work. Although this appears to be the approach taken in Montenegro (Grubišin 2015), it did not seem to have that effect there. However, it may have affected the 2004 German evaluation (Ketterle 2013), if in a different way. Since judges were asked to log time spent on a specific set of cases, this may have led to an under-accounting of other cases and work, leading to the complaint that the case weights calculated added up to more time than judges had. Where expert panels are asked to estimate rather than simply adjust logged times, there are similar and potentially more complex challenges. Panel members are usually asked to estimate times spent on cases or events, but not in the context of an entire working day. It then may take some complex calculations to adjust their estimates to the time available.

Although most first-time CWA efforts include time logs, Bulgaria appears to have been satisfied with questionnaires and panels. This is an interesting example in that the U.S.-based National Center for State Courts (NCSC), which specializes in time logs, provided external

²⁰ This tendency has been noted for other expert estimates, including for the World Bank Group’s *Doing Business*. As is often remarked, even “experts” tend to remember the one case that took 17 years rather than the 17 cases disposed in under 12 months. Although apparently made before time logs were done, the USAID comment refers to Serbia, where panel estimates were considerably higher and thus discarded for the most part in favor of the logs.

assistance. Local observers did report that some judges were not happy with the results (not that unusual for any first effort), and methodological purists may object to the survey techniques.²¹ However, according to Kalpakchiev (2016), a member of the Supreme Judicial Council, the choice had high buy-in from the local working group and has been adopted for reporting and analyzing judicial workloads.

Box 7. Bulgaria's CWA without Time Logs

In terms of coverage, this was a large project, but data collection was done very quickly and cost relatively little because it relied on a questionnaire sent out to 2,000 judges (and answered by 1,100), using a later series of expert panels to review and adjust the results. The questionnaires divided each of approximately 300 case types into separate events/tasks and then asked judges to assign a value to each of these events based on the last case of the respective type that they had examined. As Bulgaria still lacks a unified CMIS, this is an interesting example for countries in the same situation.

Drawing on this input, a methodology for case weighting was developed and approved by a judicial taskforce in 2013 that was piloted in 32 courts and applied nationwide in 2014. The results were assessed and approved by focus groups of judges. A later (2015) evaluation by NCSC experts was reported to be positive.

The resulting Judges' Case-Weighting Rules entered into force in April 2016 as the basis for reporting and analyzing judicial workloads. Much of this is included in a centralized information system developed for this purpose. Kalpakchiev (2016) notes that it will take a year to assess and analyze the results, and that for "more resolute actions," such as the closure or merging of courts, still more analysis and time will be needed. Currently the system measures case weights only for reporting purposes.

The first rigorous assessment of the use of the CWA from its launch until the end of 2016 was approved by Bulgaria's Supreme Judicial Council in September 2017 (Bulgaria 2017). Based on experience to date, it recommended some adjustments to the system. It is expected that after assessing the operation of the CWA for the whole of 2017 and introducing the recommended adjustments, the judiciary will be able to commence using the results for allocating resources between different courts, and cases between the judges within each court.

Sources: Langbroek and Kleiman (2016); Kalpakchiev (2016).

The issue in the end is less about relative accuracy and more about stakeholder preferences. These preferences can be powerful and are usually not worth combating, especially since there is no means of evaluating differences between estimates generated by panels, timekeeping, or other methods, or where their results are similar, determining how accurate they are. Much hinges on the intended use. Where it is limited to optimizing staff distribution

²¹ Clearly a response rate of only 60 percent for the 2,000 judges suggests respondent bias, but sometimes methodological purity can be waived in favor of getting some initial results. The real test is whether performance improved, and for that, see the section "Four Non-Methodological Issues for Analysis Design."

or caseloads between similar work units, simpler calculations using relative weights may be adequate. The method could start with timekeeping data, later adjusted by panels, questionnaires as in Bulgaria, or only panel estimates as in Moldova and Romania, based on levels of complexity.²² As attention turns to evaluating outputs, more accurate measures of level of effort become important, but this could be left for a second stage or type of analysis.

e. Are there advantages in combining time logs and panels?

There is a growing tendency to combine both methods. In the United States, this is usually done after an initial time log CWA as a way of cutting costs and accelerating later iterations. As described in Lombard and Krafka (2005), the mixed methodology uses the results of an earlier time log exercise, updating it with CMIS data and expert panels. In transitional countries (e.g., Serbia, Montenegro), the combination of both methods in a first-time analysis often leads to longer and/or more complex processes. This can open the door to a rejection of results that few understand in their entirety (see box 8 on Serbia and box 11 on Montenegro).

Although not always the initial choice, in countries first attempting a CWA, a timekeeping exercise may have greater credibility than panels or the alternative approaches discussed below. This is especially important if used not only to reallocate human resources but also for more controversial activities, such as setting production quotas, cutting staff, or providing input to personnel evaluation systems. It could also be more convincing to those approving budgets (in the executive or legislature) who, like judges, may have greater faith in time log accuracy. Hence, a time log based CWA, preferably starting with events rather than total case times, may be a good first step in estimating workforce needs. It can be organized to be less costly, including in judicial time invested, by:

- Holding the timekeeping periods to a reasonable four–six weeks
- Using online logs
- Incorporating available CMIS data on case and event frequency as well as trial times
- If expert panels are added, using them to define case and event types for the logging process and making final readjustments rather than requesting they provide their own time estimates

In its first effort at a CWA, Serbia followed none of these recommendations, but after an initial effort based on expert panels, decided to add a time log exercise *as a source of harder data*. This contrasts with other analyses, such as that in Montenegro (box 11), which began with time logs and added expert panels midway through the process. Neither of the CWA endeavors had its recommendations fully adopted. Serbia’s data analysis appears to have also been overtaken by events (including another court reorganization, the reinstatement of dismissed judges, the privatization of some judicial functions, and the entry into effect of a new criminal procedures code). There is now a second effort in Serbia, this time based on another mixed method: CMIS data and analysis of a sample of case files.

²² In both countries, the process started by grouping cases, but then rather than asking panels to provide time estimates, it focused on a division between three levels of complexity. “Weights” were assigned by multiplying the average time for all cases (time available divided by number of cases disposed) by a complexity factor. See Hriptievski, Gribincea, and Wittrup (2014) for a more detailed explanation.

Box 8. Serbia's First CWA with a Rolling Design

Serbia's 2010–12 CWA had the objective of develop[ing] a weighted caseload methodology that can be used to calculate the number of judges needed for each court. It began with a 12-person working group that also served as an expert panel. The group defined case types, differentiated first by substantive matter and then by complexity. Using this framework, it estimated processing times, calculated the time available to judges for case work, and made adjustments for exogenous factors like travel. The approach was event-based from the start, and to cover this element, additional experts (judges) were added to the working group. As in Montenegro, cases from some substantive areas (e.g. bankruptcy) were presumed to be complex. Others were divided into three complexity levels.

Following several months of meetings, the working group decided to add a validation analysis based on timekeeping by 386 judges in 37 courts (roughly 15 percent of the judges but 50 percent of the courts, owing to a prior reorganization that condensed court numbers from 168 to 64).²³ The entire evaluation took over two years, including a final application of the methodology to five courts to test staffing recommendations. However, the report and recommendations were never approved by the High Judicial Council for the reasons cited above.

Sources: Mircic (2012); Langbroek and Kleiman (2016).

It would be presumptuous to draw many conclusions from the limited examples (many U.S. analyses, plus several Eastern European cases). Still, it appears that a first-time effort may achieve better results by using methodological simplicity, whether it relies on timekeeping or expert panels and surveys. Neither method is exempt from challenges to its accuracy. Although a mixed methodology is often introduced to combat this flaw, it adds inevitable complications and still further controversies over the conclusions, which may make its results less comprehensible to non-participating stakeholders and thus reduce the likelihood of their acceptance. CWAs in the United States remain far less ambitious in their objectives. In Western Europe, it has taken time to add the full complement of potential CWA applications and even then, resistance has forced reconsideration of some recommendations. Despite the frequent desire of stakeholders (including donors) for rapid completion and immediate perfection, neither goal is realistic for countries just beginning the process. Even if the recommendations of an initial analysis are not accepted or implemented, introducing the concepts and analytic methods may be an important accomplishment on which later efforts can build.

²³ A subsequent reorganization increased the number of courts, but neither reorganization affected the number of judges. What did affect it was a vetting process in 2009, in which roughly 800 judges were not retained and were replaced by new appointees. The dismissed judges protested, resulting in the reinstatement (still ongoing) of many of them.

f. Should support staff be included in a CWA?

The CWAs reviewed vary on this point. In recent U.S. analyses, staff are rarely included; in other countries, they increasingly are. Obviously, excluding staff makes for a less complicated, less costly, and possibly shorter process. Three early U.S. examples incorporating staff as described by Flango and Ostrom (1996) make this abundantly clear. Pace et al. (2011) in fact advises against CWA methods for assessing staff input.

Where it can be assumed that staffing patterns (numbers, categories, and tasks) are always similar, limiting timekeeping or estimates only to judges or equivalent professionals in other organizations probably makes sense. However, in many developing and transitional systems, this is demonstrably not the case. Under conditions of variable staff numbers and input, time invested by the lead professionals will vary correspondingly. Factoring in this added complication may be too much in an organization that is conducting a CWA for the first time but at some point, it will be necessary. In the meantime, analyses that do include staff typically ignore these potential differences and their effects, reaching conclusions on staff reallocation by using average level of effort (Montenegro) or judge-to-staff ratios (Moldova). However, even they usually miss the role of “irregular” staff, which also makes a difference in the level of effort required of other workers.

Box 9. The Forgotten Workers: Interns, Volunteers, and Irregular Contracted Employees

Note should be made of the use of interns and “volunteers,” many of them unpaid, in both Eastern Europe and Latin America.²⁴ In some courts and even prosecutors’ offices, they do a good deal of work, which should be counted. However, none of the studies reviewed (including, for example, Langbroek and Kleiman (2016) or Gramckow (2011), which covered several Eastern European analyses) made any mention of their presence. Only World Bank (2014), which is not a CWA, mentions their role in Serbia. Any CWA of whatever type in countries where this practice is followed is clearly incomplete, if it does not take them into account. If one wants to know how much effort is expended in processing cases, the contributions of these individuals (as well as those of a substantial number of “contracted” personnel performing a variety of tasks) must be included.

g. For time logs, is sampling more effective than total participation?

Earlier treatments (e.g., Flango and Ostrom 1996) assume that when dealing with a large judiciary (e.g., several hundred or a few thousand judges), sampling is needed to hold down costs and reduce the burden on all judges (or other actors). The need for sampling seems less obvious today, given the shorter times required and use of data already available (e.g., event, case, and disposition frequency from a CMIS and recording of trial times) in more recent examples. However, as several authors caution, in larger organizations or federal systems, or where court sizes and other circumstances vary considerably, users of both samples and total

²⁴ The presence of both groups has two explanations. Practical work, possibly in a judicial office, is often required for graduation with a law degree. Still more commonly, people take or keep (once the required practical work is completed) these positions in the hopes that they will provide a way into a permanent paying job. Contracted personnel are often hired with discretionary budgets for whatever work needs to be done. However, this is less frequently related to case processing.

participation should beware of the impact of these differences. Samples can be structured and analyzed to reduce the influence of larger courts (more judges/cases, hence greater weight of their scores in the final calculations). A full participation timekeeping exercise may also require adjustments by analysts to avoid similar distortions of results.

Although there are indications that larger courts may be more efficient (i.e. receive and process more cases per judge), any legitimate reasons for the lower productivity of small courts, and especially those covering larger areas, should be considered in calculating final case weights. This is one reason for using expert panels to reassess the results of a timekeeping exercise. Moreover, several studies (Ketterle 2013 on German courts; APRI 2002 on prosecutors; Pace et al. 2011 on defense) caution that local court practices (especially but not exclusively in federal systems) can affect processing times for judges, prosecutors, and defenders. Some of these practices can be considered less efficient, but that is a matter to be addressed by other means. In the meantime, actors in these systems can do little to overcome these obstacles or their reflection in the time logs register. As Pace et al. (2011) conclude, the phenomenon argues against efforts to immediately impose uniform workload standards across an entire system.

It is often argued that a well-designed sample is superior to a “census” or inventory,²⁵ which is what full participation attempts. While doubtless true, sampling techniques are complex and may meet resistance from local actors not versed in their intricacies. A frequent compromise solution requires timekeeping by all judges (or other actors) in a smaller number of work units, selected to be representative of the various situations in the organization under evaluation. While not as “scientifically” constructed as a simple or stratified sample or as inclusive as across-the-board participation, this may be most acceptable to organizational leaders. As long as the weights are in turn adjusted to avoid introducing biases from larger districts, the approach is sufficient to produce a first-generation CWA.

²⁵ This is largely because a census or inventory inevitably misses something, and often in a way to bias results.

4. Are There Quicker and More Effective Substitutes for a CWA?

One presumed advantage of a CWA is its potential to resolve the series of problems listed in section 2a above. However, where the intended uses are less extensive, there are simpler, less costly, and quicker alternatives available that can produce comparable results. Moreover, as suggested above, there can be advantages to treating more limited goals sequentially, thereby reducing confusion and resistance created by pursuing multiple objectives simultaneously (for a summary of this section, see Annex 1).

Most alternatives can address both allocative and technical efficiency,²⁶ though analysts typically use one or the other. Those that track both often use comparisons of technical efficiency (e.g., output per judge) only to determine which courts have too few or too many staff, recommending personnel reallocations on this basis. Most do not differentiate between case types; those that do utilize values from a separate timekeeping or panel exercise. The methods discussed below generally produce their results more quickly and at far lower costs than any of the CWA variations. When they do not, it is usually because of data problems that would slow a CWA as well. Whether they are equally effective depends on the objectives pursued and whether case weights are considered essential to reaching them. So far, no alternative replaces a CWA in calculating the level of effort accorded to different case types.

Box 10. The Limited Role of CWAs in Solving Judicial Management Issues

Although this is a review of approaches to calculating and using case weights, section 4 covers techniques that, for the most part, ignore them. This raises a question about the importance of case weights. The issue is not their value in the abstract but as applied to solving real judicial management issues. It is hard to overstate the significance of knowing how much time judicial actors invest in processing different types of cases. However, in many organizations that are conducting a CWA for the first time, the use of this information is relatively limited, often only to staff reallocation. So far, its use for projects like those in Western Europe—court mergers, productivity quotas, and staff cutbacks—is minimal.

Given the numerous questions about accuracy, exaggerated panel estimates or logged times, and the changes likely to result from pending reforms, a faster technique (one of those discussed below) might serve as well. Whatever the quality of the CWA recommendations, case weighting is not always able to achieve its stated goals. Even the log-based German CWA faced complaints that the calculated times for judges' caseloads exceeded the time available. The Montenegro recommendations, although supported by both logs and panel estimates, were challenged. In Moldova, on the other hand, analysts skipped both logs and panel estimates, using only relative effort levels for their analysis, and had the results accepted.

²⁶ The terms come from DEA or Data Envelopment Analysis. Allocative efficiency refers to optimizing the distribution of resources relative to existing demand, whereas technical efficiency focuses on minimizing the inputs (staff, budget, and so on) required to produce a given level of outputs (cases decided). In the simplest terms, allocative efficiency compares staffing and caseload; technical efficiency compares staffing and output.

The optimal use of all the information generated by a CWA is a long-term process unlikely to be realized in a novel undertaking that takes only a year or two. Whatever method is used, if its findings and recommendations are not too revolutionary (e.g., Moldova), they may be accepted. Their basis in a thorough, probably log-based, case-weighting evaluation can increase the chances of implementation. Then it remains to be seen whether this results in better services, still another issue meriting further exploration.

a. CMIS analysis to improve allocative and technical efficiency

The CWA was initially intended to replace the use of basic statistics (cases in, cases out, and human resources available) for estimating personnel needs and planning the redistribution of staff (or, where staff cannot be moved, cases). CWA proponents argue, quite correctly, that cases vary by the level of effort required for their processing. Hence, simply counting numbers of new entries or dispositions misrepresents real work levels, since it can encourage “gaming” (actors focusing on easy cases to appear more productive) and in this way, distort assessments of over- or understaffing. The objections, however, overlook the potential offered by a second- or third-generation CMIS, as well as a few basic observations about caseloads of similar work units.

First, a more sophisticated CMIS commonly records considerable information on cases, starting with the case type and going through procedural steps to type of disposition. This means analysts are no longer limited to a simple “head count.” They can enrich their comparisons of case outputs or inputs by making simple distinctions between case and disposition types without relying on the usual CWA methods.

Second, as argued by several analysts, case weights may be less necessary because courts or other work units of the same type (as defined by instance and any further specialization) usually receive similar mixes of the same types of cases. Obviously, it would make little sense to compare the caseload of an appellate criminal tribunal with that of a first instance civil court. However, comparing work units within each category provides a good idea of existing imbalances in demand and resources. In short, if the objective is to improve allocative efficiency or equalize workload, CMIS statistics provide a quick and inexpensive solution.

CMIS data can also be used to improve technical efficiency by determining “reasonable outputs” (i.e. disposition numbers or rates), setting quotas or targets, and providing evaluation standards. The quick and dirty method used in several court systems is to compare outputs between similar units, calculate the average, and use either that number or something slightly above it as the goal. To avoid the gaming problem, a few value judgments on case complexity may be required, but these can be factored into some relatively simple calculations. Expert panels could also be used to assign a simplified set of relative case weights (and so combat the inevitable objections), but without the need for time logs. A further possibility, addressed below, is that case statistics could themselves be used to calculate level of effort.

b. Using a CMIS to update an earlier CWA

Once a CWA exists, further updating is usually recommended, though the frequency varies. Florida Legislature (1998) determined that every four years was required (though it was not specified whether by law or on the basis of further analysis). As has been argued by several participants in a variety of CWA efforts (e.g., McMillan and Temin 2011; Lombard and Krafka 2005), updating may be accomplished by using CMIS data, complemented by only partial timekeeping or Delphi exercises to cover special conditions. Such conditions might include the influx of new types of cases, changes in procedures eliminating or reducing the frequency of some events, or, as appears to be the trend in Eastern Europe, the transfer of non-litigious cases and enforcement proceedings to administrative or private actors. Except where such changes radically reduce workloads, they should not alter the amount of effort expended in the cases or events not directly touched. Hence, to save time and funds, organizations that have conducted one complete CWA might consider subsequent updates based on their CMIS contents and do a full update only after a decade or more has passed. Of course, if the changes alter entire procedural codes, an earlier update may be needed.

c. Advanced benchmarking techniques

Like simpler CMIS analyses (whose data they also use), benchmarking techniques developed in other sectors have been proposed to short-circuit the CWA process in resolving input or output disequilibria. The approach most often mentioned is Data Envelopment Analysis (DEA).²⁷ An alternative, Stochastic Frontier Analysis (SFA), is less frequently used. Debates about their relative worth for measuring technical efficiency in other sectors have been ongoing since the 1970s (Jarzebowski 2013). DEA's popularity among judicial analysts may be more a consequence of their familiarity with the technique than of any intrinsic superiority over SFA and similar alternatives.

i. Data Envelopment Analysis (DEA)

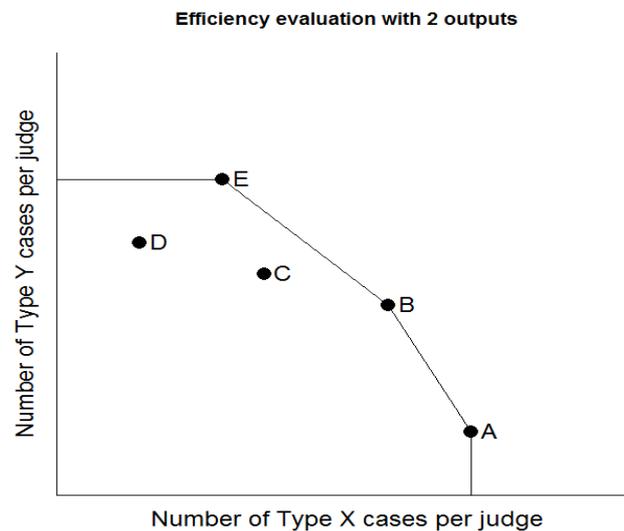
DEA has been used by academics to explore/compare the technical efficiency of courts within a single system (Nissi and Rapposelli 2010; Yeung and Furquim de Azevedo 2011). It is apparently replacing regression analysis (Rosales-López 2008) as the preferred method. In applying either approach, the academics usually ignore the basic logic of a CWA (not all cases require the same work) and evaluate relative efficiency simply by comparing some combination of output (dispositions), input (cases received), and resources (the staffing numbers) in each court.²⁸ Because they compare courts of the same type (instance and occasionally specialization) within a single country, their justification is the similar caseload composition at this level. This may work for allocative efficiency, but it does overlook judges' ability to select the cases they will dispose in a less uniform fashion. Although efficiency is usually defined as output per judge, staff input may be analyzed or included as a potential explanatory variable.

²⁷ See Hriptievschi, Gribincea, and Wittrup (2014), 116 for an illustrative list of studies.

²⁸ The examples tend to be binary comparisons—output with input or output or input against staffing numbers. Presumably DEA can do more complex multivariate comparisons, but its proponents tend to stick to the simpler ones.

Like other advanced benchmarking techniques, DEA uses a “*productivity frontier*,” defining an optimal output-to-input relationship for the existing production function and assigning productivity or efficiency coefficients to individual work units on this basis. DEA’s established methodology and the availability of specialized software for its application give it an advantage over a simple CMIS analysis, although both work with the same statistics and ratios. As compared to a CMIS analysis, the productivity frontier may tell us a little more about relative efficiency, but its coefficients and graphical representations facilitate the identification of outliers, above or below the curve. The fact that the graphical representations are more readily understood by non-experts than regression analysis’ coefficients and linear graphs may explain DEA’s greater popularity for project-related work.

The graph below (from EWMI 2013, Paper 1C, an annex drafted by Wittrup) thus compares the technical efficiency (output) of five courts handling two types of cases as a first step in recommending changes to improve resource allocation. This example assumes the same staffing patterns in each court for “simplicity’s sake.”²⁹



Only courts A, B, and E are on the “efficiency frontier”; D and C are below it. Although the graph compares technical efficiency, its proposed use is to improve staff distribution. Wittrup’s solution moves judges from less to more efficient courts, thereby reallocating resources to enhance the efficiency of both. Unfortunately, he does not explain why he predicts this outcome.³⁰ Although Wittrup suggests this is possible, the example does not assign the cases different weights, looking only at the absolute numbers processed. The EWMI study did develop relative case weights for a separate CWA effort, but Wittrup argues that in the DEA analysis, they are not easily compared across case types—a complex criminal case may have a different value or weight than a complex civil matter.

²⁹ Except for the potential role of non-judge staff, it is not clear why this is relevant, as the comparisons are output per judge and thus should not be affected by personnel numbers. However, it does facilitate the next step, the proposed reallocation of staff.

³⁰ More efficient courts may increase their outputs with additional judges, but a less efficient court might well produce less if a judge is taken from it. As opposed to Wittrup’s argument, the result might be the same efficiency level (output per judge) in each court or even a decline in one or both.

Wittrup's suggestion, without the latter caveat, was used in the Moldova analysis (Hriptievski, Gribincea, and Wittrup 2014) in which he also participated. Here, DEA was used to help courts define staffing needs and reallocate personnel, incorporating a simplified set of relative case weights³¹ as defined with a panel approach. The analysis also considers other staff, but only by deriving a judge-to-staff ratio based on current practices. The ratio was then used to recommend changes in staff numbers against what had been determined through DEA as the optimal number of judges. No questions were raised about the existing ratio, which was simply adopted as a given. The consultants added a regression model to verify data used in the DEA by predicting caseloads against various demographic and socioeconomic data. This was deemed necessary because of doubts about the accuracy of Moldova's caseload data and even the statistics on staffing.

As is evident from the project-related analyses and the academic studies, the application of DEA to courts so far utilizes very simple variables to explore a limited number of issues: optimal staffing compared to caseloads and relative efficiency of courts compared to each other. Proponents have suggested the potential for much finer analysis that might, among other objectives, explore different mixes of staff and/or cases. However, academic researchers show little interest in these questions. In the project-connected analyses in Moldova and Romania, both data problems and the requested focus only on allocative efficiency did not allow time for these pursuits, nor, even in the reports, an explanation of how they might be conducted.

ii. Stochastic Frontier Analysis (SFA)

SFA was not considered in Romania and Moldova because of data limitations, although the authors noted its utility for more complex analysis. This is evident in Castro (2011), a study lying between the academic and project related, as it was commissioned by the Brazilian government but not attached to any specific ongoing or proposed reform. Castro uses SFA to compare (numerically and graphically) efficiency ratios between and within Brazil's first instance state courts (in 26 states and the federal district), as defined by decisions on the merit produced per judge. He also compared clearance, congestion, and backlog rates and tested a series of hypotheses about factors affecting the differences. Like the academic DEA applications reviewed above, Castro ignores differences in case complexity, arguing that Brazil's first instance state courts have a fairly homogenous caseload mix. His selection of SFA is based on the size of the database (over 8,000 court units and roughly 10–12 million filings), which he argues makes it more appropriate than DEA. Constructing the database out of monthly reports, theoretically forthcoming from all judges,³² was a major undertaking, but it allowed Castro to move beyond previous analyses (e.g., Yeung and Furquim de Azevedo 2011) that only compared state averages.

³¹ Relative weights, also recommended by Wittrup, do not use real-time estimates, instead ranking cases by level of complexity, typically into three groups—simple, complex, and very complex—and then assigning weights to each group. In Moldova, this was done by calculating the time for an “average case” (total time taken divided by the number of cases) and then using complexity coefficients to determine the average time for each category.

³² It is apparent that reporting was not complete and was furthermore complicated by multiple reports from judges working in several units during the same month.

iii. DEA and SFA limitations

DEA and SFA have not been used to develop case weights, nor has anyone suggested that they can. Their proponents often critique case weighting, citing frequent changes due to ongoing reforms and controversies among those evaluating times. Instead, they highlight the greater flexibility of not being tied to fixed case weights (see Wittrup in Hriptievschi, Gribincea, and Wittrup 2014, 111–12). Analysts wishing to distinguish between cases can, as in Moldova and Romania, add a simplified CWA methodology—usually panels and relative rather than absolute weights. However, the approaches can cut to the chase in other areas—in Castro’s study, demonstrating the enormous differences in performance as measured by several indicators between similar courts across states and within each one. Their disadvantage lies in their mathematically complex methodology, even though the actual calculations can be done with computer software. Both approaches are less intuitive than suggested by Wittrup’s simple graphs, and entrusting the work and the resulting recommendations to a group of external experts may simply not sit well with stakeholders.

There are questions as to whether a method developed for analyzing efficiency in industries is appropriate for comparing courts. A first problem is the difficulty of calculating output values, the upper part (numerator) of the efficiency ratios. (The lower part is easier than in industrial applications, as staff comprises the courts’ sole factor of production,³³ although in courts, staff can operate in varying capacities, something perhaps less common in the industries to which benchmarking is usually applied.) The industrial applications use product prices, revenue, or profitability to compare outputs, but this obviously will not work for courts. Assuming equal values for all cases simply returns to the initial problem targeted by the CWA: the fact that, whether measured as level of effort or in some other way, differences between case types affect production levels. A judge who only processes cases dealing with standard traffic accidents, for example, can dispose many more cases than a judge handling complex construction cases.

A second issue is that as applied to industries, benchmarking techniques typically aim at increasing overall efficiency, not at equalizing judicial workloads. Workload equalization has been both an objective and a justifying principle for introducing the CWA into new countries. How recommendations on benchmarking may affect this goal in courts is unknown due to a lack of comparative data. Given the “black box” nature of benchmarking calculations, any difference that does exist may not be noticed but could affect applications to evaluations and quotas. Overall efficiency has been overridden in some court applications, most notably in Moldova, where the consultants followed the principle of not recommending staffing cuts, but only redistribution.

The lesser attention to workload equalization can be explained by the assumption in benchmarking exercises of a uniform production function that “*illustrates available and effectively used manufacturing techniques*” (Jarzebowski 2013, 172). Like product (or case) value, this can be assumed away by arguing that all courts of the same category use the same procedures and organizational structures (and thus have a uniform production function). At some level of generalization, this is indeed true, but it does not explain why courts with similar resources (inputs) often receive such diverse efficiency ratings. Wittrup’s solution—moving

³³ In effect there are other factors, for example, computerization and the presence of additional specialized services, but none of the project-related examples consider them. The Serbia Functional Review (World Bank 2014) notes several practices introduced by individual courts that increased both production and productivity.

judges from less to more efficient courts—might increase overall efficiency, but only if the sole determinants are staffing (something that seems highly unlikely) and work unit size (where the recommendations on court mergers are still less frequently followed).

A final issue is that as applied to courts, benchmarking techniques have not generated additional solutions for the efficiency issues. The few exceptions rarely derive their suggestions from their analysis but rather from hypotheses regarding what the analysis does not cover (e.g. management). If the methods can, as some proponents suggest, be adapted to explore the finer details of input variables, they could move the argument ahead. Counting staff is easy; constructing variables that capture their quality, what they do, and how this varies between courts and national systems is very difficult, but these are factors with potentially greater impacts on efficiency. They are also more susceptible to judicial control. For example, Rosales-López's (2008) regression analysis of Spanish courts found that certain support services (in addition to normal staff) had a positive effect, while judicial turnover was negatively associated with efficiency. Significantly, follow-ups to Costa's study (done by the same institute, Institute for Applied Economic Research (or Instituto de Pesquisa Econômica Aplicada (IPEA) in Portuguese) use approaches other than SFA to examine the results of specific reforms and organizational factors explaining differences in technical efficiency.

d. Can a CMIS be used to calculate case weights?

This is everyone's dream: using a complex CMIS to register time spent by judges and staff on each stage of a proceeding and so eliminating both time logs and panel estimates. The hope is not only the quicker, less costly delivery of results but also the avoidance of some of the biases and thus inaccuracies associated with all CWA variations.

The good news is that software is evolving rapidly and that if there is interest in this topic, someone may soon develop something. The bad news is that so far there seems to be little interest; those using the CWA in the United States or elsewhere appear content with the approach as is. There are doubtless some elements already present in a good CMIS, for example, duration of hearings to account for time spent by judges and any staff present. There is also software to monitor time spent on computers and on what task, but this would work largely for staff and has to be designed for that purpose. However, unless a judge spends her whole day in hearings or an assistant works only on the computer, none of these techniques will capture their entire level of effort.

5. Four Non-Methodological Issues for Analysis Design

The issues raised here are neither technical nor methodological. Nevertheless, they affect all approaches, even those alternatives that ignore case weights. They are infrequently, if ever, mentioned in the general or case-specific studies despite their potential consequences on the analytic work and the adoption of their recommendations.

a. Extent of stakeholder engagement in CWA implementation

Although rarely explicitly discussed, stakeholder involvement in CWA implementation is important in ensuring an analysis is understood and its recommendations adopted. Stakeholders often add complications, ranging from an overly tight time frame (see box 11 on Montenegro) to a rolling design that lengthens the process (see box 8 on Serbia). Nonetheless, the complications are less important than another issue: whether the usual participatory mechanisms are sufficient to sustain a commitment to using the analyses' conclusions and recommendations.

Except for the academic studies, all examples reviewed use some sort of working group to initiate the process, whether a first-time effort or, as in the U.S. cases, the latest in a series of iterations. However, though an unstated principle in donor-funded efforts, there are signs that this may not be enough, even when complemented by additional training for and outreach to all judicial personnel. This especially affects the post-CWA approval stage and the move to implement recommendations. Langbroek and Kleiman's (2016) review of six Balkan countries and a few others in Eastern Europe suggests that even when the reports and recommendations are approved, many are still waiting for the other shoe (implementation) to drop. Since the examples are relatively recent, it is conceivable that the results are still forthcoming.

This situation is common to both lengthy analyses (e.g. Romania and Serbia) and those taking as little as a year (see Montenegro, box 11), thereby undercutting the seemingly obvious remedy of simply allowing more time. What may be most helpful is longer preparation combined with less ambitious objectives and techniques, and a realization, as seems to occur in some developed countries, that this is an iterative endeavor best advanced gradually. When, as in a few recent cases, consultants at some stage rush ahead with their analysis, stakeholders surprised by some results may in the end reject them all. The Montenegro analysis, though admirably conducted in the required one-year time frame, is a case in point. Here, the priority to finish within the short time frame seems to have precluded the more extensive consultations that may have led to an acceptance of the methodology and recommendations. However, even in longer-lasting ventures (e.g. Romania), it appears that the technical experts got so far ahead of the stakeholder learning curve that their results were either not understood or largely ignored. Buy-in cannot be forced; stakeholders need time to consider both the techniques and the consequences adequately.

Box 11. Montenegro's Fast-Track, Complex Undertaking

The Montenegro analysis involved both timekeeping (for nine of the 26 courts and nearly half of the 254 judges) and an overlapping expert panel. The process took one year,³⁴ although initial preparation by the working group and further discussions by the government added a few months. Within that year, six months were devoted to timekeeping (done via online logs with an additional budget to develop software) and five overlapping months to expert panels. The analysis used three methods for disaggregating case categories: the traditional classification or PRIS (the Montenegrin CMIS); a slightly more detailed "CWA" categorization based on initial working group discussions; and a three-level case-complexity division developed and applied by the expert panel. The process was event based, and the same events were used for the three variations. The panel's complexity categorization used two models: one based on subject matter and applied to many case types, and the other based on process characteristics, such as number of parties, number of claims, and number of witnesses and types of evidence.

In the end, the shorter list of PRIS categories, with adjustments by the expert panels, was used to calculate the required levels of effort and then develop recommendations on staffing patterns, including judges, advisors, and administrative personnel. Combining panels and time logs is no longer unusual, but several characteristics of the process stand out: the unusually short time frame for completing all the steps, the nonetheless lengthy timekeeping portion, the use of two case categorization systems for the time logs,³⁵ calculations of staffing needs based on two scenarios (only incoming cases and entire caseloads, including pending),³⁶ and the role of expert panels not just to make adjustments but to add a case complexity component after the logs were done. The CWA results were used to prepare criteria for determining the complexity of cases and a methodology for framework standards to determine the required number of judges. However, implementation of these has since been delayed, pending the development of a new case management system in the longer term. As a result, as of the date of this printing, the Montenegrin authorities have yet to take decisions on these aspects.

Source: Grubišin (2015).

Detailed working group discussions usually generate enthusiasm among their members. However, a dozen or so participants are unlikely to sway several hundred or several thousand judges, a Ministry of Justice, a Judicial Council, and a Supreme Court unless, as in Moldova, the recommendations (more staff for most courts) represent gains for nearly everyone. To

³⁴ The time seems unusually short for what was done, and the analysis might have benefited from a longer period, especially to allow the expert panels to operate on the basis of the time log data. This was not possible, as they began to function months before the timekeeping portion was completed.

³⁵ Although the lengthy reports do not explain this, it appears that the time log respondents worked with the CWA categories, which were later condensed into the PRIS system. Many categories were the same, but the CWA system expanded a few of them and thus offered more choices.

³⁶ One issue posed by the second scenario is that if adopted and successful in eliminating the backlog, the courts would have ended up with an excess of staff against future needs. The reason for the second scenario was never explained in the report, and it is unclear what the study authors proposed that the judiciary do with it. The second scenario did soften the reductions (which extended even to judges), calculated when only incoming cases were considered.

complicate the situation in Montenegro, it appears that because of time constraints, the final analysis was not adequately vetted with the relevant stakeholders. The issues under discussion are complex, threatening, and potentially divisive; as such, they are unlikely to produce unanimous acceptance among judges anywhere. As is common in other sectors and even in other justice operations, donors (and stakeholders) promoting the CWA need to take this into account and so develop better means for encouraging wider involvement and understanding.³⁷

b. Effects of systemic inefficiencies on findings

This is principally an issue for the CWA. However, it also affects alternative methods even when not using case weights, because their analysis employs within-system definitions of efficiency—that is, what is possible under current procedures. Both logs and panel estimates are based on efforts invested by actors under existing rules and conditions. In many countries, some current practices encourage the inefficient use of time, especially, but not only, in courts with relatively little work. Some examples below illustrate the problem.

- Where programming of hearings is not well organized, all actors arrive for hearings only to find they will be canceled and reprogrammed. That is time invested, but not productively.
- Where judges do not exercise control over the proceedings, hearings may go on for much longer than is needed to cover the relevant issues. Under new criminal procedures codes, both judges and parties frequently require/request additional witnesses and evidence beyond what may be objectively necessary. Again, this is time spent, but not productively.
- Several Latin American countries use a series of prosecutors, sequentially, to handle a single case in the first instance. One may be present at the arraignment, another handles the investigation, and still a third appears during later hearings and the trial. This “horizontal organization”³⁸ is widely regarded as inefficient, as the data available on prosecutors’ output (cases disposed and especially through indictment and trial) serve to demonstrate.
- In some East European (and Latin American) countries, the transition to prosecutorial investigation often leaves an instructional judge in place to oversee the prosecutors’ work, grant arrest and search warrants, and decide on pre-trial detention. Because these judges are rarely fully occupied with these tasks, their participation in a timekeeping exercise can distort case weights.

³⁷ Two examples, having nothing to do with a CWA, come from Ethiopia and Malaysia. In Ethiopia, the Canadian International Development Agency (CIDA) team spent over a year discussing perceived performance problems with the Federal Supreme Court before introducing a case tracking system to help reduce delays (World Bank 2010). In Malaysia, a backlog reduction program, successfully designed and implemented by the Supreme Court, built on several unsuccessful efforts by the Court’s members (World Bank 2011). Sometimes failure is the mother of success.

³⁸ Horizontal organization (different prosecutors for each stage of a case) is contrasted with vertical organization, where the same prosecutor handles a case from filing of complaint to at least first instance disposition and often through any appeals. Latin American use of horizontal organization is apparently a holdover from the former inquisitorial system, in which instructional judges handled the pre-trial investigation of a criminal case.

- In Eastern Europe (and many developing nations) where judges still handle many non-litigious cases (i.e., validation of documents or agreements reached by the parties, uncontested debt collection), a sizable percentage of their time may be devoted to such quasi-administrative activities. The time invested is partly a result of the sheer number of such cases, but it also seems to stem from a tendency to give excessive attention (and effort) to handling these simple administrative matters. Where non-litigious cases must be decided quickly, this can exacerbate delays for litigious ones, but whether or not this is the case, the disproportionate division of effort means that fewer litigious cases can be resolved (or get the attention they merit) with the existing human resource base.³⁹

Under all these conditions, a CWA may still be justified, if only to provide a baseline figure. Some authors have suggested that the problem can be diminished, if not eliminated, by using data only from more productive units or relying on panels of expert judges. This will discriminate against work units with legitimate reasons for processing their cases more slowly. Still, that can be taken into account in the adjustment period. For example, if prosecutors, public defenders, and some judges must travel to a series of courts to tend to their cases, this can be (and has been) considered when case weights are used to reallocate or evaluate personnel.

The more fundamental concern is whether sector authorities recognize that their organizations are inefficient and thus that case weights, whether based on time logs or estimates, are doubtless excessive. Similarly, the alternative calculations of technical efficiency are limited to what the existing system (or production function) permits. This is a place to start and may also be reflected, if only temporarily, in personnel evaluations. However, where the inefficiencies are not acknowledged, the CWA risks legitimizing and hardwiring them for years to come. This can lead to demands for more judges, prosecutors, or defenders, when what is really needed are other kinds of change, such as procedural reform and process simplification and reengineering. This seems to be a particular problem in Eastern Europe, where for historical reasons, judge-to-population ratios are often unusually high. When comparable data are available, the “necessary” time to process specific cases or events compares very unfavorably to that in countries outside the region, and especially in much of Western Europe. Had CWA efforts been conducted in Latin America, the comparison would likely be worse, despite a far lower judge-to-population ratio.

c. Reforms’ impacts on case weights and recommendations

Changes in procedural rules, work force organization, traditional practices, overall workload and its composition, and the behavior and incentives of external actors (e.g., attorneys, parties to a dispute, or other sector institutions) can significantly alter the amount of work required. Rushing ahead with a CWA when such changes are anticipated or have occurred recently may not be a good idea. To give a few examples:

³⁹ In countries where time log data were incorporated into the reports (e.g., Montenegro and Serbia), the time dedicated to “non-litigious” cases seems excessive. However, without examining the issues covered, it is impossible to substantiate this impression. With similar cases in other countries, it is often found that judges do put in this time, but for work that seems to benefit no one.

- Where countries are transitioning from judge-led to prosecutorial investigation, the level of effort by judges, prosecutors, and public defenders in their processing of criminal cases will change considerably. Moreover, results registered during the first few years of the transition are unlikely to reflect those in effect once the transition has taken shape.
- Reforms intended to facilitate case processing also affect time requirements. These reforms range from the introduction of automated systems that allow all case events to be recorded electronically, thus facilitating tracking and eliminating redundant data entry, to the simplification of procedures to eliminate steps regarded as unnecessary. The latter includes introducing small claims proceeding or allowing a judge to make a default judgment when one of the parties does not respond or to strike out a case when neither party appears.
- When online adjudication is introduced for simple cases (e.g. collection of small debts or resolution of disputes over pension amounts), time invested in processing decreases dramatically. In São Paulo, Brazil's automated federal small claims courts, staff could process several thousand pension cases in a week or two, allowing the presiding judge to arrive weekly to push a button and release as many judgments. The judge interviewed indicated that these quick judgments required only a recalculation of the disputed amounts (World Bank 2004).
- The transfer of some non-litigious cases to administrative or private agents is an ongoing project in much of Eastern Europe. Here, courts traditionally handled "cases" that a local administrative office or private notary or bailiff might process instead. The impact on case weights is uncertain. On the one hand, it might not alter time invested in what remains, instead allowing more of these cases to be disposed; on the other, it might result in judges spending more time on cases within their normal workload.⁴⁰
- In other East European countries, the handling of some non-litigious cases occurring on a mass scale (uncontested monetary claims) was reformed by digitalizing their processing and transferring all these claims to a single specialized court unit serving the entire jurisdiction. The efficiencies achieved meant that other civil judges were relieved of these particular types of non-litigious cases, while the few judges that focused exclusively on processing them could do so at a much higher speed.⁴¹
- Decriminalization of some offenses reduces caseload, though with uncertain effects on the weights of what remains. In Sweden, the decriminalization of public drunkenness greatly reduced court caseloads (Svensson 2007). In Costa Rica, the elimination of the requirement that all traffic violations be reviewed by the courts had similar effects. In both examples, the elimination of these nuisance cases allowed

⁴⁰ Unfortunately, a few studies suggest that the second scenario is quite likely. The Judicial Functional Review in Serbia, for example (World Bank 2014), found a fairly strong relationship between case input and output, suggesting that judges calculate their necessary output in terms of what they receive (one explanation for why smaller courts in outlying areas have lower levels of productivity). Magaloni and Negrete (2001) find a similar relationship between input and output in Mexico's federal courts.

⁴¹ Such reforms were implemented in Estonia, Poland, and Slovenia. These countries digitalized and centralized the processing of uncontested monetary claims. In Estonia, for example, before the reform, uncontested claims, which constituted more than 50 percent of all first instance civil cases, were examined by civil judges throughout the entire country. Following the reform, only four assistant judges and 29 other court officers were able to process all such claims. Thus, the change had a significant impact on the workload of other first instance civil judges by dramatically cutting their caseloads (World Bank 2017).

judges to focus on more complex disputes. This probably did not reduce the time required to process what remained but should have allowed its faster disposition.

- As a contrary example in the sense of pushing demand upward, in both Eastern Europe and Latin America, initiatives such as economic austerity programs can often produce a sudden influx of cases protesting the consequences for individuals. Although these programs affect citizens throughout the country, the influx typically involves some jurisdictions or courts more than others. Aside from the fact that any prior CWA will not have captured these new cases, their arrival may impact the time judges and staff invest in processing their “normal” caseload. The burdens imposed by the extra work may mean some “normal” cases are left behind, but it might also encourage their more efficient handling (i.e. less time invested and thus lower case weights) unless the input-output ratio discussed above overrides it.

If the CWA (or other method) is conducted before or while such changes are made, it can provide a means of evaluating the impact of reforms or of introducing stakeholders to such concepts as case weighting, efficiency ratios, and the impacts of caseload distribution. Unfortunately, it will be a poor basis for planning staffing needs or productivity quotas. In transitional countries, the issue is frequently encountered because so many changes are being implemented simultaneously, some of them at EU or donor insistence. It does pose the question of whether to undertake a CWA at all, or if one is conducted, how much effort to put into ensuring the accuracy of its results, given that they will likely be invalidated by later reforms.

Box 12. Impact of Reforms on Balkan Analyses

Whether foreseeable or not, many CWA findings do not discuss potential reform impacts when reporting their results. The initial Serbia CWA does not, even though several reforms were undertaken during its implementation, which apparently explained why the undertaking was never approved. In Montenegro, although the report’s authors mention that the creation of misdemeanor courts came too late for capture in the analysis, they do not discuss any effects on their findings and recommendations.

The Moldova report (Hriptievski, Gribincea, and Wittrup 2014) constitutes an exception in listing the changes the CWA did not capture and suggesting that before the detailed recommendations on reallocation of judges and court mergers are implemented, “it might be wise” to examine them in light of these events. Their list includes the transfer of economic cases from specialized to ordinary district courts; the assignment to district courts of all first instance civil cases; the addition to all courts of positions intended to decrease the administrative burdens on judges; and amendments to the Civil Procedures Code removing the obligation for first instance judges to produce motivated opinions (unless requested by the parties). Not included in the list but mentioned elsewhere in the report were changes eliminating investigative judges as a separate category and allowing the Superior Council of Magistrates to periodically appoint judges to act in that capacity.⁴² Despite the cautions, the consultant team reported that some recommendations were

⁴² This last change is worth consideration by other countries. However, since the times registered for their duties were recorded before it went into effect, it is likely that the CWA overestimated the effort required by the investigative judge function.

adopted. Presumably they included the nearly across-the-board findings that most courts needed more judges and staff. Since Moldova has a far lower judge-to-population ratio than most other East European countries, this may not be unreasonable. However, considering the complexity of the analysis, it is likely that approval was based more on the results than on the path used to get there.⁴³

d. Consequences for court users

A CWA or one of the suggested alternatives can produce valuable information about institutional operations for use in improving the delivery of justice services to citizens and businesses. Unfortunately, this information is rarely used for this purpose. Particularly in organizations conducting a CWA for the first time, experience shows that there is a high risk that the process becomes a checkbox exercise to meet donor recommendations. Conducting any analysis should be viewed as a means to improve services and not an end in itself. Most of those reviewed stated their objectives as simply “increasing efficiency” by rebalancing workloads and staffing patterns. Very few did more than report whether a CWA had been approved. Even in the United States, reports sometimes note the impact on requests for more judges or eventual mergers and staffing cutbacks but rarely specify or track proposed improvements to output.⁴⁴

Part of the problem may be an extrapolation of developed country issues and solutions to transitional and developing regions. In the United States, the problem typically targeted is insufficient output due to allocative inefficiency or inadequate numbers and distribution of professionals and work units. In Western Europe, there has been more emphasis on raising output through quotas and benchmarking (technical efficiency) and reducing costs by restructuring the judicial map (allocative efficiency). Although an underlying mismatch between demand for and provision/cost of services affects countries outside these regions, it is not evident that a CWA or the alternatives discussed here point to the most effective solutions in their cases. Even if they do, there may be more resistance to their adoption. Thus, when consultants reach the predictable CWA conclusions—move staff, set quotas, close some work units—they may be rewarded with a rejection of their work, however well documented and conducted. And where significant changes (whether reductions or just movement of staff) are recommended, it is also a foregone conclusion that someone will object, however rigorous the analytic work. This suggests another dimension never mentioned in the analyses—the politics behind them and their likely results, from which some benefit and others lose.

The ultimate issue in the “imported” analyses is the failure to specify an explicit impact on services and thus a means of measuring success beyond assessment completion and adoption

⁴³ The Moldova analysis was still more methodologically complex than that in Romania or Montenegro and conceivably involved less stakeholder engagement in the process. However, except for the few courts identified as needing fewer judges and/or less staff, its recommendations, whether enacted or not, were less likely to meet resistance. Proposed court mergers are another question, but the recommendations here were more tentative, and in any event, the consultants followed the principle that the purpose was not to cut staffing but rather to redistribute it.

⁴⁴ One important exception is Doerner, Douglas, and Tallarico (2010) on the Oregon Appellate Court, which specifies an increase in cases decided and a reduction of decisions by AWOP as the project goals. However, though the results may be included in another report, they are not mentioned in the review of the analysis.

of any recommendations. If the problem is insufficient output, excessive delay, or accumulating backlog, an exploration of time expended on resolving different cases should be important in understanding its origins. These may have little to do with allocative inefficiency, however, or are unlikely to be remedied by moving judges and cases around or even setting production quotas based on current average outputs. The CWA, as adopted in transitional countries, seems structured to produce these recommendations if only because that was its purpose in the United States and Western Europe. In an analysis that automatically matches assumed problems with presumed remedies, the question of the benefits for court users drops out of the equation.

6. Lessons on the Uses of CWA and Alternative Approaches

a. For all approaches, a CMIS, preferably automated, is needed

All approaches require statistics on caseloads as well as on staffing numbers and locations. If these data are not available, they should be developed before any approach is attempted. Where statistical quality is in doubt (e.g., Moldova), there is no centralized database (Brazil, Bulgaria, Romania, Serbia), or data captured in the CMIS are limited, it will constrain what can be done. In some cases (e.g., Moldova), analysts had to compile data manually. In Brazil, a substitute was found in the monthly reports provided by all first instance state judges (and conceivably by others that were not included in the analysis). Still, extracting analyzable data from this source was a lengthy process. Samples of case files can also be used, though for a large system, this is less practical. In Serbia, the next iteration is using a small sample (1,000 cases), along with CMIS data and the council's annual reports, to derive the information needed to calculate allocative efficiency. Whether this is sufficient remains to be seen.⁴⁵

b. Conducting CWAs during or before major reforms is inadvisable

This is a lesson as much for donors, who often request/fund these analyses, as it is for justice sector authorities and national policy makers. Given the fate of analyses carried out when reforms are ongoing or pending, the best advice would be to hold off on their introduction. However, if for whatever reason an evaluation must be done, this is not the moment to use the most complex, costly approach or to promote the immediate adoption of any resulting recommendations. It might instead, assuming funds are available only for this purpose, represent an opportunity to introduce the case-weighting concepts via a pilot approach; use CMIS data to identify imbalances between caseload and staffing; or apply DEA or SFA to explore the differences in efficiency between court units. The purpose would be to introduce key concepts without expecting immediate action, since whatever the findings, their relevance is likely to change dramatically once the reform is completed.

c. There are design considerations beyond time and cost

Time and cost often figure as the major design determinants, and they are indeed significant. CWAs can become lengthy and expensive undertakings, and are not for the faint hearted.

There are also other factors to consider. The chart in Annex I offers estimates for the principal options, but they are based on limited information on the real duration of CWAs and even less on cost. As discussed above, there are ways to cut both cost and time, depending on what those ordering the analyses want to achieve. If the goal is simply to decide on caseload and staff allocation or demonstrate the varying levels of efficiency across work units, a less complex CWA or the methodological alternatives should be sufficient. However, if stakeholder involvement is advisable, a timekeeping or panel exercise, if at higher costs, may

⁴⁵ Even for a system as large as Brazil's with its 25 million annual filings, a sample can be relatively small (a few thousand), but it will only allow the calculation of averages for "the universe." Conceivably, it is interesting to know that the "average case" requires X hours of judicial input, but considering the vast variety of court and case types, this really says very little. Even with Serbia's much smaller universe (fewer than 2 million annual filings), similar considerations and limitations apply.

be preferred. Stakeholder participation does not guarantee greater accuracy, but it can enhance the chances of a CWA's approval by making the issues and analysis intelligible to a wider audience. Thus, in making adjustments to the design of the analysis, a first—and often neglected—step is to identify and prioritize the desired results, among which might figure: producing the most accurate analysis (whether or not it is approved and adopted); enhancing understanding of the issues among a wide range of stakeholders; moving to implement some recommendations; and increasing the quality and quantity of service output. A first-time analysis probably cannot do them all, which is why periodic follow-up actions should be part of the plan.

d. Efficiency analyses are not one-time undertakings

Although the goal is often just to do the analysis, some thought is warranted to when, how, and why it will be updated. On the one hand, no matter what the method, in a next iteration some preparatory steps (e.g., reaching agreement on design details, organizing the database, developing questionnaires and time logs) can be eliminated or significantly shortened, resulting in reduced costs. On the other, shortcuts taken the first time may have to be reconsidered in the next iteration, and if, for whatever reason, an analysis was conducted during or shortly before a major reform, its results may be invalidated. Thus, even as a CWA is being designed, it is wise to look ahead, take into account the next iteration, and make design choices accordingly. The realization of efficiency gains involves sustained efforts over time to review the efficiency issues, continually improve approaches to their resolution, and as needed, conduct follow-up analyses. Unfortunately, donor programming rarely takes these factors into account, resulting in a one-off CWA that soon becomes redundant. Future donor programming would do well to explicitly incorporate the continual improvement and periodic updating of such analyses to ensure the sustainability of their initial results.

e. Information from a CWA is underutilized

Considering how CWA results are often used, it merits asking whether their unique contribution—estimating differential case inputs—is worth the investment of time and funds. Tracking the level of effort by judges or other staff in processing different case types is an important exercise in its own right, whether described as a CWA (and thus translated into case weights) or conducted under some other title (e.g. time or time and motion analyses). This information can be used in various ways to enhance organizational efficiency. Although the CWA develops average times for a case or case event, deviations from the mean (derived from time logs) are just as important, as are comparisons of the averages (and the deviations) with those from other countries. Because time logs and panel estimates add costs and lengthen an analysis, it is puzzling that the information they provide is most often used only to determine “optimal staffing patterns,” especially since alternative methodologies also serve this purpose. That a CWA is conducted for whatever reason (e.g. donor requirements, its more participatory organization) is not the issue; the problem is rather the suboptimal utilization of the information it provides and the automatic assumption that its recommendations will be directed toward staffing redistribution. Since a CWA is often a donor requirement or recommendation, perhaps it falls to the donors to ask what they expect of it and to the affected countries to consider whether they can obtain more from the process.

f. Faster, less costly alternatives are available, but carry some risks also

A first issue is that these methods do not calculate case weights, and where that is required, some sort of CWA will still be needed, whether (see above) it is used to maximum effect. Moreover, these approaches often require that national partners hire experienced external experts to do the data analysis. Under this model, the external experts would need to go to some length to ensure that a broader range of stakeholders trust and understand the work being done to encourage their ownership and buy-in. Finally, CMIS analysis or DEA/SFA, as well as the use of relative weights in their application or in a “normal” CWA, can suffer credibility problems. Everyone knows what a time log is; the rest is likely to be poorly understood by many stakeholders. Even expert panels estimating real times may be mistrusted (e.g. as in Serbia), but not so much as those using relative weights. Whichever tool is used, the best risk mitigation is to build understanding of the process.

Credibility becomes more of an issue when the results are controversial. If, as in Moldova, the consultants soften their findings—in this example, not recommending staff reductions—they may be accepted regardless of doubts about the methods. Cynically speaking, in any “grading” exercise, the bottom line for most users is how they fare; if the results are viewed as positive, they are less likely to question how they were derived.

g. Concerns about accuracy may be exaggerated

Discussions about the comparative accuracy of different methods may overemphasize its importance, especially since there are no independent means of evaluating the results. Much depends on the intended use. If the analysis will be used to set maximum caseloads and productivity quotas, more accurate calculations of level of effort are important. However, if the immediate goal is a redistribution of caseload and staff between similar work units, using simpler estimates of relative weights, as in Romania and Moldova, should be adequate as long as they are acceptable to and understood by stakeholders.

h. Understanding technical inefficiency requires additional methods

In all the approaches reviewed, efficiency is always relative. What this means is that the same set of resources could conceivably be far more efficient, especially in the technical sense, if organized and incentivized differently and working under other operating rules. For many countries, these are the biggest problems, which are addressed by these approaches only at the margins—when an evaluation of technical efficiency indicates that some units are operating far above or below the norm. However, that is only a signal that there may be better ways to do things, without identifying what they are. As mentioned above, a timekeeping CWA should provide information to investigate this issue, but its use for this purpose is extremely rare. Other ways to explore the same issues include process mapping (to identify bottlenecks), comparison of timekeeping results with those of other countries, and cross-national analysis of caseload composition. In any case, it is likely to emerge that courts with workloads apparently equivalent to those in other jurisdictions tend to have a far larger

proportion of very simple cases (e.g. non-litigious matters) that are not generally handled by judges elsewhere.

i. Neither the CWA, nor its alternatives, address possibly more important problems

Aside from inefficient procedures, these problems include limited access, corruption, poorly qualified staff, political pressures, and other types of external interference. If these issues are not a concern or simply cannot be addressed, a review of allocative and “relative” (within-system) technical efficiency may be worthwhile, as long as it does not lull anyone into a sense that all else is well.

As long as countries understand that they have problems beyond allocative efficiency (which probably is not the worst of them) and use the CWA hammer only on the nails for which it was intended, conducting a CWA may not be time (and funds) wasted. However, where a CWA, knowingly or unknowingly, is presented as a panacea, it is not any more advisable than automation or code reform when those are introduced under the same pretenses. Unwarranted focus on CWAs may only delay attention to more vexing issues and to the problems that really reduce service quality. Moreover, doing so may also use resources that might be more productively invested elsewhere. For example, investment in a few process maps⁴⁶ or an analysis of randomly selected case files or CMIS data may determine where bottlenecks, unnecessary delays, and external interference occur.

Donors may fall into this trap as well. A CWA is a seemingly uncontroversial scientific undertaking, a good place to spend funds without rattling too many cages. It may also produce some measurable changes and even a few improvements, if not the ones most needed. This has been the problem with most reforms that are presented as silver bullets—training, code revision, automation, and so on—all of which are valuable, but only when used appropriately. It would be unfortunate if the CWA were to join this list in further postponing attention to the truly serious problems that are far more difficult to take on.

Box 13. What a CWA and Related Efficiency Approaches Can and Cannot Do

There is much hype around what a CWA can deliver. Many reformers are led to believe—by donors, consultants, or enthusiastic proponents of the methodology—that a CWA can solve more problems than it does. A completed CWA at most only ensures that both resources *and inefficiencies* will be equally distributed within the system covered, whether court, district, or nation-level. The alternatives are similarly limited. They will not resolve the issues that are impeding access, facilitating external inference, or making everyone less productive, nor will they call attention to the fact that the most efficient judge or prosecutor in country X only disposes one-fifth of the cases that the average judge does in country Y. The best designed CWA still says little about the absolute value of the cases that a judge or prosecutor does decide, which is to say that even if credited at only one-tenth the weight of a debt collection case, validation of a document in

⁴⁶ See World Bank (2015) for an example in Bulgaria.

countries where judges handle such things should not be done by a court. The same limitations apply to the “sophisticated” shortcuts that only propose to address efficiency issues without addressing other problems.

A CWA (or one of the alternatives) can direct attention to some of these problems, but only if delivered with that goal clearly stated at the start. Certainly, some of the technical efficiency work—such as Castro’s (2011) study in Brazil—should attract attention to the bigger issues, including the fact that even under similar rules and conditions, many court units (and judges) operate far under the productivity frontier. More difficult but certainly worth attempting would be to compare “case weights” to those derived elsewhere. As more analyses are done, more results are available for comparison. Certainly the 2.5 minutes it takes a Kentucky judge to decide a traffic case or even the 20 hours s/he invests in a homicide (Ostrom, Kleiman, and Lee 2016) might give pause for thought to judges who rarely spend less than an hour (and up to four hours) on the simplest cases.⁴⁷

Thus, in the end, perhaps the most critical lesson is the importance of understanding what a CWA can and cannot do in the context of the problems a judicial system currently faces. This discussion was aimed at the first part of the sentence, but its utility depends on the second—a good understanding of the existing problems. It also aimed to provide guidance on the variations in the case-weighting approach, which should be clearly understood by anyone choosing to use one.

⁴⁷ It is also notable that the greatest differences appear to affect simpler cases. The same Balkan country recording these numbers averaged more reasonable times (or closer to those recorded elsewhere) for more complex issues. However, simple cases typically represent at least 60 percent of the caseload and therein may lie the answer to how judges fill their time.

ANNEX 1: GUIDE TO USING CWA AND ALTERNATIVES

Approach	Costs ⁴⁸	Time ⁴⁹	Typical Use	Potential Uses	Comments
CWA, case-based time logs	High	12–18 months	Allocative efficiency	Technical efficiency	Although time logs may take as little as 3–4 weeks, log design, other preparatory work, and final analysis account for the rest of the time. Online apps may cut time but raise costs further. That CWA logs have not been used to estimate technical efficiency appears to be more an issue of court interest than of its potential. Event logs are faster, as cases need not be traced through their entire trajectories. Delays are common.
CWA, event-based time logs	High	12–18 months	Allocative efficiency	Technical efficiency	
Expert panel case based, real time	Moderate	8–12 months	Allocative efficiency	Technical efficiency	Panels also require preparatory time, and depending on the novelty of their use, this could make the approach nearly as lengthy as time logs. Depending on how they are organized (size, any travel for members, special training), costs could also increase. Still, the general consensus is that they are cheaper and faster than timekeeping. As with time logs, the emphasis on allocative rather than technical efficiency seems to be a product of court interest rather than real potential.
Expert panel event based, real time	Moderate	8–12 months	Allocative efficiency	Technical efficiency	
Expert panel (either type) with relative values	Moderate	8–12 months	Allocative efficiency	Technical efficiency	The pros and cons are much like those for real-time panels, but more preparation and training may be needed for introducing relative values to panel members.

⁴⁸ Since most studies did not include costs, these are estimated on the few that did.

⁴⁹ Time estimates are based on what appears to be a lowest reasonable time. Durations vary considerably, especially in transitional countries. Most of the differences are attributable to longer preparatory times and difficulties encountered in processing CMIS data.

CWA time logs plus panel	High +	12–24 months	Allocative efficiency	Technical efficiency	Clearly the longest and most expensive approach, but for its advocates, the most “accurate” method, giving an opportunity for more readjustments and with luck, consensus on the outcomes.
CWA, questionnaires or survey for time estimates	Low	12 months	Allocative efficiency	Technical efficiency	Method was used in Bulgaria and done fairly quickly at a very low cost. Additional time was required for panel review and testing, but is not added here. However, response to the questionnaire was only 60 percent and probably incorporated biases. Still, for a first try and one leading to the adoption of recommendations, it is worth noting by countries on their initial attempt, especially if, like Bulgaria, they lack a centralized CMIS database.
CMIS case statistics analysis	Low	4–6 months	Technical and allocative efficiency	Estimating case weights	As long as some effort is made to estimate case complexity (relative values, probably with a panel), a good CMIS may offer the quickest way of estimating allocative efficiency, and even without relative values may be adequate for comparing courts of the same type. The unanswered question is whether a CMIS on its own offers enough information for an estimate of case weights and thus a more complete assessment of technical efficiency.
DEA/SFA (with or without case weights)	Low +	8–12 months	Technical and allocative efficiency	Staffing ratios, technical efficiency based on case complexity or more than raw numbers	These are interesting approaches and once done, provide convincing graphics. However, as Wittrup notes, the math may be beyond the abilities of many courts (or other justice institutions). As with the CMIS-based analysis, times and costs could be extended, depending on the shape of the institutional databases. It could easily take outside consultants several months simply to organize the data needed to do their analysis, (especially) if there are translation issues.

ANNEX 2: ANNOTATED BIBLIOGRAPHY ON THE CWA AND RELATED APPROACHES TO IMPROVING SECTOR EFFICIENCY

GENERAL MATERIAL ON THE CWA

Flango, Victor E., and Brian J. Ostrom. 1996. "Assessing the Need for Judges and Court Support Staff." National Center for State Courts (NCSC), Williamsburg, VA. Available at

<http://cdm16501.contentdm.oclc.org/cdm/ref/collection/ctadmin/id/407>

This is an early but classic explanation of the CWA (case-weighting analysis) as used to assess court staffing needs. It was drafted by NCSC experts with extensive experience in the method. It also compares alternatives to the use of time logs (Delphi panels, simulation, regression) and contains sections on the application of a CWA to "quasi-judges" and ordinary staff. The authors anticipate the use of a CMIS to determine the frequency of case types and events and provide examples of case- and event-based timekeeping from several U.S. states. They offer good details on alternative methods as well, although their stated preference is for timekeeping (whole case or event log versions) to calculate the number and distribution of judges required by existing demand.

For a study that is 20 years old, it is remarkably prescient in its inclusion of alternatives and their limitations, as well as ideas on how to overcome some of the frequent caveats about a CWA (e.g., the problem of logging inefficient practices, high costs). Although providing considerable information on the mathematics of the process (ranging from sample size to dealing with rare events/case types to estimating judges' time available for case processing), it is not a recipe for CWA design. Some of its recommendations (e.g., use of samples) may also be outdated, given the increasing availability of good CMIS techniques, web-based time logs, and existing timekeeping requirements for some events and system actors.

Gramckow, Heike. 2011. "Estimating Staffing Needs in the Justice Sector." World Bank, Washington DC. Available at

<http://documents.worldbank.org/curated/en/958421468324281209/Estimating-staffing-needs-in-the-justice-sector>

Like Flango and Ostrom, this study covers alternative approaches while recommending a CWA (which it separates from Delphi panels) as the preferred method for estimating staffing needs. It does note that panels may have to be used because of time and budgetary constraints, and that the CWA, like many other approaches, may "be based on current inefficiencies in the system." It provides a good, if slightly dated, overview of other analyses and especially timekeeping exercises, and a nearly blow-by-blow account of how to do a CWA, including a detail omitted from the others relating to process mapping.⁵⁰ No time log examples are given, but there are recommendations on references that include them. The study is probably a better review of a CWA for prosecutors than the APRI document (see below). For both prosecutors and public defenders, it notes that care should be taken to include out-of-court activities when timekeeping or estimates are done. It also provides suggestions for dealing with quality and access variables. Given the date of the study and of the references cited, it

⁵⁰ Process mapping is not usually a feature of a CWA, but as Gramckow specializes in it, she has added it here. Readers can decide whether it would add value to what they intend to do.

is a little behind the curve on recent IT innovations that can facilitate timekeeping in particular.

Jacoby, Joan E. 1985. "Case Weighting Systems for the Public Defender." National Institute of Justice, U.S. Department of Justice, Washington, DC. Available at

<https://www.ncjrs.gov/pdffiles1/Digitization/100114NCJRS.pdf>

This is a manual intended to help Public Defenders Offices do their own CWA. It includes examples of manual time logs and calculations of attorney time spent on case and non-case activities. It tends toward an event-based approach, as attorneys are asked to record the activity performed and disposition type for each case. Because the article predates innovations such as online time logs and sophisticated CMIS methods, it will be most useful for defenders offices that are initiating a workload analysis but have little access to IT tools. The annexes include some statistical analyses but do not explain how and why they are used. Since the illustrative example involves a small office with few attorneys, the implication is that all will be covered. For larger offices, this might not be feasible, although U.S. practices (and those in many other countries) often require attorneys to use a billable-hours system and hence keep logs anyway.

Langbroek, Philip, and Matthew Kleiman. 2016. "Backlog Reduction Programs and Weighted Caseload Methods for South East Europe. Two Comparative Inquiries." Regional Cooperation Council, Sarajevo. Available at

<https://dspace.library.uu.nl/handle/1874/329022>

As the title indicates, both backlog reduction and the CWA are covered, though only the latter is reviewed here. This is essentially a desk review, covering not only six "beneficiary" countries (Albania, Bosnia-Herzegovina, Kosovo, FYR Macedonia, Montenegro, and Serbia) but also experiences from the United States, other parts of Europe, and the Middle East (the West Bank). Of the two principal authors, only Kleiman has significant hands-on experience, and many of the analyses in which he participated are included. The authors regard time logs as the "gold standard" for the CWA but also note that the method has been refined considerably over the past 20 years (since the Flango and Ostrom report). Sampling is now less necessary as logs can be web based, allowing virtually all judges to participate, and the length of the time-logging exercise is now between three and four weeks. However, an entire analysis may take over a year, including a long preparatory period, a second web-based questionnaire on time sufficiency, and quality adjustment sessions using Delphi panels of experienced judges.

The U.S. analyses cited were used, with some success, to request funding for more judges, and in one case, to reduce the number of judges. Experience in the six beneficiary countries with Delphi, time logs, or mixed modalities was less positive, in part due to incomplete CMIS development in several states and in part to varying interest in the process. Despite a preference for real time-based analyses, the authors recommend a Delphi approach for the beneficiary countries because it is "less burdensome, less expensive, and can be completed in less time." Although the study offers no "recipe" on how to conduct a CWA, the long list of references could provide more information (and details such as time logs, software, and so on).

Leinhard, Andreas, and Daniel Kettiger. 2011. "Research on the Caseload Management of Courts: Methodological Questions." *Utrecht Law Review* 7 (1): 66–72. Available at

www.utrechtlawreview.org/articles/10.18352/ulr.147/galley/146/download/

Leinhard, Andreas, Daniel Kettiger, Daniela Winkler, and Hanspeter Uster. 2015. "Combining a Weighted Caseload Study with an Organizational Analysis in Courts: First Experiences with a New Methodological Approach in Switzerland." *International Journal for Court Administration* 7 (1): 27–36. Available at

www.iacajournal.org/articles/10.18352/ijca.174/galley/168/download/

These two articles, both about the Swiss experience, are placed under general studies, as the first includes a discussion of alternative methods. As with the German PEBBSY (see Ketterle, below), the authors' preferred method is to measure the average time per case rather than to use an events-based approach. A case-based approach requires more time for data collection (six months, according to the authors), and as the authors admit, treats the "whole procedure...as a 'black box.'" The authors note that the Swiss remedy for this issue was to divide case-processing time into a pre-trial and judgment phase. The authors also discuss an "estimated" (i.e., relative) value approach, concluding that though difficult to compare across systems, the relative values appear to hold when compared to time estimates from logs or Delphi panels from the same one. Despite some apparent preferences, the authors conclude that "no unitary approach... is likely to emerge in the near future as the preferred method" and that "for the purpose of allocation of resources, detailed accuracy is not necessary; it is sufficient to use a rough scale of approximate values, subdivided into three to four general case types."

SPECIFIC CWA APPLICATIONS

There are numerous studies that could be cited, especially given the over 40 years of applications in the United States and the growing number in Europe. The following references are only illustrative, presenting a mix of standard approaches and a few with novel elements.

APRI (American Prosecutors Research Institute). 2002. "How Many Cases Should a Prosecutor Handle? Results of the National Workload Assessment Project." APRI, Alexandria, VA. Available at

www.ndaa.org/pdf/How%20Many%20Cases.pdf

The report documents the caseload and workload assessments done by the APRI in 56 state prosecutors' offices across the United States. One initial aim—to develop common workload standards for all such agencies—was deemed impossible because of differences in criminal codes, court structures, and crime profiles. However, the "disposition-based" model (including both case events and manner of disposition) was recommended for use by individual offices. A lengthy introduction provides information on factors affecting productivity and CWA experience for both courts and prosecutors. The portion devoted to the preferred CWA model is relatively short and recommends only that staff record prosecutors' time by case type, disposition type, and three procedural stages (pre-charge, pre-trial, and trial). Disposition type was considered essential, as most cases are disposed without a trial in the pre-charge or pre-trial stages. Resource needs (or number of full-time equivalent attorney positions) are calculated by case type, taking into account the frequency of different disposition types and the time available (subtracting vacations, leave, and non-case-related work). Support staff was not included in the estimates, although presumably the

same method could be used (with the same issues of providing standards across different districts).

Diller, Jim (listed as contact, not necessarily the author). 2017. "Weighted Caseload Measures & the Quarterly Case Status Report." Indiana Judiciary, Indianapolis, IN. Available at

<https://www.in.gov/judiciary/admin/files/pubs-trial-court-weighted-caseload.pdf>

This short (six-page) document reviews the CWA methodologies used by the Indiana courts from 1996 to 2016. Until 2015, Indiana's case-weighting analyses (1996, 2002, and 2007) used time logs from samples of courts and judges and limited the case types covered. In 2015, improvements to IT tools enabled the participation of all judicial officers in the state and coverage of all case types. The software app (CAPTURE) also recalculated (downward) the time available to judges for case-related activities. In addition to the online time logs, data came from the case status reports (which, since 2007, all Indiana courts have been required to submit quarterly online). Time estimates from the four CWAs are provided for comparison.

Doerner, John, John Douglas, and Suzanne Tallarico. 2010. "Oregon Court of Appeals Judicial and Staff Weighted Caseload Study." State Justice Institute and NCSC, Williamsburg, VA. Available at

www.courts.oregon.gov/COA/docs/orcoaworkloadfinalreport.pdf

The report is included as one of the few to incorporate non-judge staff and because it may be useful for countries that cannot use online time logs. When it began, Oregon's Court of Appeals had retained the same number of judges (10) since 1977 despite a rising workload. After considering tracking a "large number of cases" from opening to closure, the researchers chose an event-based method because it required less time and fewer resources—and because cases close at different points in their trajectories. Daily time logs were done on paper and then transferred to an online tool (which included an error check for impossible entries). A copy of the daily log divided into 10-minute intervals is included in an Appendix. Participants in the six-week timekeeping exercise included nine judges, 11 staff attorneys, 15 law clerks, seven judicial assistants, and three settlement court program staff, though the latter were subsequently discounted. Case weights for all participants were divided between affirmation without opinion (AWOP) and authored opinion cases. To reach a goal of 100 percent clearance rates and a reduction of AWOPs, increases in all personnel categories (except settlement staff) were recommended. The report includes advice on how to reduce errors in manual recording or in the transfer of data to an automated database.

Florida State Legislature. Office of Program Policy Analysis and Government Accountability (OPPAGA). 1998. "Information Brief on Weighted Caseload Methods of Assessing Judicial Workload and Certifying the Need for Additional Judges." Report 97-67, OPPAGA, Tallahassee. Available at

<http://www.oppaga.state.fl.us/MonitorDocs/Reports/pdf/9767rpt.pdf>

This very brief report, while outdated, is an excellent review of the four major alternatives considered for a new CWA in Florida's state courts. It discusses and evaluates both timekeeping and Delphi panels organized either by case type or events, and details the inputs required and cost of each. The initial cost and input estimates were prepared by a consulting firm, but the Office of the State Courts Administrator (OSCA) increased both, adding staff and

meetings that would be required to carry out the work for each alternative. On the basis of cost and time required, a Delphi event-based method was chosen.

Grubišin, Maja. 2015. "Montenegro Case Weighting Study: Final Report with Recommendations.", Podgorica. (Not available online)

This report, which includes eight attachments with statistical analysis, provides a detailed description of this CWA, including information on the Montenegrin court organization, the composition of the working group (which also served as the Delphi panel), the selection of participants in the time log exercise, the calendar for the various stages and meetings, the use and financing of an online time-logging software, and even the days of in-country presence of the EUROL⁵¹ advisors (among whom Grubišin was the principal). The eight attachments document the statistical analysis by stages, including the differences between the two case categorization systems (CWA and PRIS⁵²) used and the Delphi panel adjustments.

For better understanding, the final report text could have included additional details, such as the decision to use the PRIS categories for the final analysis (including the estimates of staffing needs), how the percentages for cases at each level of complexity were derived, and why staffing needs were calculated for two scenarios: first for only new cases and then for new and pending, with the latter significantly raising staffing estimates. Despite this increase, the final calculations for handling the entire caseload indicated an excess of assistants and administrative staff in most courts. Recommended changes for judges were more modest, including both slight reductions and increases, and under the first scenario (based only on new filings), estimates of required staff were considerably lower.

Aside from complicating readers' understanding of the analysis, these and some other details (including more on the Delphi calculations) could be useful in understanding whether more time, a more permanent advisory presence, a simpler design, or less ambitious goals might have produced less controversial results. Although probably not appropriate for a report to the client, the advisors' and participants' assessment of the online time logs would have also been useful, since had they been used for the staffing recommendations, the recommendations would have been quite different.

Ketterle, Roland. 2013. "Court Financing: The Workload Measuring System PEBBSY.", Bucharest, March 11. (Not available online)

This is a detailed PowerPoint presentation on the CWA adopted in Germany in 2001 that replaced an earlier approximation based on "estimation and tradition." The system has been used since 2004. It incorporated a new categorization of disputes, was conducted in 46 different courts of all sizes and instances, and covered 6,000 judges and other staff and 890,000 case files. Because it aimed at covering all work involved in each case, its length hinged on case trajectories and so extended over six months. Subsequent data collection has been extended to other types of proceedings. Unfortunately, the time estimated as needed by judges, *Rechtspfleger* (judicial officials), and support staff to cover their existing caseloads exceeded that available, meaning that the system has attracted complaints since its introduction. Since this is only a PowerPoint, it includes little information on the details of

⁵¹ EUROL is a European Union-funded technical assistance project helping the Ministry of Justice, State Prosecution Office, and Police Administration of Montenegro to strengthen the judiciary.

⁵² For reasons never fully explained, the team decided to use the case categories from the existing CMIS (called PRIS) and a slightly more complex version developed by the working group. However, the final (and most controversial) analysis used only PRIS and the Delphi adjustments.

CWA implementation (e.g., how participants and cases were chosen, how timekeeping was organized, whether logs were manual or online). However, it is useful for understanding the German system, if not for explaining its merits. It does demonstrate some drawbacks of the case-based approach—the longer time needed to complete the analysis, for example, and as Leinhard and Kettiger (see above) note, the tendency to treat the case trajectory as a “black box.”

Lombard, Patricia, and Carol Krafka (Project Directors). 2005. “2003–2004 District Court Case-Weighting Study. Final Report to the Subcommittee on Judicial Statistics of the Committee on Judicial Resources of the Judicial Conference of the United States.” Federal Judicial Center, Washington, DC. Available at

<https://www.fjc.gov/sites/default/files/2012/CaseWts0.pdf>

The report outlines an approach adopted by the U.S. federal courts and applied as a revision of a CWA done in 1993. Only judges’ time was covered; staff or “quasi judges” were not included. The new approach uses a combination of data from the district courts’ CMIS and other records to calculate the frequency of all events and times for some of them (especially trials). To estimate event times that the CMIS data do not provide, a Delphi technique was used. The process is covered in detail, including the conversion of raw weights (time calculated) to relative weights by case type. Event and case type frequency were tracked from 1998 to 2001. Delphi panel judges (102 between three panels) were first given forms to enter their own initial estimates of event times,⁵³ with results from the 1993 CWA provided as defaults. The CWA team recommended an iterative Delphi approach to reach consensus on times and to incorporate adjustments for more complex cases. Although the method could also be extended to prosecutors and public defenders, as other studies (APRI, Gramckow) indicate, many of their activities are not reflected in their agencies’ CMIS and would have to be estimated separately.

Mircic, Vucko. 2012. “Weighted Caseload in the Courts of Serbia.” Report submitted by the High Court Council Working Group for the Development of a Weighted Caseload System and prepared with assistance from the USAID Separation of Powers program. (Not available online)

Prepared by the 12-person Working Group Chairperson, this report describes the Serbian CWA conducted between 2010 and 2012. The two-and-a-half-year analysis received support from an international advisor and Serbian U.S. Agency for International Development (USAID) project staff. It began with a Working Group that also functioned as a Delphi panel to identify the courts, cases, and event types to be weighted, and then estimated the average times needed by judges to resolve cases divided by type and level of complexity. Input by other staff was not considered. The categories developed by the Working Group were used in a subsequent three-month manual timekeeping exercise in 37 courts (and among 386 judges). The reasons for its inclusion were doubts about the accuracy of Delphi estimates in general. The Working Group based its development of case weights on the time logs, subsequently making adjustments where it believed data quantity was low for certain case types or complexity levels. Finally, the methodology was applied to five courts, resulting in recommendations on staffing levels (all but one of which were increases).

⁵³ Insisted on by the General Accounting Office, which objected to a fully consensus-based approach that would not allow calculation of standard deviations or medians.

The report was never approved by the High Judicial Council, nor were its recommendations on the methodology's future application to determine staffing needs. One problem, which the report notes in part, involves changes during and after the exercise that would affect caseload levels and composition. These include the diversion of enforcement and non-litigious cases to private bailiffs and notaries; the reinstatement of judges dismissed during a reevaluation process; and the entrance into effect of a criminal procedures code eliminating judicial investigation of crimes. The analysis is a good example of Working Group involvement and apparent direction of a CWA, providing detailed information on each step of the process and the problems (with some solutions) encountered along the way.

Ostrom, Brian J., Matthew Kleiman, and Cynthia G. Lee. 2016. "Kentucky Judicial Workload Assessment: Interim Report to the Administrative Office of the Courts Kentucky Court of Justice." National Center for State Courts, Williamsburg, VA. Available at

courts.ky.gov/resources/publicationsresources/.../InterimReportJudicialWorkload.pdf

The report describes an unusually complex CWA that began with judges keeping logs for four weeks. The logs covered case-related and non-case-related work and were followed by a "quality adjustment" questionnaire, asking participants about time sufficiency for the activities. Three Delphi panels then did a qualitative review of the preliminary case weights. Their recommendations were oriented toward adjustments providing more time, especially for dealing with court users. The timekeeping portion involved all judges. It consolidated the cases into 33 types and divided case-related events into five categories: pre-trial; non-trial/uncontested disposition; bench trial/contested disposition; jury trial, and post-judgment/post-disposition. Data on number of filings, case types, and dispositions for the period 2012–14 were taken from the courts' CMIS. Judges tracked time in five-minute intervals using a web-based form. Case weights (time spent per case type) varied from 2.8 (traffic) to 1,118 minutes (homicide). Non-case-related work was estimated at 2.3 hours per eight-hour working day. As this is an interim report, results in terms of incrementing judicial positions were not included.

Ostrom, Brian J., Matthew Kleiman, and Cynthia G. Lee. 2013. "Virginia Judicial Workload Assessment." National Center for State Courts, Williamsburg, VA. Available at

http://www.courts.state.va.us/courts/scv/virginia_judicial_workload_assessment_report.pdf

This is a typical NCSC CWA, using event-based logs to calculate time spent by circuit and district court judges in processing different types of cases and so estimating staffing needs (based on incoming and pending caseload). The resulting recommendations included a series of alternatives for boundary realignments. The analysis used web-based logs, with entries by case type and case-related or non-case-related event. Over a four-week period, 375, or 97 percent, of all full-time judges participated, as well as 41 retired judges who were helping to deal with accumulated cases. Filing data for 2010–12 was provided by the court CMIS, disaggregated by case type and jurisdiction, and an annual average was calculated from the results. The logs consolidated cases into 16 types for circuit courts, eight for district courts, and nine for juvenile and domestic relations district courts. Case-related events were divided into a maximum of five categories (adjusted to lower numbers and types of events according to court type).

Pace, Nicholas M., Greg Ridgeway, James M. Anderson, Cha-Chi Fan, and Mariana Horta. 2011. "Case Weights for Federal Defender Organizations." RAND Safety and Justice Program, The Rand Corporation, Santa Monica. Available at

www.rand.org/content/dam/rand/pubs/technical_reports/2011/RAND_TR1007.pdf

This study was contracted by the Office of Defender Services of the Administrative Office of the U.S. courts to help estimate "the funding and staff requirements of federal defender organizations (FDOs) throughout the United States." It combined a real-time-based CWA with an examination of additional factors affecting such resource needs as region, court practices, and other exogenous variables. As data were drawn mainly from districts' CMIS and timekeeper systems (TKS), there was no need for separate time logs and the complete list of case types (284 in total over five years) was maintained. Although the TKS may have allowed an event-based calculation, this was not used; instead, averages were developed for time invested in all cases closed within the period. RAND applied some more sophisticated statistical methods for normalizing case weights, thereby reducing the impact of outliers (i.e., extreme times) on the mean average and making it possible to compare average attorney times within the same district. Although the study cautions that the role of non-attorney staff can make a difference, it also concludes that weighted caseloads "might not be the best way to make that assessment."

Washington State. Office of Public Defense (OPD). 2014. "Model Misdemeanor Case Weighting Policy." OPD, Olympia. Available at

www.opd.wa.gov/documents/0192-2014_MM_CaseWeightingPolicy.pdf

This study describes a CWA for attorneys handling misdemeanor cases. Although limited to misdemeanors, a similar system could be used for other types of cases. Prior to the CWA, attorneys were to receive a maximum of 400 new cases annually. Offices can now substitute a maximum of 300 weighted credits as developed under the analysis described. The weighted model was developed by tracking attorney time over 20 weeks in 15 different courts of limited jurisdiction. Pre-existing data from two courts were also included. Credits were based on an average of 4.5 hours per case; cases requiring nine hours received 1.5 credits and those requiring three hours received 0.5 credits, with 17 other case types ranked in between. A guide for customizing the case weights for local offices is also included, with a list of factors that might increase or decrease the weights and credits. Guidance is provided on reducing rates for early non-criminal resolutions (for localities whose courts use that practice). The report does not include details on the procedure used for tracking events, examples of time logs, or the criteria for selecting the sampled courts. It does list the out-of-court actions an attorney must perform for any type of case. Although factors such as information technology, travel, and wait times (to see clients or be called to court) were mentioned as affecting times, how they are factored into case weights was not discussed.

ALTERNATIVE MECHANISMS USED TO ESTIMATE OPTIMAL WORKLOADS AND STAFFING

The reports reviewed below were produced by project consultants and focus on ways to improve staff distributions without a full-blown CWA. They recommend reliance on data already available from a CMIS and, if case weights are desired, panel estimates. Although the goal is improving allocative efficiency, the methods could be used to track "technical efficiency" within the given court systems. However, as with the single country academic

studies (next section), this type of technical efficiency is relative and limited to output differences between courts using existing procedures and practices.

EWMI (East-West Management Institute). 2011–2013. “Determining and Implementing the Optimal Volume of Work of Judges and Court Clerks and Ensuring the Quality of the Courts’ Activity.” Romania Analysis and Assessment Report, 2011; Romania Final Report, 2013; Papers 1A, 1B, and 1C on benchmarking exercise. (Not available online)

Although the title refers to optimal volume of work, the exercise really focuses, as in Moldova (see below), on optimal court size (i.e., number of judges) relative to workload, as well as other issues (e.g., training, modifications to CMIS design, adoption of key performance indicators) not covered here. The preferred technique is Data Envelopment Analysis (DEA) as outlined in paper 1C and compared to other methods in the 2011 report. The team also developed a simplified set of case categories for use in a “normal” CWA, reducing the roughly 1,000 case types to 17 (paper 1A) and a further set of relative weights (five groups as opposed to Moldova’s three) based on case complexity (paper 1B). As in Moldova, the relative weights were derived by multiplying the complexity ratio by the average time for processing all cases. The final report lists a series of recommendations, only a few of which applied DEA to define optimal court size. However, according to Wittrup, the recommendations on staffing were not adopted, owing to a subsequent system-wide reorganization of court districts. As seems typical of DEA advocates, the discussion does not extend to more complex mixes of inputs (judges, assistants, support staff) and outputs (case and disposition types).

Interestingly, Wittrup, who applied the DEA methodology here and in Moldova, notes that given the frequent need for external consultants to apply more sophisticated benchmarking techniques, judiciaries may find it practical to use a CWA based on relative weights. As in Moldova, the study mentions issues with “case” definition and classification in the CMIS, but in the end, relied on local conventions. One issue, apparent from looking at the “cases” categorized, is that local convention mixes events and principal cases. This is a cause of caseload “inflation,” some of which is intentional, some conventional, and some inconsistent across courts.

Hriptievschi, Nadejda, Vladislav Gribincea, and Jesper Wittrup. 2014. “Study on Optimization of the Judicial Map in the Republic of Moldova.” Legal Resources Centre from Moldova, Chisinau. Available at

http://www.justice.gov.md/public/files/file/reforma_sectorul_justitiei/pilonstudiu1/Studiu_Optimiz_Hartii_Jud_-_CRJM-2014_en_2.pdf

This report describes an analysis conducted to assist the Moldovan Government in optimizing judicial efficiency through the reallocation of judicial resources (judges, staff) and the potential merger of courts. According to one of the authors, the resulting recommendations were adopted by the Parliament as part of a major court reorganization. These recommendations were intentionally conservative, as the aim was not to reduce staffing numbers but rather to effect a better distribution. The analysis used DEA, working with statistics on court workloads (numbers and types of incoming and resolved cases) taken from Ministry of Justice information and the judicial council’s annual reports. Delphi panels were used to categorize cases (already separated by legal matter and type of court) into three levels of complexity. An average processing time for all cases (total time available for case processing divided by number of cases processed) was used to assign numeric weights to cases in each of the three categories. A regression analysis and a ratio model were added to

test the relationship between population characteristics and workloads and to estimate optimal numbers of non-judge staff per court. Apparently, a companion study on prosecution services was intended or carried out but was not made available.

The report offers advice on dealing with incomplete data and checking for inconsistencies between sources. It unfortunately does not suggest how or whether DEA could be applied to more complex mixes of inputs (for example, testing different staff/judge ratios). The report's Annex 6 offers a comparison of the CWA and DEA approaches, but the main text compares only the staffing recommendations derived from the DEA analysis of different annual statistics, using simplified case weights. There is here, as in Romania, a missing step (the black box) on the efficiency coefficients used to develop these estimates.

McMillan, James E., and Carolyn E. Temin. 2011. "Dynamic Case Weighting – Using the Data We Have to Manage the Courts." *The Judges' Journal* 50 (2). Available at

<http://connection.ebscohost.com/c/articles/48918802/dynamic-case-weighting-using-data-we-have-manage-courts>

Published six years ago, the article endorses something like the approach used by the U.S. federal district courts (see Lombard and Kafka above): a combination of data available from the CMIS with estimates by experts to fill in the blanks (i.e., the time devoted to events the frequency of which can now be derived and updated from CMIS data). It promotes a quicker (than a full-blown CWA) means of gauging case complexity and thus workloads based largely on statistics already maintained by courts.

ACADEMIC STUDIES USING DEA AND OTHER MATHEMATICAL TOOLS TO STUDY COURT EFFICIENCY

The number of these studies is growing; only four examples are given here. They all focus on within-system comparisons of technical efficiency, useful for determining which courts do better but missing the question of how the national system compares to those elsewhere. Another issue is that academics tend not to include case complexity, meaning that a court could be "relatively efficient" while resolving only simple cases (and leaving the rest unattended).

Castro, Alexandre Samy de. 2011. "Indicadores Básicos e Desempenho da Justiça Estadual de Primeiro Grau no Brasil." Report 1609, Instituto de Pesquisa Econômica Aplicada (IPEA), São Paulo. Available

at http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=18243&Itemid=6 (Several related, subsequent studies are available at the IPEA website <http://www.ipea.gov.br>)

IPEA is a publicly funded research institute. The Castro study and several others following up on its findings were contracted by other Brazilian government agencies (Secretary of Strategic Matters within the Office of the Presidency, National Judicial Council, and Ministry of Justice). This work is placed with the academic studies, as it is not project connected and uses an alternative methodology—in Castro's case, stochastic frontier analysis (SFA).

Castro's study used the National Juridical Council's database of monthly reports from judges in 8,495 first instance *serventias* (essentially the lowest level single or multi-judge court unit) located in Brazil's state court systems (26 states and each state's federal district). Since state

courts receive two-thirds of Brazil's 25 million annual filings, the data taken from eight months of reports potentially included nearly 12 million cases. Only the Castro study uses SFA; it was followed by several others evaluating specific reforms and adding a combination of case file sampling, observation, interviews, and focus groups. Castro's most basic—but also most radical—findings are the enormous differences in *serventia* efficiency (as measured by merit-based decisions reached per judge) between and *within* states. Further analysis of internal and external characteristics associated with levels of efficiency (or inefficiency) are very Brazil specific, although the use of stochastic analysis to test them could interest other research efforts.

Given issues with the database (including incomplete as well as redundant reporting and no apparent effort to determine suitability for a parametric approach), a purist might object to the choice of SFA. The authors argue that its use was permitted by the sheer quantity of data and the further assumption of significant homogeneity in the production function—the basic organization and caseload composition between state court work units. Although federal agencies continued to fund IPEA research, the impact of the Castro study is unknown. Although produced at the same time as the Yeung and Furquim de Azevedo work (see below), Castro's piece represents a step forward in calling attention to the within-state differences and linking them to certain, admittedly very Brazilian, organizational characteristics.

Nissi, Eugenia, and Agnese Rapposelli. 2010. "A Data Envelopment Analysis of Italian Courts Efficiency." *Statistica Applicata – Italian Journal of Applied Statistics* 22 (2): 199–210. Available at

<https://www.statindex.org/articles/272221>

This study focuses only on Italian Courts of Appeal for the year 2008, using DEA to compare their "technical efficiency." It uses as input variables the number of judges and size of caseload (pending plus incoming cases) and as output, the number of dispositions for the year. It makes no distinction between case types and uses the caseload as a control so as not to underestimate productivity for "years in which a court is charged with a small caseload." Two different efficiency models are used: one that evaluates overall technical efficiency and the other that estimates technical efficiency at the given scale of operations. Although the efficiency quotients vary according to the model, the order of courts from most to least efficient is not changed. Given the usual impression that Italian courts are inefficient (with long delays and large backlogs), the conclusion that most appellate courts are operating "at a quite high level of efficiency" is puzzling. Still, the finding of differing levels of efficiency is suggested as a means of benchmarking courts to determine which practices make some more efficient than others. The study does note that the efficiency analysis could be improved by studying performance over time and by taking into account the relative complexity of cases.

Rosales-López, Virginia. 2008. "Economics of Court Performance: an Empirical Analysis." *European Journal of Law and Economics* 25 (3): 231–51. Available at

https://www.researchgate.net/publication/5145811_Economics_of_Court_Performance_An_Empirical_Analysis

The study addresses two questions on the Spanish courts: why some of them have higher output than others; and whether courts could produce more using their actual resources. It also examines whether high output courts have higher reversal rates on their decisions, thus indicating that greater productivity may mean lower quality. The author used a stepwise

regression model to identify five factors that explain 54 percent of the variance between courts: the court's size (staff numbers), workload, availability of Common Procedural Services (only present in half the courts), reinforcement of direct support services, and judicial turnover. Except for turnover, all the factors had a positive effect on output. No relationship was found between output and the frequency of reversal rates. Although the study concluded with some policy recommendations, there was no indication of any impact. The method does indicate means (the recommendations) for raising productivity without increasing the number of staff. Some of the results were obvious—that court size and workload would relate positively to output, for example—but the remaining factors are interesting as they are more under the control of policy makers. However, with 46 percent of the variance (and especially issues like the procedural complexity or the availability and use of technology) still not explained, there are obviously more issues to explore, and they are the more difficult to measure.

Yeung, Luciana, and Paulo Furquim de Azevedo. 2011. "Measuring the Efficiency of Brazilian Courts from 2006-2008: What do the Numbers Tell Us?" Working Paper 251/2011, Instituto de Ensino e Pesquisas, Brasilia. Available at

https://www.insper.edu.br/wp-content/uploads/2012/10/2011_wpe251.pdf

This study compares the efficiency of Brazil's 27 state court systems using an output focus, where output was measured by dividing dispositions (in both first and second instance) by total workloads (pending plus incoming cases). Case complexity was not considered. Inputs included number of judges, number of staff, and number of computers. The number of judges and staff were weighted by workload to produce relative values of inputs (i.e., the number of judges [or staff] available for every 100,000 cases in court). Using DEA, state courts were assigned an efficiency score ranging from 1.000 to 0.152, the lowest. Among the conclusions were that lack of material resources (human resources and computers) could not be blamed for the lower efficiency of some state courts and that all inefficient courts could improve their outputs even with the same input levels. As this is an academic study, it is not clear whether the conclusions or recommendations had any policy impacts.

REFERENCE LIST

- APRI (American Prosecutors Research Institute). 2002. "How Many Cases Should a Prosecutor Handle? Results of the National Workload Assessment Project." APRI, Alexandria, VA.
- Brown, Bernice B. 1968. "Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts", The Rand Corporation, Santa Monica, CA.
- Bulgaria. Supreme Judicial Council. 2017. "Analysis of the Data Collected through the System for Measuring Judges' Caseload." Adopted by a Decision of Bulgaria's Supreme Judicial Council under Protocol No. 38 of 2 October 2017 (in Bulgarian).
- Castro, Alexandre Samy de. 2011. "Indicadores Básicos e Desempenho da Justiça Estadual de Primeiro Grau no Brasil." Report 1609. Instituto de Pesquisa Económica Aplicada (IPEA), São Paulo.
- Contini, Francesco and Richard Mohr. 2007. "Reconciling Independence and Accountability in Judicial Systems." *Utrecht Law Review*, Volume 3, Issue 2.
- Dalkey, Norman and Olaf Helmer. (1963) "An Experimental Application of the Delphi Method to the Use of Experts." *Management Science*, Volume 9, Issue 3.
- Doerner, John, John Douglas, and Suzanne Tallarico. 2010. "Oregon Court of Appeals Judicial and Staff Weighted Caseload Study." State Justice Institute and National Center for State Courts, Williamsburg, VA.
- EWMI (East-West Management Institute). 2011–2013. "Determining and Implementing the Optimal Volume of Work of Judges and Court Clerks and Ensuring the Quality of the Courts' Activity." Romania Analysis and Assessment Report, 2011; Romania Final Report, 2013; Papers 1A, 1B, and 1C on benchmarking exercise. Unpublished.
- Flango, Victor E., and Brian J. Ostrom. 1996. "Assessing the Need for Judges and Court Support Staff." National Center for State Courts, Williamsburg, VA.
- Florida Legislature. Office of Program Policy Analysis and Government Accountability (OPPAGA). 1998. "Information Brief on Weighted Caseload Methods of Assessing Judicial Workload and Certifying the Need for Additional Judges." Report No 97-67, OPPAGA, Tallahassee, FL.
- Gramckow, Heike. 2011. "Estimating Staffing Needs in the Justice Sector." World Bank, Washington, DC.
- Grubišin, Maja. 2015. "Montenegro Case Weighting Study: Final Report with Recommendations.", Podgorica.
- Hriptievski, Nadejda, Vladislav Gribincea, and Jesper Wittrup. 2014. "Study on Optimization of the Judicial Map in the Republic of Moldova." Legal Resources Centre from Moldova, Chisinau.
- Jarzebowski, Sebastian. 2013. "Parametric and Non-Parametric Efficiency Measurement – The Comparison of Results." *Quantitative Methods in Economics XIV* (1): 170–79.

- Kalpakchiev, Kalin. 2016. "Rules for the Evaluation of the Workload of Judges Based on a Model Measuring the Complexity of Cases and Particular Features in the Work of the Judge." Personal blog of the author, Sofia (in Bulgarian).
- Ketterle, Roland. 2013. "Court Financing: The Workload Measuring System PEBBSY", Bucharest, March 11.
- Langbroek, Philip, and Matthew Kleiman. 2016. "Backlog Reduction Programmes and Weighted Caseload Methods for South East Europe. Two Comparative Inquiries." Regional Cooperation Council, Sarajevo.
- Leinhard, Andreas, and Daniel Kettiger. 2011. "Research on the Caseload Management of Courts: Methodological Questions." *Utrecht Law Review* 7 (1): 66–72.
- Leinhard, Andreas, Daniel Kettiger, Daniela Winkler, and Hanspeter Uster. 2015. "Combining a Weighted Caseload Study with an Organizational Analysis in Courts: First Experiences with a New Methodological Approach in Switzerland." *International Journal for Court Administration* 7 (1): 27–36.
- Lombard, Patricia, and Carol Krafka. 2005. "2003–2004 District Court Case-Weighting Study. Final Report to the Subcommittee on Judicial Statistics of the Committee on Judicial Resources of the Judicial Conference of the United States." Federal Judicial Center, Washington, DC.
- Magaloni, Ana Laura, and Layda Negrete. 2001. *El Poder Judicial y su Política de Decidir sin Resolver*. CIDE Working Paper 01. Mexico City: Centro de Investigación y Docencia Económicas (CIDE).
- McMillan, James E., and Carolyn E. Temin. 2011. "Dynamic Case Weighting – Using the Data We Have to Manage the Courts." *The Judges' Journal* 50 (2).
- Mircic, Vucko. 2012. "Weighted Caseload in the Courts of Serbia." Report submitted by the High Court Council Working Group for the Development of a Weighted Caseload System and prepared with assistance from the USAID Separation of Powers program.
- Nissi, Eugenia, and Agnese Rapposelli. 2010. "*Statistica Applicata – Italian Journal of Applied Statistics* 22 (2): 199–210.
- Ostrom, Brian J., Matthew Kleiman, and Cynthia G. Lee. 2016. "Kentucky Judicial Workload Assessment: Interim Report to the Administrative Office of the Courts Kentucky Court of Justice." National Center for State Courts, Williamsburg, VA.
- Pace, Nicholas M., Greg Ridgeway, James M. Anderson, Cha-Chi Fan, and Mariana Horta. 2011. "Case Weights for Federal Defender Organizations." RAND Safety and Justice, The Rand Corporation, Santa Monica.
- Romania. Superior Council of the Magistracy. 2014. "The Efficiency of Courts' Activity: Interim Report." Working Group on the Efficiency of Courts' Activity. (In Romanian.)
- Rosales-López, Virginia. 2008. "Economics of Court Performance: an Empirical Analysis." *European Journal of Law and Economics* 25 (3): 231–51.
- Solomon, Peter. 2012. "The Accountability of Judges in Post-Communist States: From Bureaucratic to Professional Accountability." In *Judicial Independence in Transition*, edited by Anja Seibert-Fohr, 909–36. Springer: New York.

Svensson, Bo. 2007. "Civil and Criminal Justice – Swedish Experiences." Draft report prepared for the World Bank, World Bank, Washington, DC.

Washington State. Office of Public Defense (OPD). 2014. "Model Misdemeanor Case Weighting Policy." OPD, Olympia.

World Bank. 2004. "Brazil. Making Justice Count: Measuring and Improving Judicial Performance in Brazil." Report 32789-BR, World Bank, Washington, DC.

———. 2010. "Uses and Users of Justice in Africa: The Case of Ethiopia's Federal Courts." Report 57988, World Bank, Washington, DC.

———. 2011. "The Malaysian Court Backlog and Delay Reduction Program: A Progress Report." Poverty Reduction and Economic Management Sector Unit, East Asia and Pacific Region, Federal Court of Malaysia and the World Bank, Washington, DC.

———. 2014. "Serbia Judicial Functional Review." Multi-Donor Trust Fund for Justice Sector Support in Serbia. Report 94014-YF, World Bank, Washington DC.

———. 2015. "Mapping the Way Through Court Procedures in Bulgaria." World Bank, Washington, DC.

———. 2017. "Towards Effective Enforcement of Uncontested Monetary Claims: Lessons from Eastern and Central Europe." World Bank, Washington, DC.

Yeung, Luciana, and Paulo Furquim de Azevedo. 2011. "Measuring the Efficiency of Brazilian Courts from 2006-2008: What Do the Numbers Tell Us?" Working Paper 251/2011, Instituto de Ensino e Pesquisa, Brasilia.

