Case weighting in judicial systems

CEPEJ STUDIES NO. 28

As adopted by the CEPEJ in July 2020

Table of content

1.   Introduction

The aim of this study is to review various case-weighting systems (CWS) and provide judicial systems and policy makers with tools to evaluate them and choose which is best suited for their judicial systems. For this purpose, the second chapter will define the term case-weights and list the possible uses of a case-weighting system. In this respect, it is important to distinguish between a CWS, which aims to assess the complexity of cases, and systems designed to assess the performance of judges. In essence, the weight of a case indicates how much more or less time-consuming the case is in comparison to others. This is done by assessing the amount of time and effort each case requires to be processed, on average. This average value is not intended to assess the performance of individual judges in comparison to their peers. Furthermore, we underline the fact that the evaluation of the performance of judges is a complex and sensitive issue, requiring careful handling. According to "international standards" [1], this evaluation should be based on the combination of a qualitative and quantitative analysis of a wider list of indicators than those used in CWS. For this reason, the assessment of the complexity of cases is not to be confused with the assessment of the performance of judges.

The third chapter will provide a general overview of the case-weights practices in member States of the Council of Europe, based on their replies to an on-line questionnaire and supplementary interviews. The third chapter will then proceed to an in-depth review of the case-weights models implemented in the following selected member States: Austria, Denmark, Estonia, Germany, Romania and the Netherlands. As a reference frame, the fourth chapter will examine the case-weights model implemented in the United States. The fifth chapter will then summarise our findings in a comparative manner. Finally, the last chapter will conclude with a list of recommendations. They provide non-exhaustive basic guidelines to any

---

[1] United Nations, Basic Principles on the Independence of the Judiciary, 1985; Committee of Ministers of the Council of Europe, Recommendation No. R (94) 12, adopted 13 October 1994; Council of Europe, European Charter on the Statute for Judges and Explanatory Memorandum, Strasbourg, 1998; Consultative Council of European Judges (CCJE), On Standards Concerning the Independence of the Judiciary and the Irremovability of Judges, CCJE (2001) OP N°1, Strasbourg 2001; European Network of Council for the Judiciary, Dublin Declaration on standards for the Recruitment and Appointment of Members of the Judiciary, 9-11 May 2012, Consultative Council of European Judges (CCJE), On the Evaluation of Judges' work, the Quality of Justice, and respect for Judicial Independence, CCJE (2014) OP N°17, Strasbourg 2014; International Association of Judges, The Universal Charter of the Judge, 14 November 2017.

judiciary seeking to adopt or evaluate an existing case-weights system. In this context, we turn the spotlight on the main building blocks of a successful CWS, one of which is the significant and indispensable contribution of judges as key stakeholders in the process of developing and maintaining a valid and reliable CWS. Simply put, no system will be successful without the full co-operation and active involvement of judges.

2.   Definition and objectives

2.1 What are case-weights?

The term "case-weights" (or "weighted caseload") was coined in the United States of America (U.S.) and dates back to the late 1970s[2]. Case-weights assess the complexity of different case-types, based on the understanding that one case type differs from the other in the amount of judicial time and effort required to be processed.

The weight of a case can be presented in several manners. Traditionally, case-weights are presented using numerical figures representing the average amount of work-time spent on each case-type, from filing until the day it was resolved. In this respect, case-weights do not refer to the disposition time (i.e. the number of days or years that passed since filing until disposition), but to the actual work-time (minutes and hours) spent on all case-related events in each case-type[3]. The term "case-related events" refers to judicial activities that occur in all case-types and require judicial attention, such as:

- studying the case (for example, in preparation for court hearings);
- conducting court hearings (pre-trial, trial and post-trial);
- drafting orders and judgments (including precautionary orders, final and post-judgments, etc.);
- other case-related activities carried out by the judge or by the court personnel.

---

[2] Since then, the National Centre for State Courts (NCSC) has carried out over 100 studies, and more than 35 countries currently rely upon the NCSC weighted caseload formulas. Please note that the CEPEJ Glossary defines Caseload as the "sum of pending cases at a given time plus the incoming cases in a given period". However, when weighting the caseload for the purposes of determining the number of needed judges, the common use of the caseload refers to the number of incoming cases.
[3]Kansas District Court Judicial and Clerk Staff Weighted Caseload Study (October 31, 2011).

For illustration purposes, if the weight of a small-claim case were to be "1", this would mean that this type of case requires an average of 1 hour to complete all case-related activities/events in the life course of that case. Accordingly, if the weight of a class-action case filed in the court were to be 5 hours, this would mean that a single class-action case is five times more time-consuming than a small-claim case. In other words, one class-action case equals five small-claim cases.

Another way to present the weight of a case is to convert the time-units into relative values or to assign each case-type a grade (points) or a benchmark (numerical or alphabetical). For example, the value "1" or "A" can be assigned to the case type that requires the least amount of work-time to process (from filing to disposition) and the weight of all other case-types will be determined in relation to that. However, it is important to note that when case-weights are presented in relative values, grades or benchmarks they do not provide a clear-cut assessment of the amount of judicial time necessary to resolve all the incoming cases in a single year. This can only be achieved through case-weights that estimate or measure the judicial time spent on each case-type.

To sum up, case-weights assign different weights for different types of cases to account for variations in complexity and the need for judicial attention per each case type. By doing so, case-weights indicate how much more or less time-consuming one type of case is compared to other types of cases[4].

2.2 Case-weights objectives and possible uses

Originally, the case-weighting systems were designed to identify the needs of judges. Over the past 40 years the CWS has developed and is used, among others, for:

- determining the needed number of judges, court staff, prosecutors, and/or public defenders;
- supporting funding and budgetary requests;
- allocating justice system personnel within the different work units;
- assigning cases within the courts;
- setting productivity quotas and evaluation standards;
- planning the merger or reduction of work units, etc.

The development of case-weighting systems can also be the basis for further analysis and actions, such as: identifying critical issues in the court procedures, identifying judges' good practices to manage cases and

---

[4] Federal Judicial Center, 2003-2004 District Court Case Weighting Study, 2005.

planning specific programmes for the reduction of the length of the proceedings.

3.   Overview of case-weighting practices in European judicial systems

In order to collect updated data on case-weighting practices in European judicial systems, we designed an online questionnaire consisting of 21 questions. The questionnaire design was reviewed by the working group for judicial time management (CEPEJ-SATURN) in February 2019 and submitted to the member States in March 2019.

The questionnaire was submitted online by the SATURN Secretariat to the National Correspondents of the member States of the Council of Europe periodically involved in the project "Evaluation of European judicial systems".

By the end of May 2019, 35 member States replied to the questionnaire: **Armenia**, **Austria, Azerbaijan, Belgium, Bulgaria, Croatia, Czech Republic, Cyprus, Denmark, Estonia, Finland, Germany, Hungary, Iceland, Ireland, Italy, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, North Macedonia, Poland, Republic of Moldova, Portugal, Romania, Russian Federation, Serbia, Slovak Republic, Slovenia, Sweden, Switzerland, Turkey, Ukraine** and **UK-Scotland. Bosnia and Herzegovina** responded to the questionnaire in March 2020 and the reply was incorporated into this document accordingly. The questionnaire was followed up with interviews of selected countries to better understand and further analyse the answers we received to the questionnaire. Such interviews were conducted with representatives from the following countries: Austria, Czech Republic, Denmark, Estonia, Finland, Lithuania, Romania, the Netherlands and Sweden.[5]

A first draft of this study was sent to the member States to double check the text based on the replies to the questionnaire. Most of the amendments proposed by the member States were included in the current study. A second round of checks by the member States was then carried out in May 2020.

Based on the information collected, it was decided to further review the case-weighting system in six European countries, namely: Austria, Denmark, Estonia, Germany, Romania and the Netherlands. Accordingly, a workshop was organised in Paris on 24 January 2020, attended by experts of the

---

[5] We are grateful to the people who were contacted for this additional information for their fine and kind collaboration.

selected countries, along with the members and scientific experts of the SATURN group. During the workshop each expert presented the method used in their country and discussed it extensively with the members and the scientific experts of the SATURN group.[6] A summary of the presentations and discussion held in the workshop will therefore be included in chapter 3.2.

3.1 Online questionnaire and interviews on case-weighting practices in Europe

In this part of the study, we present the replies to the questionnaire submitted to the member States. In some cases, we were able to double check the replies with interviews while in some others this was not possible. Therefore, please note that this chapter relies exclusively on the information provided by the member States.

23 out of 36 member States reported that they use a weighting system to determine the complexity of each case-type (Q1). However, not all the member States that reported using a case-weighting system fit with the definition stated at the beginning of this study.

Among the ones that did not report on a case-weighing system, **Poland** reported using a "random case allocation system" (SLPS) since 2018, which allows the assignment of cases to the various judges and takes into consideration some pre-established criteria and so provide for a balanced and transparent distribution of the different incoming case-types among the judges.

Belgium reported that its judiciary used such a system in the past (2012-2016) but the results "were not made official" (Q3). Therefore, **Belgium** is included in Table 1, even though the case-weighting system is not currently in place. The **Russian Federation, Slovak Republic, Armenia,** and **Belgium** reported that they plan to use a case-weighting system in the future (Q4). **Montenegro** reported that the implementation of such a system has been postponed until the full deployment of a new electronic case management system.

The following Table 1 shows the replies from the member States on the purposes of the case-weights systems in place (Q7).

---

[6] The expert from the Netherlands could not attend the workshop, but detailed information was gathered through a long phone interview and text exchanges.

Table 1 – Purposes of the case weight systems

State	Estimation of the number of judges needed for each court	Case assignment in the courts	Allocation of judges in the courts	Allocation of new judges in the different courts	Assessment of the court productivity	Other	Assessment of individual judges
Total	**15**	**14**	**9**	**8**	**9**	**7**	**6**
Romania	✓	✓	✓	✓	✓		✓
Azerbaijan	✓	✓	✓	✓	✓		✓
Republic of Moldova	✓	✓	✓	✓	✓		
Ukraine	✓	✓	✓	✓			
Serbia	✓	✓	✓			✓	
Bosnia and Herzegovina	✓	✓	✓[7]		✓		✓
Lithuania	✓	✓					✓
Estonia	✓	✓					
Finland	✓		✓	✓	✓		
Belgium	✓		✓				
Denmark	✓			✓	✓	✓	
Czech Republic	✓			✓	✓		

---

[7] Bosnia and Herzegovina added the following reply in a comment made after the survey was completed: "the court president uses the quota to evaluate if it is necessary to move a judge from one court department to another in order to mitigate the increased caseload of a certain type of cases".

State	Estimation of the number of judges needed for each court	Case assignment in the courts	Allocation of judges in the courts	Allocation of new judges in the different courts	Assessment of the court productivity	Other	Assessment of individual judges
Netherlands	✓				✓	✓	
North Macedonia	✓						✓
Austria	✓						
Croatia		✓			✓		✓
Italy		✓				✓	
Hungary		✓					
Iceland		✓					
Malta		✓					
Turkey		✓					
Sweden			✓			✓	
Germany				✓		✓	
Bulgaria						✓	

As Table 1 shows, some countries gave a multiple answer, with purposes that are more popular than others.

15/24 (62%) of respondents replied: for estimation of the needed number of judges in each court;
14/24 (58%) replied: for assigning cases within the courts;
8/24 (33%) replied: for allocating new judges in the courts;
9/24 (37%) replied: for allocating judges in the courts;
9/24 (37%) replied: for the assessment of the court's productivity;

7/24 (29%) replied: "other" and among them the most common explanation was for supporting funding and budgetary requests.
6/24 (25%) replied: for assessment of the individual judge performance.

Question 5 asked: what kind of methodology is used to assess the case complexity.

The following Table 2 summarises the member States replies.

Table 2– Methodologies used to assess the complexity of cases

State	Points-Based	Work-time Estimates	Other methodology	Work-time Measurement
Total	**11**	**8**	**8**	**5**
Estonia	✓	✓		
Romania	✓		✓	
Finland	✓			✓
Azerbaijan	✓			✓
Hungary	✓			
Iceland	✓			
Italy	✓			
Lithuania	✓			
Malta	✓			
Turkey	✓			
Republic of Moldova	✓			
Netherlands		✓	✓	
Bulgaria		✓	✓	
Austria		✓		✓
Ukraine		✓		✓
Belgium*		✓		
Serbia		✓		
Czech Republic		✓		
Bosnia and Herzegovina			✓	
Croatia			✓	
Denmark			✓	
North Macedonia			✓	
Sweden			✓	
Germany				✓

As table 2 shows, seven member States reported exclusively using a **point-based system** that awards case-types a grade to represent the complexity of the case (**Hungary, Iceland, Italy, Lithuania, Malta, Republic of Moldova** and **Turkey**). We will now briefly discuss the information received from each of these seven member States.

Hungary reported using a system in which "national regulations define case-types and determine the number of points that belong to each case-type". However, based on the answers we received, we cannot elaborate on the method or formula by which the score is assigned to each case-type in Hungary, or on which components are included in that formula. Additionally, Hungary replied that once a sum of points was assigned to a certain case-type and that case-type was enlisted in the IT system of the court, that number cannot be modified afterwards (comments to Q5). This raises the question as to what solution is given to a situation in which the complexity of a certain case-type has changed over the years, and now justifies examination and possible change of the score after it was logged into the IT systems.

Iceland reported using a points-based system in which "the case weight is just a grade of the complexity of the case as a whole" (Q6). However, when asked how many case-types have been graded/weighted. Iceland replied, quite inconsistently, "none". In Q11 Iceland replied that the data collection for the process of deciding the grade/weight of the case is carried out every year. However, it is unclear how the grade of the case is determined, which data items determine the grade of the case and who participates in the process of data collection.

Italy also reported using a points-based system (called *Giada*) to determine the level of complexity of different case-types. However, it seems that this point-based system is used only to balance the number of criminal cases heard in a single hearing.

Lithuania also reported introducing a point-based system and explained that "the process of evaluation of the workload of courts/judges, according to the types of cases, cases were divided on the complexity (factors from 0,3 (e.g. cases in the execution process) till 2,5 (e.g. criminal case)" (Q5 comment). According to the methodological documents we received from Lithuania[8], "the workload of a judge, which shows how many cases and stand-alone pre-trial investigation a judge has examined in the six months preceding the workload assessment, is calculated using the following formula:

---

[8] The original document is available at:
https://www.e-tar.lt/portal/lt/legalAct/c8c70e600dad11e5920c94700bb1958e/asr

$$K= \frac{1}{m} * \sum_{j=1}^{m} \left(\frac{\sum_{i=1}^{n} Vi * fi * ri}{0{,}048 * d} \right)_j$$

K - workload of the judge

m - number of months for which workload is calculated

J - each specific month for which data on cases resolved and independent pre-trial proceedings taken to calculate workload

n - number of cases resolved by the judge and independent procedural action of pre-trial investigation during the calculated month

Vi - each specific case resolved by the judge and each specific independent act of pre-trial investigation carried out

fi - the complexity rate set for each specific case and for each specific pre-trial investigation act carried out

ri - 1 when the case is referred to a single judge or the judge rapporteur

ri - 0,5 in other cases

d - the number of working days worked by the judge

A constant of 0.048 is required in working days within one month

However, Lithuania noted that the weights assigned to the cases do not allow comparison between courts of different jurisdictions (e.g. courts of general jurisdiction with administrative courts), and between courts of different levels (e.g. the courts of first instance with the courts of second instance) (comment to Q8). The weights have been calculated based on data extracted from the case management system *Liteko*.

The **Republic of Moldova** uses a point-based system in which: "The complexity of a case is determined by a fixed component [referred to as the nomenclature of complexity levels] and a variable one. The fixed component refers to the primary subject matter, whose complexity is scored as an integer between 1 and 10. It is set on the merits and remains unchanged during all procedural stages. The variable component refers to the secondary subject matters, the number of parties, the number of the trial bundles, the number of witnesses, the number of passed order, the resolution of a case through mediation, etc." (comments on Q15). According to article 13 of the Superior Council of Magistracy (SCM) decision, if during the proceedings new factors of complexity arise (e.g. more secondary subject matters), the electronic case management system will automatically amend the case weight score. Therefore, there is an initial score, an intermediate score, and a final score (complexity), when the case is disposed by the judge. For example, if the case is disposed by a full judgment, the score of the primary

subject matter is multiplied by 0,20 (20%), if it is disposed by a dismissal order, it is multiplied by 0,75 (75%).

Turkey also reported using a point-based system to have a balanced case assignment in both courts and prosecutor's offices. Several factors are taken into consideration to assign the weights. For example, "in criminal cases, the main factor is the number of parties involved in a case (suspect, defendant or victim). In addition, whether the suspect is arrested, or the victim is dead, are factors taken into account among others. All factors are given points which are gradually risen by UYAP (National Judicial Network System) automatically. For example, UYAP gives 50 points for the first suspect recorded by the system and 25 extra points for each additional suspect. If the suspect is arrested, 70 points would be assigned for the first arrested suspect and 30 for other arrested suspects. Furthermore, UYAP gives extra points if there is more than one offence recorded in a case. "In civil cases, there are clerk offices established in all courthouses to deal with the assignment of cases. These offices are in charge of inserting data relating to cases, such as the number of parties, type of disputes, information on whether legal aid is granted in the case, etc". Then the electronic case management system automatically gives points to the case (Q5 comments).

As shown in table 2 above, 3 States reported that their weighting system **combines a point-based method with a work-time measurement** methodology (**Finland** and **Azerbaijan**) **or with work-time estimates** methodology (**Estonia**). The Estonia case study will be described in the following chapter.

Finland reported in detail a point-based method based on a time-study. The case-types to be weighted in the district courts and courts of appeal were first grouped into 8 case-type categories (coercive measures, conversion of fines, criminal cases, uncontested summary cases, extensive civil cases, land court cases, petitionary matters and insolvency matters). Then each case-group was divided into three levels of complexity (basic, regular, difficult). Court personnel recorded the working time spent on the different case-types using an electronic tool called "Tarmo" in 2009. This analysis allowed the composition of the time needed to process the various cases and then, to assign different weights to the three level of complexity for each case category. For example, the three scores in criminal cases are: 1 basic, 2 regular and 6 difficult. Basic civil cases were considered 2.2 more time-consuming than a basic criminal case. Therefore, a basic civil case has 2.2. weight score, 4.4 for regular cases and 8.8 for difficult ones. The aggregated weights are then calculated by multiplying the number of cases in the various categories with their weight scores.

The court personnel are asked to record their working time annually for a week. So far, the logs have not caused a need to change the case groups, the difficulty categories, or their weighting scores. It is worth mentioning that in Finland "the set of workload indicators is the starting point for assessing resource needs, however, it is not the only grounds for distributing resources". As far as we know, this is the baseline to start a negotiation process between the Ministry of Justice, currently the brand new Court Administration Agency, and the single courts about resources allocation and performance targets.

Seven States reported that their weighting systems rely on **work-time measurement** (**Germany**) or on **work-time estimates** (**Belgium, Serbia, the Czech Republic** and **the Netherlands**) or on a **combination of work-time estimates and work-time measurements** (**Austria** and **Ukraine**). The systems used in Austria, Germany and the Netherlands will be discussed in the following chapter.

Ukraine reported using both methods and gave further explanations in writing about the system in use. For example, in 2016 the case-weights system included the following steps: 1) conducting and analysing a survey among court presidents about the time spent by judges in hearing cases; and 2) analysing the data on the length of case processed through the automated case management system. These two analyses combined helped to calculate the case-weights for different case-types and for different judges. For example, if the decision is taken by a panel, the panel members have a different task in the case decision-making process. This is the reason why it was also important to differentiate the case-weights while taking into account, on one hand, the time spent on processing a case and, on the other hand, the tasks of judges having different case-weights in the same case. The Council of Judges of Ukraine approved the case-weighting system by case category and recommended to the State Judicial Administration of Ukraine their application in courts. The analysis, which was conducted with international technical assistance, aimed to facilitate the role of the State Judicial Administration of Ukraine in allocating judges, cases and making decisions on court funding. In addition, the study findings were also used to ensure the fair distribution of cases among judges through the electronic case management system. A similar analysis was carried out in March 2020 based on the data submitted in 2019.

Serbia classified its case-weights system under "other methodology". Based on the supporting documents available on the Serbian model[9], it seems that

---

[9] Ceretto, J.D., *Weighted Caseload Formula for the Courts of Serbia*, Interim Report (2017).

this model is based on a time-motion study that combines judicial surveys, expert-panel estimations of the time required to complete case-related tasks and an analysis of case-related indicators that influence the complexity level of the case. These case-related indicators include the following: average number of hearings per case-type, average number of documents filed per case-type, average duration of disposed cases per case-type, number of defendants in criminal cases, number of plaintiffs in first-instance labour cases, the major legal basis of the case, etc. The research stages in the Serbian model include: identifying case-related tasks that require judicial work-time in each case-type, isolating case-related indicators that may predict the complexity of the case, sampling cases to extract data on the frequency with which each task accrues in the life cycle of each case-type, determining the average work-time per task based on calculated estimations provided by a sample of judges, multiplying the time value in the frequency value to calculate the weight of each case-type. Finally, the model in Serbia assigned weights to the case-types that consume most of the available judicial work-time. For example, civil, criminal and labour cases in the basic courts constituted only 26.86% of the total 2015 incoming cases. However, the study conducted in Serbia concluded that these case-types demand most of the judges work-time and therefore were awarded weights. The weights in the higher courts were awarded to 5 case-types: first instance civil, first instance criminal, and second instance civil, criminal and labour.

The **Czech Republic** reported using a Time-estimate method, in which a group of expert-judges periodically set up some targets about the number of cases per case-type that are supposed to be solved by the judges, therefore by each court, per year. This approach is quite common in several member States and is usually named "framework criteria" or "quota". The quota however is not something that the judges have to fulfil. It is primary goal is to serve as a case weight. Thus, allowing the calculation of a weighted caseload and estimate of the needed number of judges for the court. They are also testing to measuring the complexity of the case extracting data on the case management system and elaborating them through statistical applications. The current results are informally used to discuss the number of judges needed in the courts between the Ministry of Justice and the President of the courts.

As mentioned before, **Belgium** reported that they tried to use a case-weighting system in the past but the "results were not made official".

Seven States reported using **other methodologies** (**Bosnia and Herzegovina, Bulgaria, Croatia, Denmark, North Macedonia**, and **Sweden**). Romania and Denmark are described in the following chapter.

Bosnia and Herzegovina has a system which is quite typical in South East Europe. Judges are supposed to solve a certain number of cases ("quota") in different ways (e.g. default judgment, sentence on merits, etc.) in one year (Q1). Quotas are targets set up by experienced judges and then approved by the High Judicial and Prosecutorial Council. These quotas take, indirectly, into consideration the time and efforts spent by the judges on producing different types of decisions and allow the assessment of the productivity of each judge at the end of the year based on the number and the type of a decision issued. For example, the productivity quota per year for civil law case-type is as follows: 176 case with decisions on the merits, or 726 default judgment, or 968 other dispositions (Q15).

Bulgaria reported that the weight of a case is determined by a formula combining two elements: the first element is a "weight coefficient" that was assigned to 4 groups of cases (civil, commercial, criminal and administrative); the second element is a "correction coefficient" assigned to cases with higher complexity indicators (cases with above average number of parties, evidence, expert witnesses, numerous causes of action, etc.). The weight coefficients represent the estimated work-time required to complete a case belonging to one of the 4 groups of cases mentioned above. These weight coefficients are expressed by a value greater than 0, with one unit being equal to 16 working hours. The sum of the coefficients of the cases that a Bulgarian judge has examined and completed in one year forms an index of individual annual weighted caseload. It may be further explored that Bulgaria reported using this weighted caseload not for the forecast judges' need or case assignment but, among some other goals, for the assessment of judges, then "to assist in an objective and fair way of the workload of judges in disciplinary proceedings", and to "increase the motivation of judges".

Croatia reported that the weight of a case is based on a pre-defined case-quota that a single judge has to complete in a year. For instance, if a judge is expected to resolve 165 cases of a certain type within one year, then the average weight of that case-type will be 0.006 (1/165).

North Macedonia reported that the complexity of a case is assessed through the use of a list of case-related indicators including (among others): the legal area of the case, the number of parties involved, the type of crime and punishment, the amount of evidence, the monetary value of the case, the estimated number of working hours for processing the case and the involvement of a foreigner in the case. Such a list can and does provide a ranking scale of cases by their level of complexity (from simple to complex). For example, criminal cases which entail a conviction of up to three years imprisonment, and enforceable cases in civil matters are considered simple.

The ranking is different but used in all the courts with civil, criminal, and administrative jurisdiction. Currently, there is a project funded by an international donor being finalised to revise the case-weighting system, taking into consideration the "judge's assessment of the case complexity". The system is supposed to be also used to evaluate judges' performance.

It is also worth mentioning that even though some member States weight their caseload, with different approaches, **they also consider other factors to forecast the number of judges needed** (Q 20).

Armenia, Belgium, Estonia, Hungary, Lithuania, Luxembourg and **Romania** apparently compare data on case filings per judge and then forecast the number of judges needed.

Azerbaijan reported using a "complex approach", with data to forecast the future tendency and population within the jurisdiction.

Bulgaria, Cyprus, North Macedonia and **Slovenia** use data on filed, decided and pending cases.

The **Czech Republic** uses case flow data such as filing, decided, and pending cases, but also the length of judicial proceedings and some other variables that were not specified. "its goal is to find overweight courts and appoint extra judges there".

In **Italy** several indicators are taken into consideration such as: incoming and pending cases, backlogs, cases per judge, incoming cases per 100,000 inhabitants, specific features of the territory of the courts (for example, mafia crime rate, etc.).

Portugal uses an "abstract productivity benchmark" for each judge based on court filings, disposed cases and duration of the proceedings. This benchmark is revised every three years and it is used for the purpose of calculating the number of judges needed in each court.

A similar system seems to be applied in **Slovak Republic,** with a certain "quota" of cases that has to be resolved by each judge. It is our understanding that if the case filings overcome the judges' quota, then new judgeships are considered.

In the **Republic of Moldova,** the following criteria are used to determine the needed number of judges: judges' caseload of the last three years, case complexity, number of inhabitants for court jurisdiction, number of specific cases and other specific criteria that affect the court activity.

In **Sweden and the Netherlands**, a budget is allocated to each court and then the court president or the court board decide how to utilise it so as to reach the established targets, which could also mean the employment of new judges.

In **Switzerland,** the variables taken into consideration to calculate the number of judges per court are: number of incoming and disposed cases per year, disposition time, real length of the proceedings, age of pending cases with specific attention to the number of cases that are still pending after 2 years, number of incoming and disposed cases per judge.

The member States were also asked whether they collect data on the amount of work-time a single (full-time) judge has available to handle case-related work in one year (Q 17-18). The purpose of this, was to isolate the "judge-year" value and the "judge-day" value. The judge-year value represents the number of work-days that a single judge has available for case-related work in one year (365 calendar days - vacation and sick days - training days, etc.). The judge-day value represents the number of hours a judge has available in a single work-day for case-related work.

In this respect, the judge year in **Austria** is set at 1 720 hours per year; in **Azerbaijan** this value is set at 1 152 hours (192 days*6 hours per day); In **Germany**, it is set at approximately 1 700 hours and in **Denmark** it is set at approximately 1 500 hours per year. Three additional States also provided information on the number of annual work-days but did not report the number of hours available per one day for case-related work. These states are **Belgium** with 212 days in one year, **Estonia** with 200 working days per year and **Serbia** with 202 days.

3.2 In-depth review of the case-weighting systems in six European countries

3.2.1 Austria

In **Austria**,[10] the case-weighting system is called "*Personalanforderungsrechnung II*" ("PAR II"). It is based on a major time-study carried out in 2008/2009, whose weights were quite recently amended through a revision procedure made by various groups of expert-judges between 2017 and 2019. The current weights are used for cases under the jurisdiction of the 20 regional courts, which are courts of first instance with

---

[10] Special thanks go out to Ruth STRAGANZ-SCHRÖFL, Ministry of Justice, AUSTRIA, for her assistance in the drafting of this part of the study.

general jurisdiction and as courts of appeal (full time equivalent - FTE - 802 judges) and related public prosecutors offices (360 FTE), and the 115 district courts, courts of first instance with limited jurisdiction (705 FTE judges; based on data analysis from January 2020).

The declared purpose of the case-weighting system is to calculate the number of judges and prosecutors required to manage the demand of justice (incoming cases) in reasonable time and to allocate them to the different courts in line with the principle of immovability of judges.

A first "*Personalanforderungsrechnung*" was implemented in 1994 (for the district courts) and in 1996 for the regional courts. An analysis of files and an ex-post time assessment were then carried out. Afterwards the Ministry of Justice carried out a time-study in 2008 and 2009, triggered again by the Austrian Association of Judges and the involvement of an external consulting firm. The time-study survey involved 11 out of 20 regional courts (63% of all FTE judges of this jurisdiction), 30 district courts (about 35% of the judges of this jurisdiction), and 5 public prosecution offices (over 55% of the prosecutors). Judges and prosecutors were asked to fill in a form with the time they spent on handling the cases, divided on specific case categories. The survey lasted 6 months. The average minutes spent by the judges and by the prosecutors for each case category form the weights for each case category. Considering that a judge's and a prosecutor's working time per year has been set at 1 720 hours, the ratio between incoming cases multiplied by their specific weight and the judge's yearly working time resulted in the number of judges needed to address the demand of justice.

The time for non-case related activities, such as management duties, training, teaching, participation in professional bodies and other specific extra judicial activities, is then added to the case-related judicial activities.

The case-weighting method used has been criticised sometimes due to the fact that at the basis of the time-study, judges and prosecutors alone took down notes about the time they used to work on a case, implying that the time assigned to the different case related activities may be biased by this self-evaluation.

An evaluation of the 2009 system started in 2017. It was decided to proceed with an evaluation due to several factors. These are in particular the long time elapsed since the time-study was carried out, a decreased number of district courts, substantive amendments in criminal law, and a general impression that the complexity of cases was raised. The evaluation was carried out by several judges' working groups divided by case categories, which were coordinated by a Steering Board within the Ministry of Justice.

The Steering Board consisted of members of the four Regional Courts of Appeal, the Association of Judges and of the Ministry of Justice.

One of the most important factors in the assessment was the involvement of practitioners and the transparency of the whole process finalised to a wide acceptance of the new weights.

The working groups' members were first and second instances judges, public prosecutors and members of the respective associations. The weights were amended starting from the time-study carried out in 2008-2009 on two principles. Firstly, the calculation was revised taking also into account other data collected during the time-study but that were not used. Secondly, 8 sets of expert groups (the members of which were first and second instances judges or public prosecutors but were not part of the working group) made a time-estimation through a Delphi method.

Among the 84 different case categories, 44 remain with the original weights (minutes), 26 were further evaluated using a calculation model developed by the Austrian data processing centre, and 14 were readjusted, using the calculation model and the Delphi method. This is the reason why the methodology used can be considered based on both a time-study and a time-estimate, with this latter used to validate and adjust the results of the initial time-study. The time estimation took into account the need to add some time in processing the case if one or more of the following factors occurred in the case: translator, expert witness, foreign context, contentious matters, juvenile involved in a criminal case and suspect in custody. For some case-categories, a detailed business process was prepared to assess the various tasks within three broad phases: pre-trial, trial and post-trial. The agreement on the time estimation was built on a detailed methodology assessment based on a Delphi with final weights that progressively were agreed upon by the judges in the working groups.

At the time of writing the study, the Ministry of Justice debated some further issues. For example, the calculation is based on the incoming cases and it does not take into consideration the productivity of each court; the time-study carried out in 2008/2009 proved not to be very reliable for the public prosecutors as it was processed during an important change in the code of criminal procedure. It therefore fell short as a basis for the following time estimation. The local characteristics of the various courts are not sufficiently taken into consideration in the average minutes calculated to process a certain case category. As a consequence, the case-weights can be affected by this variety and they could not reflect the time used to process a case across the country.

The combined method used in Austria is very interesting because it is a practical example of the combination of a time-study and a time-estimation by expert-judges. According to the Austrian expert in charge of the development of the case-weighting system, the time estimation carried out by the expert groups with a well-established methodology can be considered a reliable alternative to the more expensive and cumbersome time-study.

3.2.2 Denmark

In **Denmark**[11], case-weights were established in 2002, implemented in 2005 and then dramatically revised in 2011. The featured case-weights are utilised in both the 24 district courts (court of general jurisdiction). The two high courts have a different set of weights but the basic methodology to establish the weights is the same. In 2002, a limited time-study was conducted on the time needed to resolve cases. This study was used as a basis for a more qualitative assessment carried out by three working groups of judges. These working groups were divided according to the following legal areas: 1) civil and criminal cases, 2) enforcement cases, and 3) insolvency cases. They were entrusted with adjusting the initial case-weights values.

In 2011 a major revision took place. The idea was to have a more transparent and simple method, widely accepted by the judges. The revisions included: a decrease in the number of case-categories, changes made in the weights of some of the case categories (for example, the criminal case categories weighted decreased from 32 to 7), and the type of judgment, which was considered in the "qualitative assessment" in 2002; it was excluded in 2011.

During the time-study, the court staff was required to register daily how they spent their workday on various case-types. In some of the district courts, the appointed judges do it half-yearly (side answer Q5) in a kind of "lump-sum" self-report of the time they spent on case-related activities and non-case-related activities (such as managerial duties and other activities). However, this is not related to a time-study but to the productivity benchmarking procedure.

The so called "primary weights" were calculated based on the time-study findings, and then increased by a multiplicator of complexity, considering specific case-related indicators that indicate a higher level of complexity, as well as a qualitative evaluation of the basic average weight of each case-type regardless of case-related indicators. These statistical case-related

---

[11] Special thanks go out to Frederik Ellesøe RASMUSSEN, Court Administration, DENMARK, for his contribution to the drafting of this part of the study.

indicators include the following: prolonged hearings that last more than one day, the use of interpreters and the number of judges involved in the case.

Resolved weighted cases are then used to assess the court productivity, and to allocate the budget to each court, and within each court to the various divisions.

The Danish court administration is aware of the limit of the current time-study which does not imply a direct registration of the time needed to deal with a case. From time to time, there are minor adjustments of specific case-weights based on a qualitative evaluation of some changes in the procedure or in the law. However, as mentioned, these primary weights are amended with complexity factors during the life cycle of the case, and weights are mainly used for the allocation of the budget. For these reasons, to have a detailed, usually quite cumbersome and expensive, time-study is currently considered not a priority, since it may not improve so much a consistent and equitable allocation of the budget.

"It is always a question how detailed a weighting system should be. A basic principle for the Danish Court Administration is that we stop to make it more detailed, when a higher level of details does not really make a difference to the end result"[12].

The goal of case-weighting in the Danish court administration is mainly to allocate the budget to the various courts. Also, for this reason, the case-weights are calculated for both judges and other court staff, and the budget is allocated accordingly. For example, the case-weights of a criminal hearing case are 10 points for both judges and court's staff, due to the fact that both are participating in conducting or preparing the hearing. However, due to the different salary level, the score used to allocate the budget to the various courts is 6,3 for judges and 3,7 for the court staff. The case subject factor is unique for each case subject. In the specific case of criminal cases, each personnel group spends close to an equal amount of time on an average criminal case, however, the salary difference is approximately 2 to 1 for judges relative to court staff, which is why the above mentioned weights for criminal weights are adjusted accordingly for each personnel group to be economically comparable. In the example of estate cases, the court staff spends about 20 times the amount of time on an average estate case as judicial employees. However, because the salary difference is approximately 2 to 1 for judges relative to court staff, the actual case subject factor for estate cases is approx. 0,20 for judicial employees and approx. 0,8 for court staff,

---

[12] Court Administration, DENMARK, *Memorandum about the Weighting of Cases in the Danish Court System*, 21 September 2015.

if the weights are to be made economically comparable. The specific case subject factors are limited below:

Case subject factor	Judicial employees	Court staff
Civil	66%	34%
Criminal	63%	37%
Enforcement claims	30%	70%
Forced sales	42%	58%
Insolvency	68%	32%
Estate	20%	80%
Notaries	6%	94%

After the weights for each case subject have been adjusted by their related case subject factor, then a combination factor is applied, which multiplies the weight by specific factor. This is done in order to make a weight point of a civil case comparable, for example, to a weight point of a criminal case.

Combination factor	Weight
Civil	4,06
Criminal	1,48
Enforcement claims	0,37
Forced sales	1,33
Insolvency	2,63
Estate cases	1,24
Notaries	0,10

The budget is allocated to each court taking into consideration two factors: the average weight of resolved cases (40% of the budget), and the number of incoming cases of each court, in order to have a plausible forecast of the expected activity of the court. This "expected activity" of the court is also used to allocate judges to the various courts, respecting the principle of immovability.

3.2.3 Estonia

Estonia[13] started to develop a case-weighting system in 2007 to ensure access to justice, reasonable timeframes, equal case distribution among judges, and to assess the time required to solve cases to explore the possibility to have more specialised judges to cope with an increase in incoming cases.

---

[13] Special thanks go out to Kaidi LIPPUS, Ministry of Justice, ESTONIA, for her contribution in the drafting of this part of the study.

For these purposes, working groups of judges assessed the time needed to solve different types of civil, criminal, and administrative cases. This estimation tried to take into consideration several factors that, in the judges' experience, may affect the time spent to solve the various cases. For example, in civil matters a counterclaim was considered as such a factor, and in criminal matters such factors were: the number of accused, the number of charges, and the need of an interpreter.

These factors were then researched using a time-study which lasted for 1 year and half. During this time-study, judges filled in timecards for each case solved. It was also calculated, taking into consideration vacation time, illness, training and institutional events, that the yearly case related working time for a judge is 1 600 hours.

At the end of the process, three different case-weighting scale were adopted:

In administrative matters the method to weight cases was different. Judges agreed on a scale ranging from 0,34 points, which were awarded to the least complex and time-consuming cases, up to 10 points that were awarded to the most complex and time-consuming cases (such as tax matters).

In criminal matters, judges agree to set a "reference for a standard criminal case", which estimates 31,1 hours to solve a case with one accused and no more than five counts, without interpreters. Then if the number of accused is higher and the counts are numerous, this standard time is multiplied for various coefficients. Other simplified proceedings were set at much lower time requirements such as 2,3 hours for summary proceedings, and 8,9 hours for penal orders. The criminal scale was very wide, starting at 1,2 estimated time for the simplest case, up to 513,2 hours for the most complex ones.

In civil matters the key factor considered to set the time scale was the kind of judge's decision to solve a case, which can be a ruling, default judgment, or full judgment. The scale spanned from 1,5 hours for the simplest cases, which need just a ruling, to 11,5 hours for the most complex ones. Family matters, which have more than one claim (for instance, rights to access of a child and a child support claim together in one case), were considered more complex and therefore a coefficient of 1,3 was added to the simple family matter.

In 2015 this approach, which was adopted only by the administrative judges, was revised, also thanks to a check of the real length of the different judicial proceedings extracted from the electronic case management system. The revision in administrative matters adopted a scale of complexity points from

2 to 40, where each point was valued at about 0,67 hours of administrative judge's time. The calculation was done in the following way. The number of cases solved for each category in the last 4 years was multiplied by the corresponding score (weight) of complexity. For example, if 100 cases of the category general procedure/data protection and public information were solved in the last 4 years, it was multiplied for the corresponding score (i.e. 20), and then added to the other solved case category multiplied for their relative score. 1 600 working hour per judge were multiplied for 4 (years) and for the number of full-time equivalent judges in service in the last 4 years.

The ratio between the total judges' working hours and the sum of the cases solved multiplied by their score of complexity results in the administrative case weight point value of 0,67 hours.

In civil matters it was also agreed about a 40-point scale, but the calculation (made using the same formula as in administrative cases), of the case weight point value ended to be 0,27 hours. This value was challenged by the judges who thought it was not realistic, also in comparison with the 0,67-value calculated for administrative cases.

In criminal matters, the amount of time estimated for the most complex cases was just increased.

In 2017 the need of a more precise case-weighting system was revamped. A working group with presidents of the courts was established by the Ministry of Justice to revise the current system and make a better estimation of the time needed to solve different case categories. Civil judges were not able to reach an agreement of the corresponding time spent for each case category. Therefore, they just listed the scores without relating them to the time spent for each case listed.

Currently, the case-weights are used by the administrative courts, and by 2 out of 4 first instance county courts. The case-weighting is mainly used for case assignment and for proposing the number of judges who should be allocated to each court or to the different branches within the same court. Some weights were also proposed for the second instance courts but they have never been fully adopted.

During the interview, it was also pointed out that, based on the statistical data, currently there is not such a practical need of a case-weighting system for a small size court (2-4 judges), because the specialisation opportunities are low and the small number of judges serving in the court cannot be decreased. However, Estonia is considering adopting a state-wide case distribution system by certain case categories. Therefore, the need for case-

weighting is more and more in demand to assess comparatively the caseload of the judges.

The judiciary in Estonia is right now in the process of revising the case-weighting methodology in order to work out the changes needed to adopt it in all the courts in a consistent manner so as to make comparable the weights among civil, criminal, and administrative cases.

3.2.4 Germany

Germany[14] completed its first time-study in the year 2002 and updated it in the year 2014. The 2014 study involved a representative sample of 16 163 participants from 70 courts and prosecutor's offices in 14 federal states (Länder). The total number of participants included: 5 057 judges and judicial officers, 8 233 non-judicial staff members and 2 855 officials from the prosecutor's office of the different jurisdictions.

All participants were asked to track and report the amount of time they devoted to different work activities during a period that lasted 6 months from 1 January to 30 June 2014. For this purpose, the German research team designed a manual reporting form (a time-log) that was pinned to each case handled during the six-month data collection period. To illustrate, if a judge devoted 30 minutes to a case-related activity in a specific case, he or she only recorded the number 30 in the time-log form pinned to that specific case. In this respect, it should be noted that the judges expressed fears that the time-study might jeopardise the independence of the judiciary. This fear was put to rest thanks to the following steps taken: anonymisation of the data set, and a commitment that the data set will not serve for individual performance evaluations nor for comparisons between courts. It was then well established that the aim of the study was to quantify the average working time per case-type as a frame of reference.

At the end of the 6 months' time-study, the research team gathered all the time-logs and calculated the total amount of minutes recorded in the time-study per each case-type category. In this respect, it should be clarified that a case-type category can group together different case-types that differ one from the other in their legal claim but require a similar amount of work-time to process and resolve. The case categories in Germany are defined by a working group of the Federal States, consisting of delegates of the justice ministries in the Länder (not including judges). In their decision, the working

---

[14] Special thanks go out to Andreas NEFF, President of the District Court in Freiburg, GERMANY, for his contribution to the drafting of this part of the study.

groups members take into account the expected time to process and resolve each case. For example, the case category in Germany - RL011 - includes the following 5 different case-types: medical malpractice cases, construction/architectural cases, personal liability and fee claims, company disputes and cartel cases.

Based on the time-study results and taking into account the "law of large numbers" it was determined that these 5 case-types require, on average, a similar amount of work-time (ranging from 1 058 to 1 815 minutes per case) and therefore they were grouped under one case category and assigned with the same case-weights (1 193 minutes). The case-weights in the German methodology is called "basic-number", also referred to as a "Pebbsy Number"[15]. The calculation of the pebbsy-basic number builds upon a preliminary formula that calculates the number of cases processed during the time-study per each case category (*"RechnerischeMenge"*). This preliminary formula is shown below:

$$Rechnerische\ Menge = \frac{(t_{Typ\ E} * M_{Typ\ E}) + (t_{Typ\ A} * M_{Typ\ A})}{(t_{Typ\ E} + t_{Typ\ A})} + M_{Typ\ V}$$

t= number of minutes reported in the time-study
M= number of cases handled during the time-study
TypE= cases filled before the time-study began but resolved during the time-study
Typ A= cases filled during the time-study but resolved after the time-study ended
Typ V= cases filled and resolved within the time-study period

To clarify, the time-study also collected data on *Typ L* cases, which are "long running" cases that were filled before the time-study began and resolved after the time-study ended. However, it was decided that the time-data on *Typ L* cases (M *Typ L*) will only be included in the pebbsy-basic-number formula that is shown below, and not in the preliminary formula. In addition, it was decided not to include the variable of "number of cases handled during the time-study" in regard to *Typ L* cases in neither formula.

The result of the preliminary formula is the denominator of the pebbsy-basic-number formula, as can be seen in the following:

---

[15]Pebbsy stands for "Personalbedarfsberechnungssystem", meaning: Personnel Requirements Calculation System.

Total minutes reported in the time-study per all the cases belonging to the same case category

---

The calculated number of cases processed in the time-study period (*RechnerischeMenge*)

To clarify, the numerator of the formula above also includes the number of minutes reported in the time-study in regard to *Typ L* cases. *Typ L* cases are "long running" cases that were filled before the time-study began and resolved after the time-study ended. As shown above, the *Typ L* cases are not included in the preliminary formula but only in the numerator of the pebbsy-basic number formula

The main purpose of the pebbsy-basic numbers is to calculate the required number of judges in the different level of jurisdictions in Germany, based on the following formula:

Number of incoming cases * relevant pebbsy-basic numbers

Annual judicial work-time available to handle cases (judge positions* daily working hours* annual workdays)[16]

The pebbsy-basic numbers are revised periodically by the Commission of the Judicial Administrations of the Federal States for calculation of staff requirements. In the future the plan is to update the pebbsy-basic numbers every 10 years while taking into account circumstances that might justify an update to be made before the 10 years have passed. Such reasons for updating the pebbsy-basic number included: changes of the law, new case law, structural reforms and modernisation measures.

As a final point, it should be noted that the pebbsy basic numbers represent the average processing time of different case-type categories. However, the processing time of an individual case (or case-type) can deviate considerably from this. Therefore, the Pebbsy system is not suitable for the following uses: distribution of cases within the court, nationwide benchmarking and drawing conclusions on the working patterns of the individual staff member and alignment of the forms of organisation in the Federal States.

---

[16] The value of the annual judicial work-time available to handle cases is determined per each federal state, separately. This value ranges around 1700 hours per year.

3.2.5 Romania

Romania[17] assesses the complexity of different cases using a grading scale that awards each individual case a sum of points to represent its level of complexity. In this respect, we emphasise that unlike countries that assign the same weight to all the cases classified under the same case-type, the Romanian system calculates the complexity grade of each individual case separately.

This grading scale was first integrated into the Romanian case management system (ECRIS) in 2003 and implemented in 4 pilot courts. Initially, the administration module of the ECRIS allowed each of the courts to assign complexity grades independently, on a local level, and set that value in the system (or modify the existing value). The side effects of this decision were the wide variance between the courts which ruled out the possibility of comparing the weighted caseload of different courts and of drawing conclusions on the number of judges needed to handle the incoming cases.

As a result, the Supreme Council of Magistracy (SCM) decided in 2005 to form ad hoc working groups that were entrusted with: establishing a nationally unified complexity scale, determining the appropriate workload per judge and setting the timeframes to process and resolve different case-types. The working groups included participants from 18 courts (6 district courts, 6 tribunals, and 6 courts of appeal). For this purpose, 570 cases were analysed (300 district courts cases, 180 tribunals cases and 90 courts of appeal cases) and conclusions were drawn in regard to 50 different case-types (both criminal and non-criminal). To our understanding, the working groups participated in some sort of a time-study to record their work-time per case. However, we do not have information on the duration of the time-study, the number of participants, and the manner of record (per judicial activity or per case as a whole/ consecutive measurement or fragmented etc.). These working groups concluded their work in 2006 and based on their conclusions the SCM adopted in 2007 a decision to unify the complexity scale on a national level.

Nowadays, the calculation of the complexity of each individual case is done by the ECRIS algorithm based on the following components:

1. The initial complexity grade awarded to the "main object of the case" and to the "secondary object of the case". To clarify, the main object of a case is the case-type or cause of action. Accordingly, the

---

secondary object is the sub-categorisation of the case-type. The complexity level of each of these objects (primary and/or secondary) was graded on a scale from 1 to 10, as a constant. To our understanding, these constants were based on the conclusions of the working groups convened in 2005-2006 and has not been changed since. Beyond that, we have no information on how the members of the working groups formulated the scale and graded each object (primary and/or secondary).

2. Number of parties involved in the case - this variable distinguishes between cases involving up to 4 parties and cases involving 5 parties and above.

3. Number of tomes (volume of the case).
   ▪ Number of witnesses exceeding 10 per case - this variable was included in the formula starting in 2016 but has yet to be implemented due to technical difficulties in the ECRIS algorithm.

The relationship among these variables is shown in the following formulas. For example, in cases that involve 4 parties or less, the ECRIS algorithm will calculate the complexity level of the individual case as follows:

[(complexity of the main object)]
+
[(aggregate complexity of secondary objects) x 10%]
+
[(complexity of main object) x (number of tomes) x 20%]

In cases that involve 5 parties and above, the ECRIS algorithm will calculate the complexity level of the case as follows:

[(complexity of main object)]
+
[(aggregate complexity of secondary objects) x 10%]
+
[(complexity of main object) x (number of tomes) x 20%]
+
[(complexity of main object) x (number of parties) x 5%]

In the future the ECRIS algorithm will take into account the variable of number of witnesses exceeding 10 per case, and calculate the complexity level of the case using the following formula:

$$[\text{the complexity of the main object}]$$
$$+$$
$$[(\text{the complexity of the secondary objects}] \times 50\%]$$
$$+$$
$$[(\text{the complexity of the main object}] \times (\text{number of parties}) \times 10\%]$$
$$+$$
$$[(\text{complexity of the main object}) \times (\text{number of witnesses exceeding the minimum of 10}) \times 10\%]$$

3.2.6 The Netherlands

The **Netherlands**[18] also use self-reports of judges and other judicial officials on their work-time use. However, the method of data collection as well as the manner, in which the final case-weights are formulated in the Netherlands, are different.

In the past, the Netherlands used the "shadow research" method, in which law students accompanied ("shadowed") selected judges for a period of time to collect data on their working hours per different judicial activities. Over time, the method of data collection changed and now relies on a combination of two methods: the first is the Work-sampling method, and the second is the time-estimates provided by a group of expert-judges based on the work-sampling results.

The Work-sampling method is a statistical technique for determining the proportion of time spent by employees on various types of activities. This is done by asking the employees at random time points to record their current activities without the need to record the duration or start and end times of those activities. Based on the Dutch study, this method works well if there is a sufficiently large number of participants that provide a sufficiently large number of moments, representative of the various types of activities performed by the participants. This technique was implemented in the Dutch judiciary in the 2014 case-weights study and in the 2017 study that followed.

The 2017 study included 1 859 participants that were randomly sampled out of the general population of approximately 5 100 judges, judicial officials (RA or GA), councellors and judges in training or legal support staff. To this end, an outside consulting firm was hired to manage the data collection using an application (app) that was installed on the participants smartphone or tablet.

---

[18] Special thanks go out to Jos PUTS, Judicial Council, the NETHERLANDS, for his contribution to the drafting of this part of the study.

The data collection via the time-app lasted a single research week per each participant, from Monday to Sunday, between 7:00 AM and midnight. Prior to the research week, the participants received a short intake questionnaire. The purpose of this questionnaire was to gain insight into: the standard working week of each participant, the availability of each participant during the research period at the relevant court (different research weeks were assigned per each court), and the access of each participant to a device that meets the technical requirements for installing the time-app (a smartphone or a tablet).

Based on the replies to this questionnaire, the participants were assigned a specific research week that aligns with their schedules as much as possible (for example, not all cantonal judges were assigned the same research week). It was also known which participants needed a loan tablet or smartphone for the duration of their research week.

To ensure the successful data collection, the research team conducted a training session in different courts (23 training sessions were held in total). In addition, the participants received an extensive manual on the app with an explanation on how to use the iPad/ smartphone/ tablet for the non-experienced user. Additionally, a short instructional video was prepared as well as a guidebook with answers to frequently asked questions regarding the research effort.

Once the app was activated by each participant, the participant was asked to indicate which type of activity he or she is performing at that moment. In response, participants could choose one out of the three following options:

1. "I am not working" (I am not performing a judicial activity right now);

2. "I am performing a non-case-related activity which is..."(selection from a pre-defined list of non-case-related activities such as: participating in work meetings, professional training and education, management and administrative duties, etc.);

3. "I am performing a case-related activity in case-type ....."(selection from a pre-defined list of case-types that were later grouped to seven legal areas/categories in the first instance (criminal, family, immigration, tax, trades, administrative, and small claim cases) and to four legal areas in the courts of appeal (criminal, family, tax and trades)).

The time-app repeated this question 12 times, in random points in time, during each day throughout the research week of each participant. The points in time in which the question appeared on the app differed per each

participant and per each day of each participant. As a result, each participant provided information on his or her time-use in 84 different and random time points during the research week (reliability of 93.3%, with a maximum margin of error of 10%).

The duration of the data collection was 61 research weeks spread over 41 calendar weeks from January to December 2017. The first research week took place in week 3 of January (16 to 22 January) at the Oost-Brabant Court and the Den Bosch Court, and the last was in week 50 (11 to 17 December 2017) at the Amsterdam Court. To our understanding, the final database was anonymised so that the identities of the participants will not be revealed.

At the end of the data collection period, the consulting firm calculated the proportion of time spent on non-case-related activities (22-30%), and the remainder was attributed to case-related activities. This remainder was calculated per each of the 10 categories/legal areas mentioned above. For example: if 5% of the measurement moments was recorded as "I am working on a criminal case", then it was concluded that "5% of the available work hours per FTE is dedicated to criminal cases". In this respect, the available work hours per FTE in the Netherlands was set at 1 821,6 annual hours in 2017 (253 working days * 7,2 hours per work-day). This figure was then multiplied in the proportions of the measurement moments per each of the ten categories as mentioned before.

As a final stage, the aggregated weight of each category was subdivided into the different case-types included in each category. This was done by 10 focus groups of expert-judges from the first and second instance (one group for each legal area). Every focus group met once and involved 11 participants from the first instance court and 4 from the court of appeal. It should be noted that all courts were invited to participate in the focus groups and that most courts sent a judge and some sent a judicial assistant. In addition, the focus groups also involved operational management persons.

Prior to the focus groups meetings, the group members received working documents that included: the results of the work-sampling, statistics on the number of cases handled in the year 2017 divided to case-types in each court, etc.

Based on their professional experience, their conclusions from the working documents and the group discussion held during the group meeting, the group members then divided the aggregated weight of each category into 6-8 case-types, thus formulating the weight of 70 different case-types. In future studies it is intended to determine the average data not only in the national level, but also locally, for each court.

4.   Review of the weighted caseload model in the United States courts

The weighted caseload model used in the United States courts consists of three elements:[19]

1. Case filings, meaning the number of cases filed to the courts during one year, and counted per each case-type separately;

2. Case-weights, which represent the number of minutes that a judge or a judicial officer is required to invest in each case-type, on average;

3. The judge-year value, meaning the number of minutes each judge or judicial officer has available for case-related work in one year.

The case-weights, or the average work-time per case-type, relies on self-reports or self-documentations collected from judges. These self-reports and/or documentations can be collected using one of two methods of data collection (or a combination of both): *the Delphi method* and *the Time-study method*. The unique characteristics, as well as the strengths and weaknesses of each of these methods, will be discussed in detail in the next chapter.

In a nutshell, the Delphi method is "a survey technique to gain consensus knowledge by questioning a panel of experts in multiple rounds"[20]. In the context of the case-weights studies in the United States, the Delphi method is used to collect calculated estimations of the average amount of work-time required per case, based on the self-reports of judges who were selected to participate in a focus group. In comparison, the time-study does not rely on estimations of work-time or on one-time/repetitive surveys. Instead, the time-study method relies on continuous and real-time self-documentation of the judges' work-time per case. Simply put, in each day of the time-study, the judges are required to document the duration of work-time spent per every

---

[19] Matthew Kleiman, Richard Y. Schauffler, Brian J. Ostrom & Cynthia G. Lee (2019) Weighted caseload: a critical element of modern court administration, International Journal of the Legal Profession, 26:1,21-32; Matthew Kleiman et al., Case Weighting as a Common Yardstick: A Comparative Review of Current Uses and Future Directions, vol.7, n.4 (2017) (Onati Socio-Legal Series, n.d).

[20] Mehdi Mozuni and Wolfgang Jonas, An Introduction to the Morphological Delphi Method for Design: A Tool for Future-Oriented Design Research, She Ji: The Journal of Design, Economics, and Innovation 3, no. 4 (2017): 303–318, doi:10.1016/j.sheji.2018.02.004.

single case that they handled that day. Furthermore, in time-studies that are more detailed, judges are required not only to indicate the type of case they are working on, but also the type of judicial activity they performed in that specific case.[21]

Until the early 2000's, the case-weights studies in the United States utilised the Delphi method to collect self-estimations of the average work-time judges spend on each case-type. However, since then it seems that the majority of the updated case-weights studies now utilise the time-study and consider it to be the most accurate method to determine the weight of the case. With that said, it is important to note that these studies still use the Delphi method, but only as a research tool aimed to evaluate and validate the findings of the time-study. As a frame of reference, table 3 in the following page will review the participation rate of judges and the duration of the time-study in 10 case-weights studies published in the United States between 2013-2018.

Table 3: the participation rate of judges and the duration of the time-study in 10 CW studies

Country	Participation rate of judge	Duration of time-study	Publication year
California[22]	98%	4 weeks	November 2018
Iowa[23]	94%	4 weeks	June 2017
Oregon[24]	96,9%	4 weeks	May 2016
Florida[25]	97%	4 weeks	May 2016
Indiana[26]	99,3%	4 weeks	April 2016
Kentucky[27]	94-96%	4 weeks	January 2016

---

[21] Brian.J. Ostrom et al., *Florida Judicial Workload Assessment* Final Report (National Center for State Courts, 2016).

[22] Judicial Council of California, *Judicial Workload Assessment: 2018 Judicial Workload Study Updated Case* weights, Report to the Judicial Council, 2019.

[23] Suzanne and John Douglas., *Iowa Judicial Officer Workload Assessment Study, 2016*, Denver, National Center for State Courts, Court Consulting Services, 2017.

[24] S. Tallarico & A. Davis, *Oregon Circuit Court Judicial Officer Workload Assessment Study,* National Center for State Courts, 2016).

[25] B.J. Ostrom, M. Kleiman, C.G. Lee & S. Roth, *Florida Judicial Workload Assessment Final Report,* National Center for State Courts, 2016.

[26] D. Sayles, S. Tallarico and J. Douglas, *Indiana Caseload Assessment Plan to Utilize Resources Efficiently*, National Center for State Courts, 2016.

[27] B.J. Ostrom, M. Kleiman& C.G. Lee, *Kentucky Judicial Workload Assessment*, National Center for State Courts, 2016.

Country	Participation rate of judge	Duration of time-study	Publication year
Vermont[28]	100%	12 weeks	August 2015
Montana[29]	91,8%	8 weeks	October 2014
West Virginia[30]	100%	9 weeks	September 2014
Tennessee[31]	95,3%	6 out of 11 weeks	October 2013

As shown in the table above, the participation rate of judges in the selected states ranges between 91,8% to 100% with an average value of 96%. Additionally, the duration of the time-study in those studies ranges between 4 weeks to 12 weeks with an average value of 5,9 weeks.

Following the completion of the time-study, the initial weights in the U.S. model are determined using one of two formulas. The first is the "events-based" formula and the second is the "annualised time-study minutes reported per case-type category" formula.

The "events-based" formula relies on the following four steps:

i.   Identifying the case-related events that occur in all case-types and require judicial work time. In this respect, the formula should include only events that are applicable to a large proportion of case-types and courts. In other words, specific or unusual events should not be taken into account (for example, different work habits and styles of individual judges and special circumstances, etc.).[32]

ii. Extracting state-wide data from the case management system, or from hand-sampled cases, to determine the average frequency with which every event occurs in each case-type.

[28] S. Tallarico, E. Friess & J. Douglas, *Vermont Trial Court System Judicial Officer and Court Staff Weighted Caseload Study*, National Center for State Courts, state justice institute, 2015.

[29] S. Tallarico, J. Douglas & E. Friess, *Montana District Court Judicial Weighted Caseload Study*, National Center for State Courts, State Justice Institute, 2014.

[30] J. Douglas, S. Tallarico, E. Friess& W. Wills, *West Virginia Circuit Judge Workload Study*, National Center for State Courts, State Justice Institute, 2014.

[31] S. Tallarico, J. Douglas & E. Friess, *Tennessee Trial Courts Judicial Weighted Caseload Study*, National Center for State Courts, 2013.

[32] V.E. Flango & B.J. Ostrom, *Assessing the Need for Judges and Court Support Staff,* National Center for State Courts, 1996.

iii.   Collecting self-reports/documentation from judges on the amount of work-time they spend handling each event in each case-type (either by Delphi method or Time-study method).

iv.   Multiplying the event frequency by the average time required for judges to handle the events leading to the disposition of the case in each case-type.

The following table will demonstrate the events-based formula on two case-types: felony case and divorce case.[33] To clarify, the work-time in this numerical example is the median value of minutes per event.

Table 4: The events-based formula for felony and divorce case

Event type	Felony case			Divorce case		
	Work-time per event	Event Frequency	Event weight	Work-time per event	Event Frequency	Event weight
Initial Appearance	5	1,05	5,25	30	0,05	1,5
Preliminary Hearing	17	0,63	10,71	---	---	---
Arraignment	7	0,64	4,48	---	---	---
Scheduling/Pre-trial	15	0,03	0,45	15	0,44	6,6
Pre-trial Hearing/Motions	15	1,83	27,45	30	0,49	14,7
Default Judgment/ Plea Acceptance	15	0,85	12,75	15	0,47	7,05
Court Trial	47	0,01	0,47	30	0,36	10,8
Jury Trial	480	0,05	24,00	---	---	---
Post Judgment/ Verdicts	15	0,18	2,70	20	0,05	1,0
Disposition/ Sentencing	18	0,73	13,14	20	0,61	12,2
Bench Warrant	5	0,39	1,95	3	0,02	0,06
Appeal/ Review	5	0,33	1,65	30	0,05	1,5
Initial weight in minutes per Felony case			105	Initial weight in minutes per Divorce case		55,4

The obvious advantage of the events-based formula lies in the fact that it allows the researcher to break down the total weight into events, identify the most time-consuming event, compare the time-consuming events among different case-type and draw conclusions based on that. By doing so, the events-based formula promotes a better understanding of the workload entailed in each case-type. However, the downside of this formula is that it requires a research effort to collect data on the frequency of each event per

---

[33] Ibid.

each case-type. Such a research effort might prolong the duration of the full study in 3-4 months due to the time needed to complete the collection and analysis of the frequency data. The alternative formula is the "annualised time-study minutes reported per case-type category" which relies on the following components:

i. The aggregated total time-study minutes, reported by all participants, per each case-type;

ii. The number of work-days out of the total time-study period;

iii. Total judicial working days available for case-related work per year;

iv. The number of case filings in the previous year per each case-type.

The table below will demonstrate the calculation of the weight for a simple misdemeanour case in Iowa[34].

Table 5: The "Annualised time-study minutes reported per case-type category" formula in Iowa:

Actual **Minutes** Recorded during Data Collection Period per Simple Misdemeanour cases	208 886
Divide by	/
Days of Data Collection Period	20
Equals	=
Statewide Minutes per Day Working on Misdemeanour Cases	10 444
Multiply by	*
Total Judicial Working Days per Year	215
Equals	=
State-wide Annualised Minutes for Simple Misdemeanour Cases	2 245 460
Divide by	/
Number of Simple Misdemeanour Cases Filings in the Previous Year	172 459
Equals	=
Initial Case Weight (minutes)	13

---

[34] Suzanne and John Douglas, *Iowa Judicial Officer Workload Assessment Study*, 2016, Denver, National Center for State Courts, Court Consulting Services, 2017.

The advantage of this formula derives from the fact that it does not require a research effort to collect and analyse frequency data on the various events, thus shortening the research duration and the time required to calculate the weights. The disadvantage of this formula is the loss of the ability to break down the weight and examine the most time-consuming events to gain a better understanding of the weight as reflection of the case-complexity.

Once the initial case-weights are calculated, they are then submitted to qualitative adjustments that may include review and validation of focus groups of expert-judges, sufficiency survey, interviews, onsite visits, etc. These measures allow the research team to adjust the initial weights, if necessary, so that it would adequately reflect the work time needed to efficiently and effectively process each case-type.

After completing the time-data collection phase, and the data analysis phase resulting in the final weight assigned to each case-type, the next phase is to calculate the needed number of full-time judge positions to handle the current weighted caseload. This is done using the following formula:

$$\frac{\text{Annual filings of each case-type x the weight assigned to that case-type}}{\text{(added up over all the case-types)}}$$

The number of minutes a single judge or a judicial officer has available for case-related work in one year.

5.   Comparative summary

This chapter will first present and discuss the common denominator of the models reviewed in this study and then proceed to present and discuss the main differences between these models. Based on these differences, we will then propose dividing the different models into four main groups, based on a two-layered classification that enriches the current classification in the literature. Finally, this chapter will end with a summary of the main reasons to choose one model/method over the other, preparing the ground for the recommendations to be discussed in the next chapter.

Accordingly, our findings show that the common denominator shared by all models is embodied in two primary elements. The first element is the "case-counts", meaning the counting of the annual number of filings per case-type in the last year. In this respect, it should be noted that there are models that count the number of cases resolved instead of the number of cases filed, or in addition to it. With that said, it seems that none of the models reviewed in

this study included case-counts of pending cases in their formula to determine the number of needed judges.

Moreover, there are models that calculate the average of the case-counts in the last three years, and other models that calculate the three-years average in a weighted arithmetic manner, in which the annual data of each year contribute to the final average in a different proportion (as done in the model in the Czech Republic). However, whether the case-counts refer to filings or to resolved cases, and whether it is done per one year or three years - the basic understanding that all case-weights models share, is that different case-types should be counted separately and awarded separate weights accordingly, to account for the differences in the amount of time and effort they require.

The second element, that all case-weights models share in common, is the self-reporting of judges and other judicial officials about the level of complexity entailed in each case-type. In this respect, it is important to differentiate between self-report, which is a research tool used in all case-weights models, and self-documentation, which is utilised only in some of the models. Essentially, self-reports are provided retrospectively, in response to questions that may later be discussed in a group setting and rely on the subjective notions and perceptions of the respondent. In comparison, self-documentation relies on real-time documentation of the work-flow of each participant in a continuous, uninterrupted and meticulous manner. This distinction is the essence of the differences between the Delphi method, that relies on self-reports in the form of calculated-estimations, and the Time-study method that relies on an empirically based self-documentation done by each participant. With this distinction in mind, it is important to note that while all case-weights models collect subjective self-reports (using the Delphi method or a variation of it), the use of these self-reports differs from one model to the other. For example, some case-weights models will use the self-reports of judges as the main source of data required to quantify the weight of the case. However, other models will use the self-reports of judges as a secondary means aimed only to validate the weights determined, based on the self-documentation of judges (for example, in a time-study). These differences will later be discussed in detail.

Coming back to the collection of self-reports, which is done in all models, it is important to note that the process of data collection demands decision-making on a long list of methodological issues. For example, the questions presented to the participants in the Delphi method (or a variation of that method) can be open-ended, allowing the respondents to reply in their own words, or closed-ended with fixed alternative answers to choose from. Additional decisions are required on the manner in which the questions will

be phrased; as well as on the number of respondents (the entire population or a sample of the population, a representative sample or a sample of experts alone etc.); and on the manner in which the questions will be presented to the participants. For example, questions can be presented via a survey or a questionnaire (in paper form or on-line), a face-to-face individual interview or a group interview followed by a group discussion, etc. Additionally, the questions can be preceded by a presentation of relevant statistical data, case-studies or a review of the findings acquired through other research tools, as a frame of reference. Understandably, all these methodological issues have an influence on the validity and reliability of the data collected and therefore require careful, informed, and well-founded decision-making. Such decision-making is also required when adopting additional or alternatives methods of data collection that raise similar and other methodological issues.

Following this review of the common denominators of all models, the main differences between the various models can be summarised under four key-characteristics as follows:

1. *The single weight-unit used to assess and present the complexity of the case* – meaning either time-units (i.e. minutes per case) or other numerical-relative-values, such as points or benchmarks, that grade the complexity of the case, but do not necessarily convert into work-time per case.

2. *The method of data collection* – traditionally, the common classification distinguishes between two primary methods of data collection: the Delphi method (estimations) and the Time-study method (empirical measurements). To this common classification, we add a third method which is the Work-sampling method used to determine the proportion of time spent by the participants on various activities by sampling random moments along their day. As discussed above, this method was implemented in the Dutch model and was later combined with the Delphi method.

3. *The transparency of the case-weights,* meaning the ability to break-down the weight of the case to the different work-generating factors, of which it is composed, and to analyse the effect of each of these factors on the total weight. Such factors include time-consuming events and tasks that judges perform in cases, and/or other case-related indicators that try to predict the complexity of the case. By doing so, a transparent weight provides a better and in-depth understanding of the complexity of the judicial work.

4. *The number of separate and independent case-weights in each model.* In this respect, a broad list of separate and independent case-weights can indicate a higher level of detail and accuracy of the case-weights model. In other words, grouping different case-types into fewer and fewer categories, and awarding the same weight to all case-types grouped under the same category, eliminates the differences between those case-types. For this reason, a narrow list of might undermine the purpose of reliably reflecting and quantifying the differences in complexity of different case-types.

These four characteristics will now be discussed in detail.

1. *The single weight-unit used to assess and present the complexity of the case*

In this respect, some models assess and present the weight of the case using time-units (most often minutes) or relative numerical figures that can later be converted into time-units. Such weights are implemented, among others, in Austria, Germany, the Netherlands, Serbia, Israel[35] and in the United States. On the other hand, there are models that choose to weight the complexity of different case-types using benchmarks on an open or closed grading-scale, that do not convert into time-units. Such models are implemented, for example, in Finland and Lithuania that assign cases with weighting coefficients, or in Romania that assigns cases with an initial complexity grade.

It is important to explain that choosing to present the case-weights using time-units allows for the completion of the following three calculations:

- *Calculation of the proportionality of different case-types* - for example, if the weight of a civil case is 80 minutes and the weight of a criminal case is 40 minutes, then the ratio of civil cases to criminal cases is 1 to 2, meaning a civil case is twice more complex and time-consuming than a criminal case. Another implication derived from this is that for every civil case assigned to judge A, two criminal cases should be assigned to judge B in order to create an equitable and balanced assignment of cases.

- *Calculation of the difference between different case-types* - for example, using the same numerical figures above will result in

---

[35] In the Israeli case-weights study conducted in 2012, a single weight-unit was the equal of 2.8 minutes.

conclusion that a civil case requires 40 minutes of judicial work-time more than a criminal case.

- *Calculation of the aggregated judicial working-time required to process different case-types* - for example, if the number of incoming civil cases is 100 and the number of incoming criminal cases is also 100, then this means that the aggregated judicial work-time required to process these cases is 12 000 minutes which are equal to 200 working hours in total. This is as a result of the following calculation:

  [(40 minutes * 100 filings) + (80 minutes* 100 filings)] \ 60 minutes = 200 hours of judicial work. By comparing this result to the judicial work-time that is available to the court system, it is then possible to calculate the number of full-time judge positions required to handle the current weighted cases.

While most relative-values maintain the proportionality attribute of the time-units, they may not allow the calculations of the difference between the different case-types or the aggregated work-time and the needed number of judges in accordance to it. For example, the weighting coefficients in the Finnish model maintain the proportionality attribute of the time-units, thus allowing to determine, for example, that the ratio of civil cases to criminal cases is 2,2 to 1. However, the Finnish weighting coefficient does not convert into time-units, and therefore cannot be used to calculate the difference between different case-types. To illustrate, the difference between a civil case and a criminal case in the Finnish model is 1,2 (2,2-1). However, this difference is seemingly meaningless without the key to convert it into time-units. Respectively, these weighting coefficients are not aimed at calculating the aggregated judicial working-time required to process the number of incoming cases of different case-types. This is also the case in the Lithuanian and Romanian models. Additionally, as noted in the review of the Lithuanian model, the proportionality attribute of the Lithuanian weighting coefficients is limited to courts of the same jurisdiction or level alone. The meaning of this is that the case-weights of different jurisdictions cannot be compared and evaluated to ensure a balanced and equitable allocation of cases between the different jurisdictions. These examples showcase the possible advantages of the time-units in comparison to the benchmarks and other grades. This is also true for the Estonian model.

The exception to the classification of the single weight-unit (as either minutes of work-time per case or relative grades per case) is found in the Czech and Croatian models. In both these models the so-called "weight" is a derivative of a prior determination of the quota of cases that each judge is expected to

resolve in one year. For example, in the Czech model, a district court judge is expected to resolve 282 criminal cases per year. Meaning the "weight" value of 282 does not represent the number of minutes per a single criminal case. Alternatively, this numerical value can be considered as some sort of an aggregated weight assigned to the entire category of criminal cases per one judge-year. This aggregated weight is therefore the equivalent of a single judge position. In other words, the ratio of district court judges per criminal cases in the Czech Republic is 1 to 282.

Similarly, the Croatian model also relies on case-quotas, but takes them one step further by calculating the weight of the single case. To illustrate, if the annual quota per judge is 165 general-civil cases, then the weight of a single general-civil case is calculated using the following formula: 1 year/165 cases = 0.006. In other words, a single general-civil case requires 0.006 judge-year, on average. Theoretically, this weight can be converted into minutes per case, given a determination of the judge-year value (number of minutes per year that a single judge has available for case-related work). However, to our knowledge this use is not formally applied neither in the Croatian model nor the Czech model. To sum up, the single weight-unit in the Czech and Croatian models may be considered as a time-unit. However, it is not the common time-unit of minutes per a single case and is therefore the exception to the dichotomous classification distinguishing between time-units of minutes per case or grades and benchmarks that do not convert into time-units.

2. _The method of data collection_

Traditionally, the common classification of the case-weights models divides them into two: models that rely on the Delphi method and models that rely on the Time-study method. However, this dichotomous classification may not apply to models that implement a mix of the two methods or to models that implement a third method that does not fit the unique characteristics of neither the Delphi method nor the Time-study method. That is the case in the Dutch model that relies on a method that analyses the time-use of judges by sampling their work in random points along their daily work-flow. This method is called "work sampling" (other names for it include snap-shot sampling, occurrence sampling, multi moment analysis, etc.).

Essentially, all three methods collect data from participants in order to quantify the level of complexity of different case-types. However, these methods differ in the manner of the data collection, in the duration of the data collection, in the number of participants and in the level of involvement of each participant, as well as in the single-weight unit that will eventually

represent the weight of the case. These differences are summarised in the table below.

Table 6: Differences between the methods of data collection

	Delphi method (Calculated estimate)	Work-sampling method	Time-study method (Empirical Measurement)
The type of datas collected from the participants and the manner of data collection	**Subjective self-reports** that estimate the complexity level of the case. These self-reports are provided **retrospectively**, in response to structured questions, later discussed in a **group setting** in order to reach a consensus.	**Segmented real-time reporting** of the current activity performed by each participant, without the need to report the start and end time of the activity, or its' duration. The reporting is done to a third party (computer app or phone/ online reviewer), **in random moments.**	**Real-time and continuous self-documentation of the duration** of each activity, as measured by each participant individually in a manual or an on-line time-log form. *The level of detail is of course dictated by the design of the judicial time-log.
The duration of the data collection (not including event-frequency data collection)	Short-term period, often limited to 1-2 focus groups meetings.	Relatively Short-term period, most often ranging from 1 week to 3 weeks.	Relatively long-term period, ranging from 4 weeks up to several months.
The number of participants	A sample of the population (requires decision on the sampling method).	A sample of the population (requires decision on the sampling method).	A representative sample of the population or the entire population
The level of involvement required from the participants	Minimum involvement. Subjective self-reports and estimations provided in a group setting.	Medium level of involvement. Requires partial availability for Fragmented time-use reports, at random points in time, throughout one week and up to three weeks.	High level of involvement. Judges are required to self-document the specific work-time they spend on each judicial activity in a continuous and uninterrupted manner for several weeks or months.
The single weight-unit	Grades/ Benchmarks or time-units.	Time-units.	Time-units.

As previously explained, it seems that all case-weights models utilise the Delphi method (or a variation of it) to collect self-reports of judges and other relevant personnel. However, in the models that apply the Time-study method, as the primary research-tool and main source of data, the Delphi method is used only as a secondary research tool aimed to validating the findings of the time-study and amend them, if necessary. This is the case in the Austrian and German models as well in the updated studies conducted in the U.S., among others. In principle, this can also be true of models that rely on the Work-sampling method as the primary source of data and use the Delphi method as a secondary research tool. However, it seems that in the Dutch case-weights model, it might not be accurate to define which of the two methods (Delphi or work-sampling) is the primary source of data. That is because the contribution of both methods to the final weights in the Dutch model may be considered as equal. For this reason, we choose not to classify the Dutch model under the Delphi method but under a third classification which is a mix of the Delphi with the Work-sampling method.

Accordingly, the following table proposes to classify the different case-weights models based on a two layered classification. The first layer will divide the models according to the weight-unit they use (time-units of minutes vs. grades or benchmarks that do not necessarily convert into minutes). The second layer will then divide the models according to the primary method they use to determine the initial case-weights. This layer will be based on the three-fold classification of the methods previously discussed: Time-study, a combination of the Work sampling method with the Delphi method, and the Delphi method alone (or a variation of it). To clarify, models that apply the time-study to determine the initial case-weights and use a post-time-study Delphi to validate the time-study findings will be classified under time-study.

Table 7: The proposed two-layered classification of the different case-weights models

The **weight unit** of the model	Time unites of minutes			Grades and benchmarks that do not necessarily convert to minutes
The **main** method used to collect the primary input of judges	A. Time-study method	B. Work sampling method + Delphi method	C. Delphi method (or a variation of it)	D. Delphi method (or a variation of it)

As shown in the table above, the classification of the different models divides them into four major groups:

A. Group A includes all models, that present the weight in **time-units**, and are based on a **time-study**. This group of models include, among others, the models implemented in Austria, Denmark, and Germany. This group also include the studies conducted in both the United States model and in the Israeli model which switched from the previously used Delphi method to the Time-study method.

B. Group B includes all models, that present the weight in **time-units**, but are based on a combination of **the Work-sampling method and the Delphi method**. This group refers to the Dutch case-weights model.

C. Group C includes all models, that present the weight in **time-units**, but are based on the **Delphi method (or a variation of it)**. This group of models include the Serbian model, the 2013 Israeli case-weights model and the U.S. model previously based on the Delphi method.

D. Group D includes all models, that present the weight in **grades or benchmarks**, and are based on the **Delphi method**. This group of models include, among others, the models implemented in: Croatia, Czech Republic, Finland, Lithuania, and Romania. This list is not exhaustive and based on the partial data we collected. It seems that the majority of the member States that responded to the questionnaire may be classified in this group.

As for the reasons of each judiciary for choosing one method/model over the other, these rely on a careful examination of the strengths and weaknesses of each method. Respectively, the main strengths and weaknesses of each method will be presented in a separate table as follows.

Table 8: Main strengths and weaknesses of the Delphi method

The Delphi method	
Main strengths	Main weaknesses
1. Shot-term process of data collection (often limited to one or two daily focus group meetings) and relatively low cost (no external services required).	1. The classic Delphi method uses focus-groups of expert-judges and is therefore not representative and does not account for newly appointed and inexperienced judges.
2. Achieves a consensus based on the opinions of expert-judges or a representative sample of peer judges.	2. The reliability of these subjective self-reports is highly compromised due to cognitive biases (e.g. under or over-estimation), group dynamics influences and human perception errors.
3. Simple data acquisition, that requires minimal effort to participate, as well as minimal interruption of the workflow of the participants.	3. The design of the questions presented to the participants can also influence their responses and jeopardise the reliability and validity of the data[36].

Table 9: Main strengths and weaknesses of the Work-sampling method

The Work-sampling method	
Main strengths	Main weaknesses
1. Unlike the Delphi method, it provides a statistical technique that allows the calculation of the proportion of the time spent per each activity.	1. It does not record the actual duration of the sampled activity as done in the time-study.
2. It relies on an empirically based, real-time data, collected in random moments, thus increasing the reliability and accuracy of the data set in comparison to the Delphi method. Additionally, this manner of data collection also enables the monitoring and supporting during data collection as well as interpretation of the data in real-time to ensure its reliability.	2. It does not allow a small breakdown of the observed activities in the workflow or enable the identification of the most time-consuming event as made possible in the time-study (depending on the research design).

---

[36] Ibid.

The Work-sampling method	
Main strengths	Main weaknesses
3. In comparison to the time-study method, it requires relatively short-term process of data collection (several weeks, pending on the project design), and requires relatively simple data acquisition and low level of effort and interruption of the participants' workflow.	3. It requires a sufficient number of participants and a sufficiently long period of data collection to ensure a representative data set. Therefore, pending on the project design, the data collection may be as time-consuming and expensive as the Time-study method[37].

Table 10: Main strengths and weaknesses of the Time-study method

The Time-study method	
Main strengths	Main weaknesses
1. It provides a direct measure in the form of first-hand account of the individuals' work-time based on real-time and continuous records.	1. Considered to be more expensive and time-consuming than the Delphi method.
2. It provides a robust and empirically-based source of data that allows analysis in various and in-depth dimensions that other methods do not provide. These dimensions include providing average/ median/ minimum and maximum values of the different activities' duration and a variety of analysis options per court, per judicial position, per case, per activity, etc. For these reasons, the time-study method is considered as "the gold standard for case-weighting studies"[38].	2. It is more burdensome to the judges tasked with meticulous and continuous documenting of their work-time for several weeks and even months.

---

[37] For example, the scheduling of individual observation week per each participant in the Dutch model was spread over one year, thus equalizing the timeframe to complete the research to that of the average model that relies on the time-study method.

[38] Kleiman, M.: Lee, Cynthia G.; Ostrom, Brian J.; Schauffler, Richard Y. 2017. "*Case-Weighting as a Common Yardstick: A Comparative Review of Current Uses and Future Directions*", Onati Socio-legal Series [online], 7 (4), 640-660. Available from: http://ssrn.com/abstract=3047725

To sum up, the decision which method is best-suited for a given judiciary should take into account the following factors, among others: the timeframe set to finalise the case-weights, the budget and the human resources available to conduct the project; the technological and administrative capacities of the organisation; the level of willingness of judges to participate in a long-term research effort etc. To illustrate how these factors come into play, studies found that the "Delphi method is typically employed in contexts where administrative data are limited, project time-lines are short, budgets are tight and cultural and political barriers reduce the likelihood of high judicial participation rates in a time-study"[39].

3. _The transparency of the final case-weights_

In this respect, there are models that present the weight of the case as the sum of all its parts and allow the drawing of conclusions on the relative impact of each factor or component on the end-result of the case-weights. A good example for a transparent weight is that of the "events-based case-weights formula" used, among others, in the United States, Serbia and Israel. As previously discussed, the events-based formula allows the researcher to break down the total weight into events, compare the time-consuming events among different case-types, identify the most time-consuming event and draw additional conclusions based on that. By doing so, the events-based formula promotes a better understanding of the unique complexity entailed in each case-type.

With that said, the events-based formula relies on the ability of the research-team to extract accurate, up to date and reliable data on the frequency with which each event accrues in each case-type (for example, the number of hearings per case, the number of decisions per case, the number of final decisions per case and the type of decision, etc.). This ability relies on two alternative data sources. The first, which is preferable to the other, is the case management system (CMS) or other statistical and computerised data-systems that collect the relevant data-items. Simply put, if these systems (CMS or others) continuously and reliably collect events-frequency-data, the process of data collection and analysis is fairly easy, fast and efficient. However, if the relevant frequency-data cannot be extracted in a computerised manner (because it is not computerised or its reliability is insufficient) the alternative source of data is via a manual sampling of the paper court-cases. This will of course require access to the paper copy of the cases and a sufficient number of qualified researchers with legal-education to sample a representative amount of cases of each case-type. Naturally, manual sampling of cases will prolong the process of data

---

[39] Ibid.

collection, arise reliability issues, and delay the completion of the study. For these reasons, it is essential to apply a long-term strategic approach to ensure that all the required data will be computerised in a timely, accurate, reliable, complete and accessible manner.

An alternative way to present the weight of the case, which does not require frequency-data (or other case-related items of data) is referred to as the "black box" weight. In the black box weight, the weight of the case is a non-transparent value that does not provide information on the specific events (or other indicators) compiling the weight and their relative impact on that weight. As discussed in previous chapters, the case-weights in the Austrian and German model are presented as such, and so are the weighing coefficients in the Finnish, Lithuanian and Romanian models. With that said, it is interesting to see that the case-weights formula in the Romanian and Finnish models include transparent indicators, alongside the black box coefficients. For example, the formula in the Romanian model includes transparent indicators such as: the number of tomes, the number of parties and the number of witnesses in each individual case. Thus, promoting a more accurate and empirically based depiction of the complexity level of the case.

In summation, the transparency attribute of the weight is one of the indicators of the level of detail and accuracy of the case-weights model. In other words, models that utilise transparent weights are considered to be more detailed and accurate than others.

4. *The number of separate and independent case-weights in each model*

In this respect, it is essential to understand that the number of case-types or case-type-categories that are eventually awarded a separate and independent weight can also indicate the level of detail and accuracy of the case-weights model. Simply put, the more case-types that are weighted separately, the more detailed and precise the model will be.

With that said, including more case-types that require independent weights directly influence the amount of time and research effort required to extract the relevant data per each case-type and to analyse said data (i.e. number of filings and other relevant case-related indicators such as events-frequency). For this reason, some models may prefer to group different case-types into fewer and fewer categories in order to ease the level of research-effort and to shorten the time-set required to calculate the weights. However, the grouping of different case-types into one category will cancel out the differences between those case-types. Simply put, different case-types that were grouped in the same category will be awarded the same weight.

As a frame of reference, our findings show that the Dutch model, for example, assigned separate weights to 70 different case-type categories; in the Lithuanian model, weights were awarded to 45 case-types; and in the Serbian model weights were assigned to 8 case-type-categories. In comparison, the weighted caseload model in the United States also group case-types into weighted categories but most often assign weights to a limited list of categories. For example, 13 case-type-categories were awarded weights in Montana District Court in 2014; 18 case-type-categories were assigned weights in Oregon Circuit Court in 2016; 28 case-type-categories were weighted in Iowa in 2017; and 33 case-type-categories were weighted in Kentucky in 2016. It is also interesting to note that the 2013 Israeli case-weights model assigned 101 case-types with a separate and independent weights.

To sum up, the decision of how many case-types to include in the model and whether they should be grouped into fewer categories, should carefully be balanced between the interest of ensuring a detailed and accurate depiction of as many case-types, and the interest of completing the research in a short period that requires minimum effort.

Cross-referencing the four key-characteristics discussed above (weight-unit, method of data collection, weight-transparency and number of weights) allows the ranking of the different models according to the level of detail and accuracy of each model as well as the level of research effort that each model requires (in terms of time, budget and the level of involvement required from the participants). In this respect, it seems that the level of accuracy and detail correlates with the level of research effort entailed in each model. For example, a case-weights model that will rank high on the scale of accuracy and detail will be a model that: is empirically-based and relies on continuous self-documentation of the actual work-time invested in different case-types; presents the weights in a transparent manner; and yields a large number of independent case-weights. This model, however, will also be ranked high on the scale measuring the level of research effort due to the fact that this model is more expensive, time-consuming and burdensome in comparison to models that rely on other methods. Accordingly, a relatively low level of accuracy and detail will be attributed to a model that: relies mainly on retrospective and subjective self-estimations; presents the weights as grades or benchmarks, in a black-box manner; and groups different case-types into one category in a manner that cancels out the distinct differences between those case-types. Respectively, the research effort entailed in such a model will also be ranked at a relatively low level.

The following figure will illustrate the proposed scale to assess the influence of different methodological issues on the level of accuracy and detail and the correlating level of research effort required.

Figure 1: A scale to assess CW models based on their level of accuracy, detail and research effort
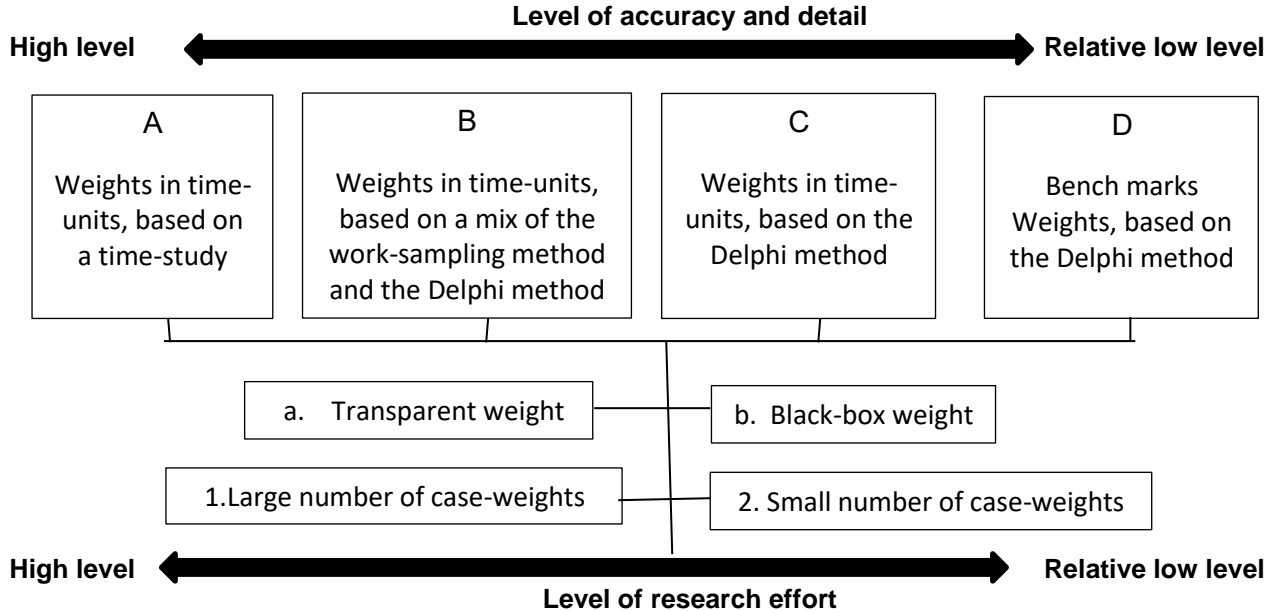
As shown in the figure above, a research design that will be classified as B+a+1 is considered to be more accurate and detailed as well as demanding more research-effort than a model classified as C+b+1, for example. Another possible example will scale model C+a+1 at a higher level of both accuracy and research-effort demand in comparison to model D+b+1. These are just two examples of the numerous comparisons that this scale enables between models that rely on different combinations of methodological choices.

Based on this comparative summary, it is evident that the task of developing a successful case-weights system requires comprehensive preparation work, careful implementation, quality control measures and periodical review to ensure that the weights will remain valid, reliable, accurate and up to date. For this purpose, the following chapter will offer concise recommendations for judicial systems that consider adopting a case-weights system or revising their current system.

6.   Recommendations

The following list of recommendations is based on the extended review presented in this study and is not intended to be an exhaustive one. The aim of this list is to provide basic guidelines to judicial systems seeking to develop a case-weighting system or re-evaluate the system that is currently in use.

Recommendation 1: Engaging and encouraging judges to actively participate in the research
Judges are key stakeholders in the process of case-weighting. Their active participation and indispensable input are vital for the success of any case-weights model. For this reason, it is essential to engage and encourage judges to voice their opinions and actively participate in all stages of the research. Doing so will promote transparent and collaborative work as well as increase the judges' willingness to participate in the process; ensure that the data set will be representative and reliable; and increase the judges' trust in the research methodology and findings. In this respect, a possible way to ease the acceptance of the process is to anonymise the identity of the participants and ensure judges that the data will not be used to assess their individual performance. Additionally, the fear of misuse of data collected by internal research bodies can also be eased by working with external research institutions or independent consultants[40].

Recommendation 2: Appointing an experienced research project manager
The research project manager should have relevant professional experience in this field of research and be able to command the respect of judges. The research project manager will be entrusted with the planning, design and execution of the research project from the start to end, subject to the review and approval of an advisory committee.

---

[40] Footnote 36 above.

Recommendation 3: Forming the advisory committee
The advisory committee will consist of experienced members of the judiciary and may include among others: expert-judges from courts that differ in size, geographic location or jurisdiction, researchers, statisticians, court clerks and court administrators, etc. The advisory committee will be entrusted with providing guidance on policy matters, overseeing the research project and approving the research findings[41].

Recommendation 4: Completing a comprehensive assessment of available resources
The success of any case-weights model, which will be tailored to the measurements of a specific judiciary, requires a comprehensive and preliminary assessment of the available resources in order to fulfill the purposes of the model. This assessment will include:
- Compiling the list of case-types that will be included in the research;
- Identifying the time-consuming tasks and activities/events to be classified as case-related work and non-case-related work;
- Testing the availability, accessibility and reliability of the computerised statistical data on relevant case-related indicators (number of annual filings per case-type and/or events-frequency per case-type, etc.) [42];
- Determining the need to complete the data by hand-sampling court cases and the capacity to do so;
- Assessing the needed number of researchers, statisticians, technical support team and other personnel required for the successful execution of the research design;
- Assessing the need for external services and available resources to finance these services;
- Setting appropriate, sufficient and realistic timeframes that consider the unique requirements of each stage of the research;
- Assessing the level of willingness of judges to participate in a long-term research effort.

Recommendation 5: Deciding how many case-types will be included in the research
Determining the list of individual case-types that will be awarded an independent weight, or the grouping of several case-types into one

---

[41] Matthew Kleiman et al., *An Assessment of Louisiana's Judicial Workload Model,* National Center for State Courts, 2014.
[42] See footnote 36.

category that will be awarded the same weight, should be based on the following requirements:

- It is imperative to ensure that the case management system records data on the number of filings per each identified case-type in a consistent and reliable manner [43].
- The number of filings per case-type should be sufficient to ensure the reliability of the weight.
- The classification into case-types, or into case-types categories that group different case-types, should rely on legally and logically clear-cut distinctions between the types of cases. In other words, this classification should be both exclusive and collectively exhaustive to ensure that any given case falls into one, and only one, case-type or case-type category [44].
- The grouping of different case-types into one category can also rely on similarities in the average amount of judicial work required to process the single case of each type. However, this manner of grouping can be done only after the weights have been decided.

Recommendation 6: Deciding the research design
The research design should delicately balance between the different cost-benefit considerations. This balance will take into account the desired level of detail and accuracy of the chosen research design and the purposes the case-weights aim to fulfil versus the timeframe set for the project, the financial and human resources available for the project and the level of involvement required from the judges. The list of methodological questions that will determine the research design includes:

- The number of participants – in this respect, using the entire population will ensure a robust data set with representative findings which do not require sampling methods. However, it also requires more effort and money to collect the data than in an adequate and representative sampling of the population.
- The sampling technique – naturally, this is relevant only to studies that have had to base their data set on a sample. In this respect, the rule of thumb is to use stratified/representative sampling of the entire population to ensure that the findings will be reliable and representative. Other sampling techniques include: random sampling, convenience sampling, voluntary sampling, experienced judges, alone etc. However, the

---

[43] Ibid.

[44] Brian.J. Ostrom et al., *Florida Judicial Workload Assessment Final Report*, National Center for State Courts, 2016.

reliability of the findings based on these techniques is considered questionable.

- The methodology of data collection – this decision should be based on the unique characteristics of each method (Delphi method, Time-study method or a combination of the work-sampling method and the Delphi method) as discussed in the comparative summary.
- The single weight-unit (minutes of work or benchmarks and grades) and the transparency of the weight as discussed in the comparative summary.
- The duration of data collection – this decision should be made in accordance to the number of participants and the chosen data collection methodology.
- The design of the questions (in the Delphi method) or the design of the time-documentation forms for the time-study (paper and online form), or the manner of data collection in the work-sampling method (phone reviewer or computer/ smartphone app).
- The "judge-year" and "judge-day" value – these values are an integral part in models that weigh cases in time-units and allow for the calculation of the needed number of judge-positions to adequately and efficiently handle the current weighted caseload. These values should be set by the advisory committee.

Recommendation 7: Scheduling the period for data collection
The collection of the judge's self-reports or self-documentation should be scheduled sensitively, and preferably not before or during major reforms that might risk the accuracy, the validity and the reliability of the data collected. It is also advisable to schedule the data collection in a timing that will not overlap with official holidays and leaves of absence[45].

Recommendation 8: Developing a training programme for the participants
In large-scaled data collection research projects, it is imperative to develop a training programme to ensure, in advance, that the data will be collected in a unified, reliable and accurate manner. In this respect, the training of participants on how to document time per activity (in a time-study) or how to report time-use (in a work-sampling study) can be done, using on-site training presentations provided in the courts, instructional videos and written guidelines, etc. An additional tool is to conduct a preliminary test-run of the time-study forms or the work-

---

[45] Footnote 38 above.

sampling computer-app to ensure that the participants are correctly applying the guidelines.

Recommendation 9: Establishing a technical and professional support team
In a time-study or work-sampling study, it is essential to provide the participants with technical and professional support for the duration of data collection. The technical support will address difficulties, such as failure in logging-in data onto the on-line website or computer/ smartphone app. Respectively, the professional support will deal with questions on proper classifications of the observed activities, over-lapping reports and other research policy issues that the participants encounter during data collection.

Recommendation 10: Applying measures to validate the initial weights and determine the final weights
A quality adjustment process should be undertaken to ensure that the weights adequately reflect the complexity level of each case-type. Such quality adjustment measures may include post-data collection surveys asking the participating judges to assess whether the initial weights reflect sufficient time to provide high quality services, or only reflect the work-time spent given the current time-constraints regarding various case-types. These surveys may also include open-ended questions in which judges will identify the weights that require adjustments and justify the specific adjustment needed. An additional post-data collection measure is conducting focus group discussions (the Delphi method or a version of it) to evaluate the initial weights and the required adjustments to approve the final weights. Additional measures include individual and group interviews with judges and judicial officers, as well as on-site visits in courts to gain insight on good practices and effective work-patterns[46].

Recommendation 11: Setting appropriate uses of the case-weights
In principle, case-weights quantify the complexity level of different case-types by assessing the amount of time and effort required, on average**,** to process a single case of a certain case-type. This assessment of the average time and effort per a single case is not to be confused with assessing the performance of the single judge. To clarify, the use of the average value acknowledges that there is, for example, a small-claim case that requires 10 minutes to process and another small-claim case that requires 30 minutes to process. However, the weight of both these

---

[46] Matthew Kleiman et al., An Assessment of Louisiana's Judicial Workload Model (January 2014), National Center for State Courts.

cases will be identical and set at the average value of 20 minutes per case. This means that case-weights are not suited to determine that a judge that devotes the case its required 30 minutes is inefficient in comparison to the average value of 20 minutes. For this reason, it is not recommended to use weights to assess the performance of judges. In this respect, it should also be taken into account that using case-weights to assess performance might discourage judges to participate in the data collection and impair the ability to achieve a representative data set and determine a reliable weight.

Recommendation 12: Carrying out periodical reviews to ensure that the case-weights remain up to date and reliable
Over time, the integrity of any case-weighting system may be affected by external factors, such as changes in legislation, case law, legal practices, court technologies and administrative policies, etc. For this reason, case-weighting systems should be reviewed on an annual basis and updated every 5-7 years, or every 3 years in studies that rely on work-sampling. With this said, if such major changes in the law, technology or policy occur in between regular updates of the model, a Delphi focus group may be convened to discuss the affected case weight(s) and decide on the proper steps to be taken[47].

7.   Concluding remarks

It is evident that quantifying the complexity of the judicial work, that is required in each case, on average, is a highly demanding and complex process in itself. However, this process is essential in order to make educated and data-driven decisions that rely on tangible and measurable standards. As the famous quote by Peter Drucker says: "you cannot manage what you cannot measure", and in the context of case-complexity assessment, the measuring tool is the case-weighting system.

Respectively, this study provided an overview of different case-weighting practices in Europe and zoomed-in on the case-weights models implemented in Austria, Denmark, Estonia, Germany, Romania and the Netherlands. As additional grounds for comparison, we also reviewed the case-weights model in the United States.

Based on this extensive review, we then compared the different models and pointed out their similarities and differences. In this respect, our findings

---

[47] Brian.J. Ostrom et al., Florida Judicial Workload Assessment Final Report (National Center for State Courts, 2016).

show that all the models reviewed in this study share a common denominator in the form of two elements. The first is the use of case-counts (counting the number of cases per case-type), and the second is the use of subjective self-reports of judges on their perception of the complexity level of different case-types (using the Delphi method or a variation of it).

Furthermore, we concluded that the models differ one from the other in four main characteristics. The first characteristic is the single weight-unit which divides the models into those that represent the weight of the case in time-units vs. models that use grades or benchmarks that do not necessarily convert to time-units. The second characteristic is the primary method used in each model to collect the data from judges. In this respect, we added to the common classification of the Delphi method vs. the Time-study method, a third classification of the Work-sampling method combined with the Delphi method. To our knowledge, the use of this unique combination of methods in the field of judicial case-weights has been implemented only in the Netherlands, thus adding on to the traditional dichotomous classification known in professional literature (Time-study vs. Delphi). The third characteristic we analysed was the ability to break down the weight into the different work-generating factors it is composed of, as an indicator of the level of accuracy and detail of the weights. In this respect, we divided the models between those that use transparent weights vs. those which utilise black-box weights. The last characteristic discussed was the number of case-types that were awarded separate and independent case-weights in each model. This characteristic was also proposed as an indicator of the level of detail and accuracy of each model.

After cross-referencing our findings regarding the four key-characteristics previously discussed, we then suggested a ranking scale of the different models according to the level of detail and accuracy of each model as well as the level of effort that each model requires (in terms of time, budget, and the level of involvement required from the participants).

In the final chapter we provided a non-exhaustive list of recommendations aimed at guiding judicial systems and policy makers interested in developing a case-weighting system or seeking to re-evaluate the system that is currently in use.

We hope that this initial tool-box will fulfill its purpose and lay the ground for further advancement of this field of research and practice.

Bibliography

Benton County Washington State, *Case Weighting System*, Washington State, 2018.

Ceretto J.D., *Weighted caseload formula for the courts of Serbia, Interim Report*, Judicial Efficiency Project, British Council, 2017.

Committee of Ministers of the Council of Europe, *Recommendation No. R (94) 12*, adopted 13 October 1994.

Consultative Council of European Judges (CCJE), *On Standards Concerning the Independence of the Judiciary and the Irremovability of Judges*, CCJE (2001) OP N°1, Strasbourg 2001.

Consultative Council of European Judges (CCJE), *On the Evaluation of Judges' work, the Quality of Justice, and respect for Judicial Independence*, CCJE (2014) OP N°17, Strasbourg 2014.

Council of Europe, *European Charter on the Statute for Judges and Explanatory Memorandum*, Strasbourg, 1998.

Dalkey, Norman & Olaf Helmer, "An Experimental Application of the Delphi Method to the Use of Experts", *Management Science* 9 (3): 458-467, 1963. Dalkey, Norman, B. Brown and S.W. Cochran, "The Delphi Method III: Use of Self Ratings to Improve Group Estimates," *Technological Forecasting and Social Change*, 1, 1970.

Diller, J., *Weighted Caseload Measures & The Quarterly Case Status Report*, Indianapolis, Illinois, 2018.

Douglas, J., Tallarico, S., Friess E. & W. Wills, *West Virginia Circuit Judge Workload Study*, National Center for State Courts, State Justice Institute, 2014.

Dougals S. & J. Douglas, *Iowa Judicial Officer Workload Assessment Study*, National Center for State Courts, Court Consulting Services, Denver, 2017. European Commission for the Efficiency of Justice, *CEPEJ Glossary*, Strasbourg, Council of Europe, 6 December 2019/5.

European Network of Council for the Judiciary, *Dublin Declaration on Standards for the Recruitment and Appointment of Members of the Judiciary*, 9-11 May 2012.

Federal Judicial Center, *2003-2004 District Court Case Weighting Study*, Washington D.C., 2005.

Finland Ministry of Justice. Department of Judicial Administration. Court Administration Unit, *Principles for Calculation Workloads and Performance Targets According to the Workload Indicators of General Courts*, Memorandum, 1 July 2013.

Flango, E. & B.J. Ostrom, *Assessing the Need for Judges and Court Support Staff*, National Center for State Courts, Williamsburg, VA, 1996.

Florida Legislator, Office of Program Policy Analysis and Government Accountability (OPPAGA), *Information Brief on Weighted Caseload Methods of Assessing Judicial Workload and Certifying the Need for Judges*. Report No 97-96, OPPAGA, Tallahassee, FL, 1998.

Gordon, T.J., *The Delphi Method*, in Glenn J.C. & T.J. Gordon (eds.), *Futures Research Methodology*, CD-ROM Version 3.0. Washington: The Millennium Project, American Council for the United Nations University, 2009.

Gramckow, H., *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

Hammergren L., Harley G. & S. Petkova, *Case Weighting Analyses a Tool to Promote Judicial Efficiency: Lesson, Substitutes and Guidance*, The World Bank Group, Washington DC., 2017.

Harold A. Linstone & Murray Turoff (eds.), *Introduction to the Delphi Method: Techniques and Applications*, 2002, available at: https://web.njit.edu/~turoff/pubs/delphibook/delphibook.pdf.

Hart, S.G., Wickens, C.D., *Workload Assessment and Prediction*, in Harold R. Booher (ed.), *Manprint, an approach to systems integration*, Van Nostrand Reinhold, New York, 1990, pp. 257-258.

Helmer, O., *Analysis of the Future: The Delphi Method*, The Rand Corporation, Santa Monica, California, 1967.

Judicial Council of California, *The Need for New Judgeships in the Superior Courts: Preliminary 2018 Update of the Judicial Needs,* Assessment report to the legislature, 3 November 2018, available at:

https://www.courts.ca.gov/documents/lr-2018-JC-judicial-needs-assessment-GC69614_c_1-and-3.pdf

Kleiman, M. et al., *An assessment of Louisiana's Judicial workload model*, National Center for State Courts, 2014.

Kleiman, M., "Weighted Caseload: a Critical Element of Modern Court Administration", *International Journal of the Legal Profession,* 26, January, pp. 21-32, 2019.

Kleiman, M., et al. "Case Weighting a Common Yardstick: A Comparative Review of Current Uses and Future Direction", *Onati Socio-Legal Series*, vol. 4, 2017.

Krejčí, J, "Approaching Quality in Survey Research: Towards a Comprehensive Perspective". *Sociologický časopis / Czech Sociological Review*,n. 6.

Langbroek, P. & M. Kleiman, *Backlog Reduction Programs and Weighted Caseload Methods for South East Europe, Two Comparative Inquiries*, Sarajevo, 2016.

Lienhard, A. & D. Kettiger, "Research on the Caseload Management of Courts: Methodological Questions", *Utrecht Law Review*, 7(1), 66, 68, 2011. McDonald, H.G. and C.P. Kirsch, "Use of the Delphi Method as a Means of Assessing Judicial Manpower Needs", *The Justice System Journal* 3 (3), 314-321.

Ministry of Justice of Finland, Department of Judicial Administration, *Memorandum: Principles for calculating workloads and performance targets according to the workload indicators of general courts*, 2013.

Morales Campoya Á. F., "Development of the NASA-TLX Multi Equation Tool to Assess Workload.," *International Journal of Combinatorial Optimization Problems & Informatics* 11, n. 1, January 2020, 50.

Mozuni, M. and W. Jonas, "An Introduction to the Morphological Delphi Method for Design: A Tool for Future-Oriented Design Research," *The Journal of Design, Economics, and Innovation* 3, n. 4 (2017): 303–318, doi:10.1016/j.sheji.2018.02.004.

Reidel, J., *Caseload and Weight of Cases in Special Types of Cases First Instance Criminal Cases in Regional Courts in Germany*, Paper presented at the EGPA Annual Conference, Edinburgh, September 2013.

National Center for State Courts, *Kansas District Court Judicial and Clerk Staff Weighted Caseload Study*, October 31, 2011.

Okloli, C. and S.D. Pawloski, "The Delphi method as a research tool: an example, design considerations and applications*", Information and Management* 42(1) pp. 15-29.

Partovi, F. Y., & Burton, J. (1993). "Using the analytic hierarchy process for ABC analysis". *International Journal of Operations & Production Management*,13(9), 29-44.

Ostrom, B. J.& M. Kleiman, *Minnesota Judicial Workload Assessment. Final Report,* National Center for State Courts, Williamsburg, VA, 2010.

Ostrom, B. J., Kleiman, M. & C.G. Lee, *Kentucky Judicial Workload Assessment*, National Center for State Courts, Williamsburg, VA, 2016.

Ostrom, B.J., Kleiman, M.C., Lee G. & S. Roth, *Florida Judicial Workload Assessment*, Final Report, National Center for State Courts, Williamsburg, VA, 2016.

Ostrom, B. J., Kleiman, M., L& S. Roth, *Virginia Judicial Workload Assessment*, National Center for State Courts, Williamsburg, VA, 2017.

Pace, N. et al, *Case-weights for Federal Defender Organizations*, Rand Corporation, Santa Monica, California, 2011.

Sayles, D., S. Tallarico & J. Douglas, *Indiana Caseload Assessment Plan to Utilize Resources Efficiently*, National Center for State Courts, Williamsburg, VA, 2016.

Tallarico, S., *Kansas District Court Judicial and Clerk Staff Weighted Caseload Study,* National Center for State Courts, October 31, 2011.

Tallarico, S., Douglas J. & E. Friess, *Judicial Workload Assessment Superior Court of Arizona*, Mohave County, National Center for State Courts, Williamsburg, VA, 2013.

Tallarico, S., Douglas J. & E. Friess, *Tennessee Trial Courts Judicial Weighted Caseload Study*, National Center for State Courts, Williamsburg, VA, 2013.

Tallarico, S., Douglas J.& E. Friess, *Montana District Court Judicial Weighted Caseload Study*, National Center for State Courts, State and State Justice Institute, Williamsburg, VA, 2014.

Tallarico, S., Friess E. & J. Douglas, *Vermont Trial Court System Judicial Officer and Court Staff Weighted Caseload Study*, National Center for State Courts, State Justice Institute, Williamsburg, VA, 2015.

Tallarico, S. & A. Davis, *Oregon Circuit Court Judicial Officer Workload Assessment Study*, National Center for State Courts, Williamsburg, VA, 2016.

Tippett L.H.C., statistical methods in textile research. part 3—a snap-reading method of making time-studies of machines and operatives in factory surveys," Journal of the Textile Institute Transactions 26, no. 2, February 1935.

Washington State Office of Public Defense, *Model Misdemeanour Case Weighting Policy*, 2014.

Weinshall-Margel, K., Galon I. & I. Taraboulos, *Case Weights for the Assessment of Judicial Workloads in Israel*, The Israeli Courts Research Division, 2013.

Winkler D. et al., "Combining a weighted caseload Study with an organizational analysis in courts: first experiences with a new methodological approach in Switzerland", *International Journal of Control and Automation,* 7, no.1, July 15, 2015: 27.

Wittrup, W., *Benefit of doubt approach to case weighting. Improving and simplifying the assessment of workload in courts*, Copenaghen Business School, July 2017.

The Study on case weighting in judicial systems provides an overview of this internal management tool aimed at assessing the "weight" of a court case, by detailing, by various methods, the court time necessary for its treatment. The information deriving from case weighting can be used to optimise the allocation of human and budgetary resources, cases among judges or to support a broader service management programme within courts. Based on the in-depth analysis of the existing bibliography on the subject as well as on the methodology used in 6 European countries, the Study formulates 12 recommendations for the setting up of case weighting. The first recommendation invites judges to participate actively in the process. The Study was prepared by the Working Group on judicial time management (CEPEJ-SATURN) on the basis of an analysis carried out by two experts – Shanee Benkin (Israel) and Marco Fabri (Italy).

cepej

European Commission for the Efficiency of Justice

Commission européenne pour l'efficacité de la justice

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE