Legal Metrics Extracted from Court Decisions. A Focus on Personal Injury Compensation

Christophe Quézel-Ambrunaz* and Vincent Rivollier**

Abstract

This article explores the extraction of legal metrics from court decisions, with a focus on personal injury compensation in France. It begins by discussing the challenges of accessing and structuring data from judicial decisions, highlighting legal and non-legal barriers. Despite recent legislative efforts to open up access to judicial decisions as open data, significant obstacles remain in the extraction and processing of relevant information. The article delves into the specific case of personal injury compensation, where empirical and quantitative approaches have been widely utilised due to the absence of official guidelines and the diverse nature of compensation methods. It also discusses the failure of the Datajust project, which aimed to create a tool for modeling judges' decisions on personal injury compensation but was ultimately abandoned due to technical and regulatory challenges. Looking ahead, the article discusses prospects for the future of legal metrics, including ongoing government initiatives to improve access to judicial data and harness artificial intelligence for case orientation. It also highlights the potential of reforms in civil information systems, such as the Portalis project, to provide new insights and standardise the structure of court decisions.

I.   Access to Data Derived from Judicial Decisions

The book at the core of this symposium deals with 'legal metrics' in a variety of different settings. In what follows, we will complement the analysis offered in the volume with a study of the potential benefits and actual problems arising out of extracting quantitative data from judicial decisions, using personal injury compensation in France as a case study.

In order to use data derived from judicial decisions, the first step is to gain access to these decisions, and then to extract sufficiently structured data from them.[1] In France, judicial decisions are made in the name of the French people. Strictly speaking, there is no copyright on to their content. Any interested party can request a copy of a judgment from the court registry. In principle, there is nothing to prevent judicial decisions from being freely accessible in France. In practice, however, open data on judicial decisions has not been the norm,[2] and to date, such access remains

---

* Professor, University Savoie Mont Blanc, Institut universitaire de France.
**Associate Professor, University Savoie Mont Blanc, Centre national de la recherche scientifique.
[1] I. Sayn and V. Rivollier eds, *Justice et numérique, Quels rapports?* (Chambéry: Les cahiers de jurimétrie, Presses de l'Université Savoie Mont Blanc, 2024).
[2] A. Dunes, 'La non-publication des décisions de justice' *Revue internationale de droit comparé*,

partial.[3] There are both legal and non-legal reasons for this.

From a legal point of view, data protection now prevents information allowing the identification of parties from being freely accessible. In the past, reproductions of judgments in law journals or in computer databases have always omitted data such as addresses, but names were present and indeed used to identify landmark judgments. Now, pseudonymisation and redaction measures are carried out directly by a service of the Court of Cassation, removing names, first names, addresses, dates of birth, and dates of death.

From a non-legal perspective, the production of judicial decisions was not designed for the creation of databases.[4] There is heterogeneity in practices, both in the software used for drafting and in the conditions for storing decisions. As a result, no one, not even the Ministry of Justice, has access to all the decisions handed down in France. Thus, the computerisation of courts and tribunals does not automatically lead to open access to judicial decisions.[5]

Another element of complexity must be added: in France, two orders of jurisdiction coexist, the administrative order and the judicial order. These two orders, completely independent and separate, have different practices regarding the digitisation of their practices and the dissemination of their decisions. This article will focus on the judicial order, as it undoubtedly represents the largest pool of data to be exploited. In fact, every year the number of judicial decisions exceeds the number of administrative decisions. Furthermore, they are more diverse, both in terms of their subject matter and the data they contain. In the administrative order, a large part of the litigation concerns appeals against administrative decisions, resulting in a decision to annul or not to annul an administrative act. Therefore, data exploitation would essentially be Boolean for the administrative order, whereas more quantitative approaches would be possible in the judicial order.

It was a political will, driven in particular by legal publishers and newcomers to the legal access market, that determined the opening up of judicial decisions as open data.[6] These are public data that must be open by default, according to the provisions of the Digital Republic Act of 2016.[7] Given the sensitive nature of

---

757 (1986).

[3] E. Serverin, 'Plaidoyer pour l'exhaustivité des bases de données des décisions du fond (à propos de l'ouverture à la recherche de la base JuriCa)' *Recueil Dalloz*, 2882 (2009); Y. Meneceur, 'Open data des décisions de justice. Pour une distinction affirmée entre les régimes de publicité et de publication' 37 *La Semaine Juridique Édition Entreprise et affaires*, 31 (2019).

[4] A.-J. Arnaud, 'Le droit, un ensemble peu convivial' 11-12 *Droit et société*, 79 (1989).

[5] C. Bordere, 'Que reste-t-il de la première jurimétrie?' *Jurimetrics. Journal of the Measurement of Legal Phenomenon*, 27 (2022).

[6] G. Deroubaix, 'L'édition juridique et la diffusion du droit. La problématique particulière de la diffusion de la jurisprudence' *La Semaine Juridique Édition générale*, supplément au n 9, 92, 93 (2017).

[7] I. Sayn, 'L'accès aux documents issus des activités des autorités publiques dans le monde du droit et de la justice', in Id and V. Rivollier eds, n 1 above, 21; L. Cadiet ed, 'L'open data des décisions de justice - Mission d'étude et de préfiguration sur l'ouverture au public des décisions de justice', *Rapport à Mme la garde des Sceaux*, 75 (November 2017).

the data contained in judicial decision, a precise framework was needed. This was provided by statute 23 March 2019 no 2019-222, and by decree 29 June 2020 no 2020-797, and decree 30 September 2021 no 2021-1276. A decree of 28 April 2021 provides the schedule for the gradual online publication of decisions. This schedule has generally been adhered to: at the time of writing this article, amongst the judicial decisions, only decisions rendered by the Court of Cassation, and those rendered by civil courts of appeal, as well as decisions of some judicial courts in civil matters, are available online. By the end of 2025, decisions in criminal matters, and all decisions of the judicial courts, should be added. The pseudonymised data are accessible either through an interface with a search engine for the general public or through an API that allows third-party applications to access the decisions.

However, obtaining decisions is only the first step; it is also necessary to know how to process them. Undoubtedly, legal science today is at a rare moment of methodological transition. It is possible to hypothesise that legal doctrine takes as its main object of study the sources it which it has access. Looking back, from the perspective of the French jurist, from the rediscovery of Roman law to the Napoleonic codification, the available material was essentially the *Corpus Juris Civilis* and the glosses made on it: it was the age of pandectism. In the 19th century, the Civil Code adopted in 1804 was the main object of study: it was the age of exegesis. The 20th century saw the dissemination, notably through the creation of the *Revue Trimestrielle de droit civil*, of the case law of the Court of Cassation: the commentary on rulings then were prevalent. In the 21st century, the mass of decisions of first instance is now available for doctrine; these decisions show the living law in its ecosystem.[8] The tools and methods are still to be perfected, but it can be wagered that this century will be that of the analysis of the mass of decisions.

This availability of litigation should not obscure the fact that a large part of compensations, particularly in cases of personal injury cases, results from transactions that are not freely available. Figures are uncertain and difficult to obtain, but it is said that 90% to 95% of personal injury cases end in a transaction. However, this would mainly involve small claims that, so that significant damages would often be decided by the judge. In fact, in the case of traffic accidents, insurers are obliged to make a statistical file available to the public, including transactions.[9] They only fulfil this obligation minimally, and the file is notoriously insufficient. Moreover, it is very difficult to exploit for statistics or legal analysis.

Returning to the issue of decisions, once they are available to researchers or the public, the first question to be asked is what can be done with them.[10] The

---

[8] M.-A. Frison-Roche and S. Bories, 'La jurisprudence massive' *Recueil Dalloz*, 287-290 (1993); F. Rouvière, '*Jus ex machina*: la normativité de l'intelligence artificielle' *Revue trimestrielle de droit civil*, 217-219 (2019).

[9] Loi 5 July 1985 no 85-677 tendant à l'amélioration de la situation des victimes d'accidents de la circulation et à l'accélération des procédures d'indemnisation, *Journal Officiel de la République Française* (6 July 1985).

[10] I. Sayn, 'Des modes algorithmiques d'analyse des décisions de justice, pour quoi faire?' 5

primary purpose of a judicial decision is not to be an object of knowledge but rather a legal instrument intended to produce legal effects. Nevertheless, the growing importance of case law as a source of law has led to a reconsideration of its status: a judicial decision is of interest not only to the parties involved, but also to researchers or litigants, since it contains, at least potentially, a norm. It is therefore a matter of identifying and retrieving it. The interesting data then becomes what has been called the 'titling' of the decision: the identification of the jurisdiction, the chamber, the date, the subject, the key words, and possibly, through coding, certain legal concepts that have been retained or not.

This was the focus of what could be called the 'first jurimetrics'.[11] Given with the complexity of law, the challenge – which is still relevant today – was to preserve its comprehensibility. The primary task was to create files of decisions and to index them. Access to relevant decisions was an important issue, so a search system was needed. However, the trend soon moved towards extracting data from trial courts, especially appellate courts, in order to produce statistics.

In this initial phase, the university initiatives were quickly superseded by private databases – the most prominent being *Jurisdata*. For technical and methodological reasons, abstracts took precedence over full texts. Although *Legifrance*[12] popularized access to the full text, it was initially only for decisions of the Court of Cassation and a few appellate court decisions. The larger a full-text search engine traverses a significant collection of data, the more efficient it needs to be to rank results in a relevant order.

Legal publishers and legaltech companies have significantly worked to make search results more relevant and easier to navigate.[13] They have regularly improved their search engines. In addition, many have developed statistical databases in various areas of law to provide numerical information on certain damages or compensation, or even to estimate the probability of obtaining certain amounts in certain jurisdictions.[14] This second 'jurimetrics', or second generation jurimetrics, was probably more the work of private actors selling specific services to users.

The first name given to these services was '*justice prédictive*'; probably a poor translation of the English 'predictive justice'. A better French translation would have been '*justice prévisible*'. A certain anxiety gripped the legal professional,[15] repelled by the fear or fantasy of a 'robot judge' figure, replacing judicial debate.[16]

Management & Datascience (2021), available at https://tinyurl.com/3kx83jkm (last visited 30 September 2024).

[11] C. Bordere, n 5 above.

[12] I. Sayn, 'L'accès aux documents' n 7 above.

[13] ibid

[14] G. Zambrano, 'Précédents et prédictions jurisprudentielles à l'ère des gid data: parier sur le résultat (probable) d'un procès', 6 (2015) available at https://tinyurl.com/yc44xt7w (last visited 30 September 2024).

[15] N. Roret and G. Accomando, 'Avocats et magistrats en 2050: quelle justice demain?' 2 *Revue pratique de la prospective et de l'innovation*, 5-6 (2021).

[16] B. Barraud, 'Le droit en datas: comment l'intelligence artificielle redessine le monde juridique.

According to a recommendation from the National Bar Council, the term 'jurimetrics' should be preferred. This semantic shift is partly questionable. Indeed, it mainly concerns the use of natural language processing algorithms,[17] statistical analysis, and probability calculations. It is certainly about jurimetrics. However, if one refers to the definition of jurimetrics adopted when was created Jurimetrics, Journal of the Measurement of Legal Phenomena, jurimetrics is much broader.[18] The quantification of legislative production, linguistic research with a quantitative aspect on judicial decisions or laws, metric analysis of the quality of the law, and many other elements, undoubtedly belong to jurimetrics, but not in the restricted sense given by the National Bar Council,[19] replacing '*justice prédictive*'.[20] There is a tension here between two ways of integrating mathematical analysis into law: the one hand, the simple search for knowledge of existing law;[21] on the other hand, the search for a transformation of the production and implementation of law.[22] Legal metrics can thus be directed either towards the past or towards the future.

The volume of judicial decisions is considerable: according to the Ministry of Justice figures,[23] in 2022, 1,872,458 decisions will be rendered in civil and commercial matters (excluding criminal matters), and 281,405 cases will be settled by administrative courts. Focusing on civil matters and new cases in 2022, the figures are as follows: 15,479 decisions by the Court of Cassation. Therefore, it is reasonable to assume that a researcher or practitioner focusing on a particular type of litigation, which represents a small percentage of the total litigation, has the possibility of becoming aware of all the decisions of the Court of Cassation in their field. For appellate courts, there were 196,261 decisions. Even in a limited field, it is unrealistic for one person to read all decisions in their field of expertise, but the analysis is still

Partie II: Les nouvelles technologies juridiques ou l'intelligence artificielle au service du droit' *Revue Lamy droit de l'immatériel*, 44-50 (décembre 2019); J.-H. Stahl, 'Le juge, le robot et la boule de cristal' 1 *Droit administratif* (août 2019), 1; C. Pavillon, 'Justice alternative et numérique: des expériences mitigées aux Pays-Bas' *La Semaine Juridique Édition générale*, 51-55 (17 décembre 2019); C. Byk, 'L'humain prédit par le droit: un chemin sinueux entre illusion et nécessité' 3 *Revue de la Recherche Juridique*, 995-1001, (2018); A. Coletta, '*La prédiction judiciaire par les algorithmes*' (Thèse de doctorat Sciences Juridiques, Université De Nîmes, 2021).

[17] F. Muhlenbach, 'Les techniques d'intelligence artificielle applicables au domaine juridique', in I. Sayn and V. Rivollier eds, n 1 above, 37.

[18] C. Quézel-Ambrunaz and V. Rivollier, 'For a reasoned study of jurimetics' 7 *Jurimetrics. Journal of the Measurement of Legal Phenomenon* (2022).

[19] Groupe de travail Legaltech, Conseil national des barreaux, *Legaltechs du domaine de la jurimétrie, préconisations d'actions*, Report 9 octobre 2020, 9.

[20] C. Quézel-Ambrunaz, 'À la recherche d'une définition de la jurimétrie' 15 *Jurimetrics. Journal of the Measurement of Legal Phenomenon* (2022).

[21] F. Colonna d'Istria, 'La possibilité d'une objectivité interne dans la connaissance du droit' 2 *Revue interdisciplinaire d'études juridiques*, 109-130 (2007).

[22] A. Supiot, '*La Gouvernance par les nombres*' (Cours au Collège de France, Fayard, collection Pluriel, 2020); S. Deakin, 'Droit et statistique: représentation mathématique des lois; méthodologie de l'analyse empirique du droit' (2019), available at https://tinyurl.com/2dftzvsf (last visited 30 September 2024).

[23] https://tinyurl.com/ypknz7e8 (last visited 30 September 2024).

within the reach of well-structured teams. For judicial tribunals, there were 1,452,693 decisions. This mass of decisions can only be exploited with the use of artificial intelligence.[24]

One application of artificial intelligence is indeed to extract data from the mass of these decisions, or even to exploit it. This is not the only possible application: an actor uses artificial intelligence with neural networks to simulate the judge's reasoning and indicate the probabilities of success or failure of a claim based on the parameters of that claim. Returning to the use of artificial intelligence to extract and exploit decision data, several initiatives have been launched. For example, legal publishers offer, probabilities of success for certain claims or averages of amounts that can be obtained in court, depending on to the jurisdiction of the appellate court, based on the analysis of judicial decisions. These aggregated data provide plausible orders of magnitude, but they are not verifiable – they do not allow for the comparison of each judicial decision with the data extracted from it.

Indeed, extracting data from judicial decisions proves to be a perilous exercise.

II.   The Difficulties of Extracting Legal Metrics from Judicial Decisions

Judicial decisions do not constitute a structured dataset that can be processed directly.[25] They are merely the outcome of a judicial process, offering only a brief, imperfect, and distorted glimpse of it.

Resolving a dispute often means accepting either the plaintiff's or the defendant's position. However, even if they have submitted pages of conclusions or pleaded at the hearing, which is increasingly rare, the judicial decision often only reflects the claims of the parties – what they want from the judge – rather than their arguments, which can have significant informative value. Stylistic clauses are often used to prevent the decisions from being censured by the Court of Cassation, but they do not provide any insight into how the court reached its decision.

Furthermore, decisions inevitably overlook unlawful or even illegal determinants.[26] For example, in divorce cases, the respective faults of the spouses during the marriage should not, in principle, influence either the compensatory allowance that the wealthier spouse owes to the other or the child support payment. Nevertheless, can we be absolutely sure that moral considerations will never influence the determination of such sums? Similarly, although it is well-known that guidelines or calculation software exist for determining these sums, and they are used by both plaintiff and defence lawyers as well as judges, decisions do not

---

[24] F. Muhlenbach, n 17 above.

[25] J. Barnier et al, 'Extraire des informations fiables des décisions de justice dans une perspective prédictive: des obstacles techniques et des obstacles théoriques' 89 *Jurimetrics. Journal of the Measurement of Legal Phenomenon* (2022).

[26] P. Brunet. 'Analyse réaliste du jugement juridique' 147 *Cahiers philosophiques*, 9-25, (2016); S. Danziger et al, 'Qu'a mangé le juge à son petit-déjeuner? De l'impact des conditions de travail sur la décision de justice' 4 *Les Cahiers de la Justice*, 579-587 (2015).

reference to such tools.

In France, there is no prescribed style for judgments to be followed by every judge. The way in which judgments are written and presents is unique to each judge, resulting in considerable heterogeneity in the way judgments are presented.[27] For example, the claims of the parties may be restated at the beginning of the decision but not in the reasons, or they may be predominantly developed in the reasons. It is also possible that, where several amounts are awarded to the claimant on different grounds, the *dispositif* (operative part) may either contain detailed breakdowns or only the total amount. The only commonality among all decisions is their structure, typically starting with the presentation of the parties, followed by the facts, then the reasons, and finally the *dispositif*. The existence of 'mandatory passages' – for example, the reasons and the *dispositif* are separated by the formula '*par ces motifs*' (for these reasons) – allows an expert system to structure the decisions into their main parts quite easily.[28]

The pseudonymisation and obfuscation of certain date makes it difficult, if not impossible, for even a human expert in the field to extract certain data – and is a major concern for researchers.[29] In the course of certain studies, it has been revealed that in some published decisions, it was impossible to determine the sex or age of individuals, even though this information may seem essential for extraction – not to mention the exact date of death, which is always obscured. Another difficulty related to online publication is that decisions sometimes include tables, in particular to compare the claims and offers of each party. However, the computational treatment applied does not recognise these tables, which then become a series of unintelligible figures.

These are just some difficulties that can be arise and mislead even excellent artificial intelligence systems used to extract and structure data from judicial decisions. It may be worth noting that France is only partially in the eurozone, and New Caledonia, Wallis and Futuna, and French Polynesia use the CFP franc. A further difficulty is that, in many disputes involving sums of money, the sums at issue may be only a part of the total, whereas the total is important. These may involve liability cases where the victim's fault reduces the liability of the responsible party, cases of loss of chances, cases where only one spouse's interests in community property are at issue, or cases where only one co-owner disputes the value of a right. It needs to be clarified what data should be extracted in such cases: is it the fraction of the right in dispute or the full value of that right?

Procedural rules further complicate data exploitation. Like many legal systems,

---

[27] J.-P. Ancel, 'La rédaction de la décision de justice en France' 841 *Revue internationale de droit comparé* (1998); P. Mimin, '*Le style des jugements (Vocabulaire – construction – dialectique – formes juridiques)*' (Paris: Librairies techniques, 4th ed, 1978).

[28] Even if, in first instance, some judgements use other words to introduce the *dispositive*, see. J. Barnier, 'Extraire automatiquement des informations de décisions des juges aux affaires familiales?', in I. Sayn and V. Rivollier eds, n 1 above, 49.

[29] I. Sayn, 'L'accès aux documents' n 7 above.

French law holds that the dispute belongs to the parties, a principle known as *principe dispositive*.[30] While judges have certain powers, such as restoring a claim to its correct legal basis or requalifying a legal act, they are limited by this *principe dispositif*. In particular, they cannot rule *infra petita* or *ultra petita*; in other words, their decision must necessarily fall within the bounds set by the claim and the offer. How should be treated a decision in which a judge deems that a head of damage is not to be compensated because it is insufficiently characterized, but nevertheless awards damages to the extent of the offer made on that head? How to treat a decision that calculates compensation but allocates a lower sum to limit it to the claim? Furthermore, the parties are subject to constraints such as time limits and the concentration of resources. It is necessary to ensure that, when data is extracted, a distinction is made between a point that has been discussed on the merits and then rejected, and another that has only been rejected on the basis of a procedural exception, without having been discussed on the merits.

It is often difficult to identify the relevant data that needs to be extracted.[31] For example, in the case of penalties, the amount per day of delay is certainly more informative than the total amount of the penalty, which depends on factual elements such as the promptness with which the debtor complied, but also on the possible use of the judge's power of moderation. If the amounts can be paid either as an annuity or as a lump sum, there may be some hesitation. Sometimes, for instance in the case of *viager* transactions, the annuity appears only as a way of staggering the payment of a lump sum, and the most important variable is undoubtedly the lump sum. In other cases, such as personal injury compensation, the lump sum is only one method of payment, and the interesting data is the annuity awarded.

The procedural difficulties are also evident in other areas. Although the appeal is said to have a devolutionary effect, it does not necessarily cover the entire decision rendered at first instance. Thus, when data is extracted from appellate decisions, it is possible that they are only partial, as the appellant may have been satisfied with a part of the first judge's decision and limited his appeal to certain points. Even if an appellate decision covers all the points decided at first instance, the appellate court may simply write that it upholds what the first judges decided, without repeating the substance of that decision. Thus, extracting relevant data from the appellate judgment actually requires searching for data in the first instance judgment.

Working on the basis of first instance decisions present other difficulties: there is no indication of whether a judgment has been appealed, or what the outcome of that appeal was. The only solution would be to monitor appeal decisions in order to look for references to judgments and make connections between the information, but this involves a time lag: this consolidation of data on judgments can only take place after several years! Therefore, using first-instance data either risks working

---

[30] Code de procédure civile, Art 5.

[31] B. Barraud, 'Le coup de data permanent: la loi des algorithmes' 35 *Revue des droits et libertés fondamentaux*, (2017) available at https://tinyurl.com/td573jna (last visited 30 September 2024).

with data invalidated on appeal or using outdated data. The same reasoning applies to appellate judgments that may be overturned by the Court of Cassation.

Even at a more macroscopic level, the transition from one level of jurisdiction to another is problematic. First and second instances courts use a classification system called the Nomenclature of Civil Affairs (NAC), which could certainly be improved, but which categorises cases according to their main area. This at least makes it possible to analyse of the importance of litigation in different areas of law and to follow this importance between the first and second levels of jurisdiction. It is thus possible to determine, for example, the areas with the highest appeal rates. However, until recently the Court of Cassation does not use this same nomenclature, making it impossible to track case volumes by litigation domain.[32]

The data that can be extracted from judicial decisions is vast; however, it is difficult to speak of big data comparable to what, for example, consumer applications or websites, or even certain connected objects, can collect from users. To predict the time it will take to drive between two cities by car, well-known application algorithms can rely on data from thousands of journeys between those cities. However, to predict the amount of alimony a spouse will be awarded in a particular appellate jurisdiction, with contextual elements such as a certain length of marriage, a certain family structure, a certain income level, a certain age, etc, it will not be possible to rely on thousands of similar decisions. In general, attempts at prediction in the legal field have been disappointing.[33]

This situation should not discourage data analysis initiatives. It seems that the main obstacle to conducting jurimetrics on data from judicial decisions is the fact that the data from these decisions is not structured. Perhaps in the future, public initiatives will encourage the courts to add more metadata to their decisions using specialised software, or even to structure them in a way that not only facilitates execution, but also enables easy data exploitation.[34]

III. What Knowledge to Draw? The Example of Compensation for Personal Injury

Compensation for personal injury is one of the areas of law where data can be extracted from court decisions in the form of numerical and monetary values. In

---

[32] B. Munoz-Perez and E. Serverin, 'Éléments pour une statistique qualitative des affaires civiles traitées par la Cour de cassation' 87 *Cour de cassation*, (2020), available at https://tinyurl.com/ywrk2bpr (last visited 30 September 2024); E. Serverin et al, 'La Nomenclature des affaires orientées dans les chambres civiles de la Cour de cassation (NAO): l'élaboration collective d'un outil de connaissance et d'action' 130 *Cour de cassation*, (2021), available at https://tinyurl.com/4ren9aze (last visited 30 September 2024).

[33] J. Barnier, 'Extraire automatiquement' n 28 above; B. Jeandidier, 'L'hétérogénéité des décisions de justice réduit leur prévisibilité', in I. Sayn and V. Rivollier eds, n 1 above, 65.

[34] M. Cottin, 'Vers la standardisation dans la rédaction des décisions de justice?', in I. Sayn and V. Rivollier eds, n 1 above, 147.

fact, compensation is not regulated by any official scale, and its level is only weakly controlled by the higher courts, especially regarding non-pecuniary damages. The reality of compensation can therefore only be known by analysing court decisions of the first or second instances. The large number of court decisions available requires the implementation of empirical methods and, in particular quantitative analyses.

For this reason, empirical and quantitative approaches have been particularly mobilised in research on personal injury compensation. The reasons are manifold. Firstly, the very subject of the dispute lies in the calculation of compensation, which is divided into numerous heads of damage that are assessed separately. Quantitative approaches are also possible due to the significant number of court decisions in this area and to the compensation methods used. These methods allow the objectification of the damage and the comparison of very different concrete situations: physical and psychological injuries are described in percentages or on a seven-degree scale. Nomenclatures of damages and unofficial compensation guidelines tend to standardise methods and compensation amounts. Furthermore, the determination of the compensation amount may depend on actuarial methods, particularly concerning the capitalisation of annuities. Thus, mathematical approaches to personal injury compensation and empirical quantitative methods have regularly been used. More exceptionally, qualitative methods have also been used.

Through this section, we propose an analysis of empirical studies in personal injury compensation matter in France. To do so, we will take into account all the studies that have come to our attention in this field in recent years.[35] We will describe, on the one hand the data mobilised (1) and, on the other hand, the knowledge derived from this data (2).

1. The Analysed Data

Most of the studies were based on data extracted from court decisions, but

---

[35] E. Serverin et al, 'L'accident corporel de la circulation, entre transactionnel et juridictionnel' *Report, Ministère de la Justice* (1997); S. Porchy-Simon et al eds, 'Étude comparative des indemnisations des dommages corporels devant les juridictions judiciaires et administratives en matière d'accidents médicaux' *Rapport pour la Mission de recherche Droit & Justice* (2016); L. Carayon et al, 'Réflexions autour du préjudice sexuel. Analyse de jurisprudence sous l'angle du genre' *Recueil Dalloz*, 2257 ff (2017); C. Quézel-Ambrunaz et al, 'De la responsabilité civile à la socialisation des risques: études statistiques' *Rapport dans le cadre du projet ANR RCSR*, (2019); N. De Jong, 'L'indemnisation du dommage corporel. Les barèmes dans les décisions de justice de première instance', in I Sayn et al eds, *Les barèmes (et autres outils techniques d'aide à la décision) dans le fonctionnement de la justice* (Mission de recherche Droit & Justice, 2019), 75; C. Quézel-Ambrunaz, 'La réparation des préjudices laissés par les cicatrices. Étude statistique' *Recueil Dalloz*, 2248 ff (2020); E. Belz et al, 'Bodily Injury Claims in France: Negociation or Court?', in Id, *Économétrie des données imparfaites: méthodes et applications*, 81 ff (thèse Rennes 1, 2021); C. Quézel-Ambrunaz, 'Demandes, offres, décisions en matière de dommage corporel: étude statistique' *Report, Institut Universitaire de France*, (2021); V. Rivollier, 'Le montant de l'indemnisation du préjudice d'affection devant les cours d'appel. Essai de mesure de l'influence du montant fixé en première instance, du montant demandé et du montant offert sur l'indemnisation devant le juge d'appel' *Jurimetrics. Journal of the Measurement of Legal Phenomenon*, 107-125 (2022).

data from other areas can also be mobilised.

a) Data from Court Decisions

Until recently, access to decisions of lower courts has not been easy. Studies seeking a form of comprehensiveness have therefore been limited by the inherent limitations of existing databases. Several studies[36] have relied on two officials but non-public databases: JuriCa database, which collects decisions of civil chambers of judicial appellate courts,[37] and Ariane Archives database, which collects decisions of administrative courts.[38] While decisions of first instance courts were available in administrative matters, they were not available in judicial matters. Therefore, studies were generally limited to appellate decisions in both types of courts in order to maintain a parallelism between the two corpora.[39]

Other studies, either because they did not have access to these two databases or because they sought to study first instance court decisions, used other paths to access decisions. Court decisions could be obtained from commercial databases,[40] directly from local courts[41] or from law firm.[42] There is no exhaustivity and no guarantee of representativity, so the robustness of the results is lower, but the scarcity of such studies, especially on first instance court decisions, maintains their interest.

The provision of open data on court decisions is likely to encourage other studies based on a more complete dataset. Since the end of 2024, decisions from eight courts of first instance have been made available.

Studies based on this material suffer from several limitations. There is currently no access to decisions from criminal courts, and the open data for these decisions will be available later. However, a significant proportion of personal injuries claims are compensated in these courts.[43] There are also technical limitations: automated analysis of decisions using natural language processing techniques only allows for the extraction of relatively simple data. A thorough analysis must necessarily

[36] S. Porchy-Simon et al, n 35 above; C. Quézel-Ambrunaz et al, n 35 above; and, taking up and completing the data from the previous research, V. Rivollier, n 35 above. Using a commercial database, itself fed by the JuriCa database, C. Quézel-Ambrunaz, n 35 above.

[37] E. Serverin, 'Plaidoyer' n 3 above; S. Bories, 'JuriCA: un outil de communication et de recherche' *RecueilDalloz*, 1242, (2011); X. Henry, 'Vidons les greffes de la République! De l'exhaustivité d'accès aux arrêts civils des cours d'appel' *Recueil Dalloz,* 2609 ff (2011).

[38] F. Alhama, 'Vers une plus grande accessibilité des décisions rendues par les juridictions administratives' *Revue française de droit administratif*, 695 (2019).

[39] S. Porchy-Simon et al, n 35 above; C. Quézel-Ambrunaz et al, n 35 above; V. Rivollier, n 35 above. Taking into account only the decisions of certain courts of appeal: C. Quézel-Ambrunaz, n 35 above.

[40] L. Carayon et al, n 35 above.

[41] N. De Jong, n 35 above. This is an exhaustive study of the decisions of three first instance judicial courts, over a given period and for certain categories of the civil cases.

[42] C. Quézel-Ambrunaz, n 35 above.

[43] Only one study included certain decisions on civil matters handed down by criminal courts, obtained directly from lawyers: C. Quézel-Ambrunaz, n 35 above.

be done manually.[44]

b) Other Data from the Field

In order to overcome the limitations of access and content of court decisions, some studies have examined the field of personal injury compensation using other data. Statistical data can be used, in particular from the registry of civil cases maintained by the Ministry of Justice. However, access to this data requires the involvement of the Ministry in the study.[45]

Other data are collected and held by insurance companies, which are often involved in personal injury compensation. This data is mostly confidential. However, in the case of road traffic accidents, the *Loi Badinter*[46] has established a special regime for compensation and liability and defines a mandatory compensation offer process to promote out-of-court settlement. This Act notably provides that 'under the control of the public authority, a periodic publication reports compensations fixed by judgments and settlements'.[47] This publication is implemented by the Association for the Management of Information on Automotive Risk (*Association pour la gestion des informations sur le risque automobile* – AGIRA)[48] and takes the form of an online database.[49] The quality of the information available to the public is limited: only certain heads of damages are reported; the precise age of the victim is unknown; when the compensation is judicial, the jurisdiction that ruled is unknown; no precise description of the damage is given (only the degrees of certain medico-legal scales are presented). Even if the data are limited, they are so numerous that their use can be valuable.[50] Another study, albeit old, had access to complete insurance files on compensation for personal injuries resulting from traffic accidents. This study was thus able to analyse in depth the compensation process and the determinants of its transactional or judicial orientation.[51]

Data may also come directly from professionals working in the field of personal injury law, in particular judges and lawyers. They are then collected through questionnaires or interviews but do not allow for a quantitative approach.[52]

---

[44] J. Barnier et al, 'Extraire des informations' n 25 above; Id, 'Extraire automatiquement' n 28 above.

[45] E. Serverin et al, n 35 above.

[46] Loi 5 July 1985 no 85-677 tendant à l'amélioration de la situation des victimes d'accidents de la circulation et à l'accélération des procédures d'indemnisation.

[47] Art 26 loi 5 July 1985 no 85-677.

[48] On the creation of the AGIRA file, see E. Serverin et al, n 35 above, 81 ff.

[49] Available at https://formulaire.victimesindemnisees-fvi.fr/(last visited 30 September 2024).

[50] E. Belz et al, n 35 above. The author acceded more completed data than the online database.

[51] E. Serverin et al, n 35 above.

[52] I. Sayn et al, n 35 above (interviews with magistrates); C. Quézel-Ambrunaz et al, n 35 above (questionnaire with all types of professionals involved in personal injury cases, the majority of respondents being lawyers).

2. Knowledge Extracted from Data

The analysis of court decisions alone does not allow any conclusions regarding the transactional or judicial orientation of compensation. However, such knowledge can be derived from studies analysing insurers' data on traffic accidents. These studies come to a similar conclusion: the proportion of out-of-court settlements decreases with the severity of the personal injury; the more serious the injury, the less compensation is settled out of court.[53] Other determinants of the type of settlement have been identified, including the age of the victim and the geographical area. Other comparisons have been made between the two types of settlement, in particular with regard to the time between the accident and compensation (longer when the resolution is judicial) or regarding the compensation amounts (higher in courts).[54]

Other studies, based only on the analysis of court decisions, look at judicial practice in terms of the depth of compensation for personal injuries. Depending on the corpus studied, a comparison can be made between different jurisdictions and orders of jurisdiction can be made.[55]

Firstly, studies have examined the structure of compensation, ie, the division of damages into different headings, especially since the introduction of the Dintilhac nomenclature. One study noted a difference between judicial courts of appeal and administrative courts of appeal. The former largely applied the nomenclature, unlike the latter, whose practice of grouping or venting heads of damages appeared highly heterogeneous.[56] The subsequent studies had no longer observed such a difference[57] since the evolution of administrative case law led to accept the application of this nomenclature.[58]

The studies are significantly devoted to the analysis of compensation amounts and their determinants. These determinants relate firstly to the characteristics of the injuries, in particular their severity, as medically assessed by the expert and the court. However, other criteria are also considered. The age and gender of the victims influence compensation, especially because women receive less compensation than men.[59] Courts and geographical variations are also examined: in certain geographical areas, compensation appears to be higher than elsewhere,[60] and

---

[53] E. Serverin et al, n 35 above; E. Belz et al, n 35 above.

[54] ibid

[55] Particularly in the area of medical accidents. The judicial courts have a monopoly on actions for compensation of the consequences of traffic accidents.

[56] S. Porchy-Simon et al, n 35 above: the decisions studied in the corpus were pronounced in 2011, 2012 and 2013.

[57] C. Quézel-Ambrunaz et al, n 35 above.

[58] Conseil d'État, 16 December 2013, no 346575, reversing the 'Lagier opinion' in which the *Conseil d'État* had proposed a different nomenclature of damages (Conseil d'État, Section du contentieux, 4 June 2007, nos 303422 and 304214).

[59] L. Carayon et al, n 35 above; C. Quézel-Ambrunaz et al, n 35 above; E. Belz et al, n 35 above.

[60] E. Belz et al, n 35 above, establishes the link between higher amounts of compensation in out-of-courts settlement compensation and an higher rate of out-of-courts settlements. See

administrative courts award lower amounts than judicial courts.[61] The influence of procedural rules has also been examined; indeed, the court cannot, in principle,[62] determine the compensation amount beyond the claims of the parties. The exaggeration of the parties' claims could have a deterrent or repellent effect on the judge's behaviour.[63] However, except in extreme cases, this effect has not been found, and regardless of the parties' demands, appeal courts frequently award the same amount as that awarded at first instance.[64]

Several studies have also examined the influence of soft law instruments on the amounts awarded by courts. These tools can be either compensation guidelines or capitalisation scales.

Court decisions cannot explicitly mention the use of compensation guidelines (even though judges may use them). In order to assess the implicit use of these guidelines, several studies have therefore sought to establish a correspondence between the amount of compensation for certain heads of damage provided by courts and the amount suggested by the different guidelines. In 2016, a study, based on a corpus of decisions from 2011 to 2013, has highlighted the influence of two different guidelines: the ONIAM guidelines[65] were particularly used in decisions by administrative appeal courts, and the B. Mornet guidelines by judicial appeal courts.[66] Studies using first-instance decisions had more difficulty establishing the use of these guidelines.[67] Nevertheless, one of them was able to establish the correspondence between the compensation for various types of damages and different versions of the Mornet Guidelines.[68]

The role of capitalisation scales is easier to assess because their use is more often explicitly mentioned in court decisions. Thus, studies are successful in tracing them.[69]

The study of data extracted from court decisions certainly allows for the extraction of knowledge from the corpus of decisions studied. However, the potential of new technologies makes it possible to go beyond mere understanding

---

also S. Porchy-Simon et al, n 35 above, making comparisons based on the geographical areas of the three main administrative courts of appeal and the three main judicial courts of appeal.

[61] S. Porchy-Simon et al, n 35 above. Indirectly, C. Quézel-Ambrunaz et al, n 35 above.

[62] For court decisions that do not follow the rule, cf C. Quézel-Ambrunaz, n 35 above.

[63] On the 'compensation smile', see. A. Gayte-Papon de Lameigné et al, 'La modélisation de l'indemnisation du préjudice corporel. Un exemple de "justice quantitative" au service de l'équité', in F. G'sell ed, *Le big data et le droit* (Paris: Dalloz, 2020), 45; L. Belleil and J. Lévy-Véhel, 'Sur la modélisation des décisions de justice', in J.-P. Clavier ed, *L'algorithmisation de la justice* (Bruxelles: Larcier, 2020), 23 ff.

[64] V. Rivollier, n 35 above.

[65] ONIAM is a compensation funds dedicated to medical accidents and health issues. It publishes on Internet its own compensation guidelines.

[66] S. Porchy-Simon et al, n 35 above, does not establish a link concerning the 'permanent functional deficit', link only sought with the Mornet guidelines (141), but the comparison is more conclusive concerning the suffering (152) and the permanent aesthetic loss (177-178).

[67] N. De Jong, n 35 above.

[68] C. Quézel-Ambrunaz, n 35 above.

[69] N. De Jong, n 35 above; C. Quézel-Ambrunaz et al, n 35 above, V. Rivollier, n 35 above.

and attempt to predict the outcome of litigation. This is what the French Ministry of Justice has tried to do in the field of personal injury compensation through the Datajust project.

IV. Predicting the Outcome of a Dispute Through Legal Metrics? The Failure of the Datajust Project

The Datajust project, led by the Ministry of Justice between 2020 and 2022, perfectly illustrates the difficulties encountered in analysing of a large corpus of court decisions and constructing a public tool based on them. This aim of the project was to create a personal injury compensation tool based on an accessible corpus of decisions, which would be used to model judges' previous decisions.

In fact, compensation for personal injury is often based on compensation guidelines that are multiple and based on non-transparent development methods. Several compensation guidelines coexist and lead to quite different results. None of them has any official value. Their authors generally pretend that these guidelines reflect common compensation practices. Yet, it is unclear how these practices have been measured or assessed.[70] Thus, the idea of a single set of compensation guidelines emanating from the Ministry of Justice and based on transparent and reliable analysis does not seem absurd. This was indeed the initial ambition of the Datajust project when it was implemented.

The Datajust project was born through the regulatory act (decree) of 27 March 2020, establishing the automated processing of personal data called 'DataJust'.[71]

---

[70] B. Mornet, 'Le référentiel indicatif régional d'indemnisation du préjudice corporel', in I. Sayn ed, *Le droit mis en barèmes ?* (Paris, Dalloz, Thèmes et commentaires, 2014), 213; D. Martin, 'La politique d'indemnisation de l'ONIAM' 46 *La Gazette du Palais* (19 avril 2008).

[71] 'Décret 27 March 2020 no 2020-356 portant création d'un traitement automatisé de données à caractère personnel dénommé «DataJust»' 77-2 *Journal Officiel de la République Française* (29 mars 2020). About this decree, see. A. Bensaoun and T. Douville, 'DataJust, une contribution à la transformation numérique de la justice' *La Semaine Juridique édition générale* (2020), 907-910; R. Bigot, 'DataJust alias Thémis.I.A.: les premiers pas officiels de l'intelligence artificielle dans les salles des pas perdus' *Lexbase Avocats* (mai 2020); J. Bourdoiseau, 'Datajust ou la réforme dudroit de la responsabilité civile à la découpe?' *La lettre juridique, Lexbase* (avril 2020); M. Fathisalout-Bollon and V. Rivollier, 'À propos de DataJust: justesse de l'outil numérique, juste indemnisation des victimes?' *Revue Lamy de droit civil*, 6819, (2020); Y. Meneceur, 'DataJust, face aux défis de l'intelligence artificielle' *La Semaine Juridique édition générale,* 1978, (2020); S. Merabet, ' "DataJust" et l'effet papillon. À propos du décret du 27 mars 2020' *Revue pratique de la prospective et de l'innovation* (2020), 582. See also on this project, J. Bourdoiseau, 'Le recours à l'intelligence artificielle pour évaluer les préjudices. Rapport de synthèse', in O. Gout ed, *Responsabilité civile et intelligence artificielle* (Bruxelles: Bruylant, coll. du GRERCA, 2022), 635-645; S. Desmoulin, 'Le diable se cache-t-il dans les détails? Réflexions à propos du traitement automatisé de données à caractère personnel "datajust" ', in J.-P. Clavier ed, *L'algorithmisation de la justice* (Bruxelles, Larcier, 2020), 143-159; E. Petitprez and R. Bigot, 'Standard humain ou standardisation algorithmique de l'évaluation du dommage corporel?' *Lexbase Avocats* (janvier 2021); V. Rivollier and M. Viglino, 'Le recours à l'intelligence artificielle pour évaluer les préjudices. Rapport français', in O. Gout ed, *Responsabilité civile et intelligence artificielle*, 675-696; L. Viaut, 'L'évaluation des préjudices corporels par algorithmes'

The purpose of this decree was not to create a tool but to establish the regulatory conditions for its construction and experimentation. Indeed, the project required access to personal data contained in judgments collected for this purpose. The project aimed to analyse all

> 'judgments rendered on appeal between 1 January 2017, and 31 December 2019, by administrative courts of appeal and civil chambers of judicial courts of appeal in disputes concerning the compensation of personal injury.'[72]

Before creating a tool intended to propose a method for evaluating such damages, the ministry services had to undertake a vast analysis of existing practices in this area. The purpose of the decree was therefore mainly to make these analyses possible, without knowing exactly what the envisaged tool would consist of. The decree was subject to appeals for annulment, which were rejected.[73]

The techniques used for data extraction are not mentioned in the decree. Yet, algorithmic techniques for natural language processing have been implemented to extract information from the *corpus*: sequencing of parts of the judgments, identification of significant dates, identification of heads of damages whose compensation is discussed, identification of amounts proposed by the parties and decided by the courts, etc. However, the decree remains silent on these tools and only considers the algorithm as the purpose of the project. The envisaged tool has probably never been clearly defined by the Ministry's services. The data extraction work was an essential and considerable prerequisite.

While the data extraction work seems to have been at least largely completed, the construction of the envisaged tool did not materialise. In accordance with the 2020 decree authorising the use of data for two years, and in the absence of an extension, the project was abandoned in March 2022.[74] No tool for litigants, legal professionals, or lawyers emerges, but it is not even certain that the ministry's work went beyond a simple analysis of the corpus of decisions. None of these elements were made public.

We will present successively the context of the project (1), the difficulties

---

10 *Les Petites Affiches* (31 mai 2021); V. Rivollier, 'L'aventure Datajust: histoire d'un échec', in I. Sayn and V. Rivollier eds, n 1 above, 85.

72 Art 2 décret 27 March 2020, n 70 above.

73 Conseil d'État, 30 December 2021 no 440376, unpublished; in *Dalloz IP/IT*, 6-7 (2022), with a comment by C. Crichton; in *La Semaine Juridique édition générale*, 760 (2022), with remarks by L. Cluzel-Métayer; in *La Gazette du Palais* GPL434m6, 9 (12 avril 2022), with a comment by T. Douville.

74 On the abandonnent of the project, see L. Bloch, 'Datajust-DataJust : ni fleurs, ni couronnes' *Responsabilité civile et assurances*, mark 3 (2022); É. Marzolf, 'Le ministère de la Justice renonce à son algorithme Datajust' (14 January 2022), available at https://acteurspublics.fr (last visited 30 September 2024); S. Merabet, 'Hommage posthume à l'abandon de DataJust: des principes directeurs de la justice numérique' *Revue pratique de la prospective et de l'innovation*, 18-21 (2022); V. Rivollier et al, 'Le retrait de DataJust, ou la fausse défaite des barèmes' *Recueil Dalloz*, 467 (2022).

encountered in extracting and using legal metrics (2), and its subsequent failure (3).

1. The Datajust Project Context

Datajust aligns with the timeframe of the implementation process of open data for court decisions, initiated by the Statute of 7 October 2017 (Act for a Digital Republic) and the 2018-2022 Programming and Justice Reform Act. Formally, Datajust does not rely on decisions from open data, which were not yet available when the project was initiated. However, it uses decisions that subsequently became the first accessible in open data. Like the research projects mentioned earlier, this project relies on the exploitation of the JuriCa and Ariane Archives databases, and thus the decisions of civil chambers of judicial courts and decisions of administrative appeal courts rendered in 2017, 2018, and 2019. The exclusion of criminal judgment chambers may draw the same criticism as the aforementioned projects. The appeals filed against the decree also criticised the fact that decisions subject to appeal in cassation and possibly annulment were not excluded from the corpus. It would have compromised the accuracy principle according to personal data law. However, the Council of State rejected this point. The extent of the decision corpus has not been disclosed, but it is estimated that it comprises between 3000 and 4000 judicial decisions and between 300 and 400 administrative decisions.

Through this project, the ministry has shown its 'voluntarism': it does not just make court decisions available to the public, leaving economic operators to seize them, but seeks to develop internal uses of this data itself.

Furthermore, the Datajust project was part of the civil liability reform process, initiated in the 2000s. Indeed, in 2016 and 2017, the ministry disseminated reform projects envisaging rules and tools specific to the compensation of personal injuries.[75] Among these tools was an indicative compensation guideline for non-pecuniary damages. The guidelines were developed based on

> 'a database bringing together, under the state's control and under conditions defined by decree in the Council of State, final decisions rendered by courts of appeal in compensation for personal injury to victims of a traffic accident'.[76]

Even though the Datajust project differs slightly from the ministerial project in the scope of decisions considered, it can be seen as anticipating the implementation of a civil liability reform. Moreover, such guidelines would not need legislative reform to be adopted by regulatory means.

Moreover, the project is based on the idea that the exploitation of legal metrics derived from court decisions would enable the construction of tools facilitating recourse to out-of-court settlement. This is particularly evident in the decree of

---

[75] Ministère de la Justice, *Avant-projet de réforme de la responsabilité civile* (avril 2016); Ministère de la Justice, *Projet de réforme de la responsabilité civile* (mars 2017).
[76] Art 1271 projet de réforme de la responsabilité civile (2017).

27 March 2020, according to which the envisaged tool must serve

> 'the information of the parties and assistance in evaluating the amount of compensation to which victims may claim in order to promote amicable settlement of disputes'.[77]

Numerous mechanisms encouraging out-of-court compensation settlement already exist in the field of personal injury, especially in cases resulting from traffic accidents or medical accidents. Yet, no evaluation of these mechanisms has been conducted beforehand.

Furthermore, transitioning from an analysis of judicial decisions to guidelines is not straightforward. Similarly, the existence of official guidelines does not necessarily lead to increased recourse to out-of-court settlement. Indeed, identifying the empirical determinants of personal injury compensation depends on judicial decisions with heterogeneous drafting and content. Moreover, the identified determinants may not align with the legal criteria for compensation: one cannot validly base the amount of compensation solely on the gender of the victim, even if differences in compensation are identified. Furthermore, the link between the predictability of the decision, which would be reinforced by guidelines, and out-of-court resolution is not as evident as it seems. Theoretical models developed in conflict economics do not definitively answer the question. Models based on risk aversion consider that the uncertainty about the outcome of the trial encourages parties to reach an agreement. By reducing judicial uncertainty, guidelines would then be a tool reducing incentives to negotiate: the risks of being disappointed by the outcome of the legal action are reduced in the presence of such guidelines.[78]

2. The Difficulties Encountered in the Extraction and Use of Legal Metrics

As previously discussed, extracting data from legal decisions is made difficult by the variable structure of these decisions and the heterogeneous information they contain.[79] In cases of personal injury, these difficulties are multiplied: understanding the isolated court decision when it makes references to parties' submissions or expert reports, the division of compensation into multiple heads of damages, and the omission of key dates essential for understanding the decision, among other factors. Despite the development of a data extraction tool within the framework of Datajust, human analysis was still necessary. Given the long-term nature of the project, sustaining Datajust requires a dedicated team over time.

Developing an 'algorithm', as proposed in the decree, assumes regularities can

---

[77] Art 1 décret 27 March 2020, n 70 above.

[78] C. Bourreau-Dubois et al, 'Les barèmes, outils d'aide à la décision pour les justiciables et les juges' *Revue d'économie politique*,199-222 (2021-22).

[79] See M. Cottin, n 34 above.

be identified in the available corpus. While an algorithm can reveal or objectify the rationalities within a data *corpus*, it's uncertain whether a few thousand decisions, particularly heterogeneous ones, suffice to identify determinants in a sufficiently representative manner. Even with a hypothetical corpus of 4,400 decisions, this would be relatively small given the diversity of injuries and victims that could be included. Some heads of damages are present in only a few decisions, such as 'educational and training loss' which primarily affects young victims, making it difficult to discern regularities. Moreover, even if some damages are present in numerous decisions, the diversity of personal circumstances makes comparison challenging. For instance, how can regularities be sought in the corpus to assess the compensation for a 35-year-old woman, a mother of three, divorced, working as a nurse, and having suffered a head injury in a cycling accident? Comparable situations may be scarce, making it challenging to discern regularities, particularly if decisions vary significantly in similar circumstances. Additionally, a judge's awarded compensation only makes sense when compared to the parties' claims, which are not always reproduced in decisions.

In personal injury compensation law, neither legislation nor case-law from higher courts establishes criteria for determining compensation amounts for many non-pecuniary damages. While case law largely adopts definitions proposed by the Dintilhac report, it doesn't provide calculation methods. Furthermore, the Supreme Court's oversight is minimal, granting trial judges sovereignty not only in determining compensation amounts but also in selecting criteria. For instance, one court might consider a child's cohabitation significant in determining the compensation of his affliction due to a parent's death, while another might prioritize the child's age. Moreover, drafting rules may lead to the omission of hidden motivational elements; although judges widely use compensation guidelines, they are prohibited from mentioning this in their decisions.[80] If an algorithm simply reproduces these guidelines, its necessity may be questioned.

Given the uncertain intrinsic rationality of the studied corpus, the analyst's role appears fundamental. Since analysis criteria are not entirely predefined, analysts largely construct them. Thus, one may question to what extent the corpus's rationality is constructed by the analyst. The fact that the proposed tool pretends to observe a practice doesn't prevent a certain degree of construction by the observer. Methodological precautions, sufficient when describing litigation and existing practices, may not be adequate when the tool aims to predict dispute outcomes.

---

[80] Prohibiting *de facto* reference to scales in court rulings, eg in personal injury cases: Cour de cassation, deuxième chambre civile, 24 October 2019 no 18-20.818, unpublished; Cour de cassation, deuxième chambre civile, 22 November 2012 no 11-25.988, unpublished; in family cases: Cour de cassation, première chambre civile, 23 October 2013 no 12-25.301, published in the Bulletin. On this paradox, see V. Rivollier, 'L'indemnisation du dommage corporel. Les barèmes dans le discours des magistrats', in I. Sayn et al eds, *Les barèmes (et autres outils techniques d'aide à la décision) dans le fonctionnement de la justice* (Paris, Mission de Recherche Droit et Justice, 2019), 69-71.

3. Failure of the Datajust Project

Although the ministry has not officially communicated on the abandonment of the project, it results from the expiry of the two-year deadline set by the initial decree. This has also been confirmed by certain specialized media.[81] Excessive optimism regarding the extraction of data and the possibility of identifying regularities probably collided with the reality of the analysed decisions and the extreme heterogeneity of the contained information. As empirical research projects in the field of personal injury have shown, extracting knowledge from such a corpus is possible, but not (yet) using an automated data extraction method.[82] Furthermore, knowing is not predicting, and the gap between knowledge and prediction is difficult to overcome.[83]

Moreover, some legal professionals demonstrated a hostility as soon as the decree of 27 March 2020 was published. Several professional associations have expressed their concerns.[84] These concerns were also raised in certain parliamentary questions.[85] These criticisms fit into a rhetoric rejecting scales and compensation guidelines in the field of personal injury, with these tools being presented as contrary to the personalisation of compensation.[86]

The discontinuation of this project also reflects a step back of public intervention. As a result, civil liability reform projects appear to have a standstill, so the soft law tools they sought to introduce are forgotten. From the perspective of exploiting open data from judicial decisions and the legal metrics that could be derived from them, no other project as ambitious as Datajust has emerged: the data is made available to private actors, with the state not going any further. Surprisingly, it does not propose any further regulation of the use of this data by legal tech companies: no certification, labelling, or quality control of the services offered is carried out. Yet several experiences abroad and in France of using massive data in the legal field have highlighted possible biases and resulting risks.[87]

---

[81] É. Marzolf, n 74 above.

[82] J. Barnier, 'Extraire automatiquement' n 28 above.

[83] See C. Quézel-Ambrunaz, 'À la recherche d'une définition de la jurimétrie'n 20 above.

[84] See the reaction of the National association of victims lawyers: C. Berneld and F. Bibal, 'DataJust: quand le spectre du barème surgit des brumes numériques' 17 *LaGazette du Palais*, 79 (2020); A. Coviaux, 'Sans soin ni loi: l'inquiétant projet DataJust' 17 *La Gazette du Palais*, 83 (2020); See also Conseil national des barreaux, Motion sur le décret du 27 mars 2020, Assemblée générale du 3 avril 2020; Syndicat de la magistrature, Courrier à la ministre de la Justice, 3 avril 2020.

[85] See, at the Senate, the ministry of Justice answer to the written question no 16942, J.-M. Mizzon, 'Algorithmes et justice prédictive' *Journal officiel du Sénat*, 2899 (25 June 2020); à l'Assemblée nationale, la réponse du ministère de la Justice à la question écrite no 29640, J. Corneloup, 'Mise en œuvre Datajust' *Journal officiel de l'Assemblée nationale* (18 August 2020).

[86] C. Quézel-Ambrunaz, *Le droit du dommage corporel* (Paris: Librairie générale de droit et de jurisprudence, 2nd ed, 2023), 444.

[87] The discriminatory biaises of Compas software in US are well documented. See also the biases revealed in the fraud detection software in Netherlands, Agency for Fundamental Rights, *Bias in Algorithms. Artificial Intelligence and Discrimination* (Vienne, Luxembourg: Publications Office of the European Union, 2022). The same happened in France: G. Geiger et al, 'Profilage et

The incomplete status of the Datajust project does not necessarily indicate a failure. The ministry was aware of the uncertainties associated with the project from the beginning. However, the lack of communication regarding the lessons learned from the project is concerning. Additionally, the absence of government intervention may raise concerns about potential abuses. Indeed, the corpus of mobilised court decisions is now freely accessible, and any private operator can use it to offer similar tools. Certainly, not all operators will have the same scruples as the ministry regarding the methodological difficulties involved in such a project.

V. Prospects

Legal metrics is still a relatively new field, especially when considering that rupture occurred between the early applications of computing to law and the current world, opened up by the interconnection of networks and artificial intelligence.[88]

A relatively effective application of artificial intelligence, implemented by both private and public actors, is the consolidation of legislative texts (in the broad sense). Indeed, modifying texts are drafted in the manner of 'in such article, after such word, such word is added'. Artificial intelligence automates the process of updating texts and navigating through the different versions of the texts: past, in force, and forthcoming.

In France, the government currently has mainly two less ambitious projects than some they have sketched out in the past. The first is to complete open data, ie, the provision to everyone of all court decisions rendered – without any history prior to the publication of the first decisions. The second is the use of artificial intelligence, but in a less ambitious manner than what was planned for DataJust. Two modules are currently being successfully exploited within the competent department of the Court of Cassation. The first, linked with open data, is an algorithm for pseudonymizing court decisions: the volume of decisions from the courts below requires almost human-free processing. The second is the orientation of appeals. As already mentioned, the French Court of Cassation receives a significant number of appeals each year. It is divided into chambers: 3 civil chambers, each with its own areas of competence, a social chamber, a commercial chamber, and a criminal chamber. Each appeal must therefore be directed to the appropriate chamber, which requires analysing not only the facts of the case but especially the legal question at issue, to determine the area of law in which it arises.[89] Artificial intelligence performs this orientation, however, monitored by a team of legal experts.[90]

---

discriminations: enquête sur les dérives de l'algorithme des caisses d'allocations familiales' *Le Monde* (4 décembre 2023), available at https://tinyurl.com/m7x82rv8 (last visited 30 September 2024).

[88] C. Bordere, 'Que reste-t-il de la première jurimétrie?' n 5 above.

[89] E. Serverin et al, 'La nomenclature des affaires' n 32 above.

[90] See H. Abdine, 'JuriBERT: un modèle linguistique pré-entrainé pour le domaine juridique français', in I. Sayn and V. Rivollier eds, n 1 above, 133.

Another initiative from the Ministry of Justice may provide access to new legal metrics. A large reform of information systems in civil matters has been initiated. This reform process named 'Portalis' is progressively developed. Its purpose is to update and merge the many out-dated information systems actually coexisting. This new system will attribute a unique number for any case, from first instance to cassation court (and even referring court), allowing to link first instance, appeal and cassation decisions on the same case. From the system, many statistical new information may be extracted and provide new knowledge.[91] And Portalis should also provide some frame or model of redaction for court decision, so the structure of court decision may be more homogeneous.[92]

As for the initiatives of private publishers, the current trend is less towards extracting data from court decisions than towards exploiting advances in generative artificial intelligence and the use of natural language models. Training datasets systematically include legal texts, but also, depending on the publishers or startups entering this market, blog articles, or all court decisions available in open data. Regarding these decisions, they are exploited as text, by language processing tools, but not as a source of quantifiable data, which limits the relevance of responses to questions calling for a numerical answer.

The fields of legal metrics are vast. If this article had been written a year ago, its content would have been entirely different. It is possible that by the time it is read, it will already be partly outdated. There is a problem to be solved: how to convert performative texts into structured data?[93] This project can mobilise computer scientists, data scientists, and legal experts.

The future of jurimetrics, in a broad sense, is promising due to the strong demand for knowledge and predictability of judicial decisions.[94]

---

[91] See P. Ghaleh-Marzban, 'PORTALIS: le projet de modernisation de la justice' *Dalloz IP/IT*, 152 (2018).

[92] M. Cottin, n 34 above.

[93] ibid

[94] Which is a long term seeking: N. Bernoulli: 'De usu Artis Conjectandi in jure' (Basilea: Conradus, 1709).