

8-31-2022

If I Had More Time, Would I Have Written a Shorter and Faster Decision? An Empirical Examination of the Evolution of Trial Court Decisions

Jon Khan
Osgoode Hall Law School

Follow this and additional works at: <https://digitalcommons.schulichlaw.dal.ca/dlj>



Part of the [Courts Commons](#), [Judges Commons](#), and the [Law and Society Commons](#)



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

Recommended Citation

Jon Khan, "If I Had More Time, Would I Have Written a Shorter and Faster Decision? An Empirical Examination of the Evolution of Trial Court Decisions" (2022) 45:2 Dal LJ.

This Article is brought to you for free and open access by the Journals at Schulich Law Scholars. It has been accepted for inclusion in Dalhousie Law Journal by an authorized editor of Schulich Law Scholars. For more information, please contact hannah.steeves@dal.ca.

Jon Khan*

If I Had More Time, Would I Have Written a Shorter and Faster Decision? An Empirical Examination of the Evolution of Trial Court Decisions

This article draws from my 2019 LLM thesis on Canadian judicial decisions, where I sought to understand two things: how current approaches to judicial decision-writing may impact access to justice and how might we make decisions a better source of data while also making them more timely, concise, accessible, and consistent. It presents the results and analysis of an original empirical study of the evolution of British Columbia trial decisions over 40 years (1980–2018). It argues that the current process for writing Canadian judicial decisions likely does not further the goals of access to justice and may even hinder them. Further study and targeted reforms are urgently needed to address delay, timeliness, accessibility, and consistency in Canadian judicial decisions. But reforms must not be based on anecdote, intuition, one-off examples, or single empirical studies. Instead, proposed reforms should be based on more deliberate design strategies such as those that human-centred design employs. For example, courts could and should generate extensive, transparent data on judicial decision-writing, judicial decisions, and the judicial process; rely on interdisciplinary methods to better understand current problems; ideate new ways of writing decisions that respond to that research; prototype and iterate those new ideas; and finally, extensively consult users about writing and reading decisions.

Cet article s'inspire de ma thèse de maîtrise en droit de 2019 portant sur les décisions judiciaires canadiennes, où j'ai cherché à comprendre deux choses : comment les approches actuelles en matière de rédaction des décisions judiciaires peuvent avoir un impact sur l'accès à la justice et comment nous pourrions faire de ces décisions une meilleure source de données tout en les rendant plus opportunes, concises, accessibles et cohérentes. Il présente les résultats et l'analyse d'une étude empirique originale de l'évolution des décisions de justice rendues en Colombie-Britannique sur 40 ans (1980–2018). Il soutient que le processus actuel de rédaction de décisions judiciaires canadiennes ne favorise sans doute pas les objectifs d'accès à la justice et peut même les entraver. Des études plus approfondies et des réformes ciblées s'imposent de toute urgence pour régler les problèmes de retard, de rapidité, d'accessibilité et de cohérence des décisions judiciaires canadiennes. Mais les réformes ne doivent pas être fondées sur des anecdotes, des intuitions, des exemples ponctuels ou des études empiriques uniques. Les réformes proposées devraient plutôt être fondées sur des stratégies de conception plus délibérées, comme celles qu'emploie la conception centrée sur l'humain. Par exemple, les tribunaux pourraient et devraient produire des données exhaustives et transparentes sur la rédaction des décisions judiciaires, les décisions judiciaires elles-mêmes et le processus judiciaire; s'appuyer sur des méthodes interdisciplinaires pour mieux comprendre les problèmes actuels; imaginer de nouvelles façons de rédiger les décisions qui répondent à cette recherche; prototyper et énoncer ces nouvelles idées; et enfin, consulter largement les utilisateurs sur la rédaction et la lecture des décisions.

* JD (University of Ottawa), LLM (University of Toronto), PhD Candidate (Osgoode Hall Law School). Correspondence: jonkhan@yorku.ca // Research website: <www.deliberatelegaldesign.com>. Many people shaped my thinking during my thesis and supported this research. Without their guidance, help, and support, I would not have produced this work. Thanks to Professor Andrew Green for your advice, direction, guidance during my LLM; Professors Gillian Hadfield and Dan Ryan for motivating me to think differently about law; Professor Sean Rehaag for your guidance and help long before you were obligated to; Jonathan Kroening for your perspicacity and endless conversations about data science, statistical science, machine learning, and our desire to live in a more fair world; Tom Posyniak, Lars Brusven and Steven Davis for always being willing to talk about the law and the implications of my all my ideas; Nicole Gilewicz for supporting my curiosity, work, and pursuits; and finally, the anonymous reviewers for recommendations that greatly improved this article. All errors and omissions are mine.

Introduction

- I. *Explaining the empirical study's design and the machine learning methodology*
- II. *Exploring descriptive trends and predictions about decision length and delivery time*
 1. *Descriptive trends: on average, BCSC trial decisions' current length rivals that of US Supreme Court decisions, and parties wait much longer than they used to*
 - a. *BCSC court time per matter: on average, no real increase; but now, judges encounter more long and "mega trials"*
 - b. *BCSC decision length and delivery time: on average, judges in 2018 take three times longer to issue decisions three times longer than their 1980 colleagues*
 - c. *BCSC decisions taking longer than 180 days: 1.1 per cent of decisions used to take 180 days or more; now, eight per cent of decisions do*
 - d. *BCSC self-representation: three per cent of litigants used to appear without lawyers; now, at least 16 per cent of reported decisions involve some type of self-representation*
 - e. *BCSC decisions' structure: historically, almost no decisions used headings; now 80 per cent of decisions do*
 - f. *Final takeaways: we have much more to learn about oral decisions, judges' writing practices, and finding ways to solve simple errors in trial decisions*
 2. *Predictive analysis: if your litigation requires more than a few court days, you will likely have to wait a long time for a lengthy decision*
 - a. *Length: the longer your trial, the longer your decision*
 - b. *Delivery time: the longer your trial and the longer your decision, the longer your wait*
- III. *Analyzing the granular evolution of BCSC decisions' length and delivery time and judges' workload over the last 40 years, and suggesting reforms to test*
 1. *Length: decisions are no longer relatively short*
 2. *Delivery time: decisions are no longer delivered relatively fast*
 3. *Judges' workload: a clear and subjectivity exists in judges' outputs—some judges take longer and are more expressive than their colleagues (either in their words or what they decide to publish or both)*
- IV. *Limitations (this research has limited application) and future research suggestions (we need more analysis of trial-level judicial decisions and decision-making)*

Conclusion

What role does the current uniform allocation of judgment writing weeks in this Court play in such delays? Must existing scheduling principles be re-thought? Is it time to re-jig the allocation of judgment writing so that a more direct relationship exists between the complexity of a matter heard and the time allocated to a judge to write the resulting decision in a timely manner? If we are to restore the health of Ontario's ailing civil litigation system, as judges we must not only call on those who appear before us to change their litigation culture, we also must look at our own internal scheduling culture and change it to meet the realities of our times. No sacred cows, judicial or otherwise, should stand immune from scrutiny behind some firewall. Everything must be on the table for examination and reform. Restoring the health of our province's civil litigation system requires no less.¹

Introduction

Canadian governments, courts, and scholars have a paucity of data about the Canadian legal system. Simply put, Canada has massive data deficits about its legal system² and the common legal issues that Canadians face every day.³ You currently cannot even find out the national divorce or guilty plea rates.⁴

1. *Western Larch Limited v Di Poce Management Limited*, 2012 ONSC 7014 at paras 269-277 [paragraph numbers removed] [*Western Larch*].

2. CBA Access to Justice Committee, *Equal Justice: Balancing the Scales* (Canadian Bar Association, 2013), online (pdf): <www.cba.org/CBAMediaLibrary/cba_na/images/Equal%20Justice%20-%20Microsite/PDFs/EqualJusticeFinalReport-eng.pdf> [perma.cc/9KDL-9HUA] (“there are still many gaps in our knowledge and these gaps impact our capacity for reform...the empirical basis for decision making is still extremely limited compared to what is known about health and education. The justice system has a long way to go in terms of what information is collected, how it is collected and how open it is” at 51); Access to Justice Coordinating Committee, *Final Report: 2018* (Halifax: A2JCC, 2018), online (pdf): <courts.ns.ca/News_of_Courts/documents/A2JCCreport_WEB.pdf> [perma.cc/732A-Z77P] “[m]uch of the feedback on the functionality of the legal system is based on anecdotal evidence. The lack of reliable statistics to support decision-making is an impediment” at 13); Senate of Canada, *Delaying Justice is Denying Justice: Final Report of the Standing Senate Committee on Legal and Constitutional Affairs* (June 2017) (Chair: Bob Runciman) at 29-31, online (pdf): <sencanada.ca/content/sen/committee/421/LCJC/reports/Court_Delays_Final_Report_e.pdf> [perma.cc/2BLW-WXS8]; Canadian Centre for Justice Statistics, *Civil Courts Study Reports*, by C Brookbank, B Kingsley & T Leonard, Catalogue No 85-549-XIE (Ottawa: Statistics Canada, 1999) at 16-17, online (pdf): <publications.gc.ca/Collection-R/Statcan/85-549-XIE/0009985-549-XIE.pdf> [perma.cc/QKE4-FVT5].

3. Tavia Grant & Eric Andrew-Gee, “In the information age, Canada is falling behind,” *The Globe and Mail* (25 January 2019), online: <www.theglobeandmail.com/canada/article-in-the-information-age-canada-is-falling-behind/> [perma.cc/M2WH-VC4Y] (“[o]n everything from public health to housing, the economy and education, the country does not have the data it needs to make smart decisions. We are tracking the gaps and their effect on your everyday life”); Eric Andrew-Gee & Tavia Grant, “In the dark: The cost of Canada's data deficit,” *The Globe and Mail* (26 January 2019), online: <www.theglobeandmail.com/canada/article-in-the-dark-the-cost-of-canadas-data-deficit/> [perma.cc/Z9ZN-R334]; Tavia Grant & Eric Andrew-Gee, “Experts urge Ottawa to fix Canada's data deficit,” (27 January 2019) *The Globe and Mail*, online: <www.theglobeandmail.com/canada/article-experts-make-recommendations-about-how-canada-can-fix-its-data-deficit/> [perma.cc/QV9E-8GR5] (“[b]y data gap, we mean areas at the national level in which data are not collected or readily accessible. These could be areas where there is no ability to compare across provinces or cities, where the existing information is years out of date, published infrequently or not comparable with prior years”).

4. *Ibid*; Department of Justice Canada, *Guilty Pleas Among Indigenous People in Canada*, by Angela Bressan & Kyle Coady (Ottawa: Department of Justice Canada, 2017), online (pdf): <www.

In 2012, a Canadian judge did something novel that is closely related to Canada's legal data deficit and a largely under-researched topic: Justice David Brown tracked how long it took him to write one judicial decision. His tracking revealed a surprising, disproportionate result that showcases the need for comprehensive data, analysis, and potential reform of the Canadian judicial decision-writing process.⁵

To address a 1.5-day dispute that occupied .86 per cent of Justice Brown's allocated yearly sitting time, he had to use 21 per cent of his allocated yearly writing time.⁶ Unsurprisingly, after finishing this exercise and issuing his decision, Justice Brown asked if more allocated writing time would be one way "to restore the health of Ontario's ailing civil litigation system."⁷ But would more writing time really help restore the health of Ontario or Canada's litigation system? It might.⁸ But no one really knows, and little has happened since Justice Brown's call for help.

Despite judicial delay being an overwhelming problem in Canada's legal system,⁹ Canada's federal and provincial governments do not seem to know how long parties must wait for most court decisions.¹⁰ This ignorance contrasts leading comparator jurisdictions—e.g. New Zealand¹¹

justice.gc.ca/eng/rp-pr/jr/gp-pc/gp-pc.pdf> [perma.cc/6Y56-ZEQL].

5. During my LLM, I attempted to replicate Justice Brown's time-tracking by asking other judges to also track their time anonymously. Unfortunately, I was not successful in my first attempt: Jon Khan, "*The Life of a Reserve*": *How Might We Improve the Structure, Content, Accessibility, Length, and Timeliness of Judicial Decisions* (LLM Thesis, University of Toronto, 2019) Part IV at 61-62, DOI: <1807/98120>. But I am currently attempting to conduct the same time-tracking survey in my PhD research as well as further research on the judicial decision-writing and -making process.

6. *Western Larch*, *supra* note 1 at paras 269-277 (Brown J did this analysis).

7. *Ibid* at paras 276-277.

8. While Justice Brown's anecdote and analysis are powerful, they prove nothing other than the need to test and further analyze this issue.

9. See Senate of Canada, *supra* note 2.

10. Canada's federal government has embarked on the pursuit of gathering better data about its 96 courts. But to date, the data is not readily amenable to empirical analysis—e.g. the current data does not offer information about the evolution of decisions, including their length or delivery time. See "Civil Court Survey (CCS)" (last modified 25 March 2019), online: *Statistics Canada* <www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=1230519> [perma.cc/PTT3-ED5X].

11. Canada's federal government has almost no aggregated data on this subject. In contrast to New Zealand, e.g. who tracks delivery time for all its trial courts (see "Judgment Delivery Statistics 2012–2019" (last visited 14 July 2022), online: *Courts of New Zealand* <www.courtsofnz.govt.nz/the-courts/high-court/judgment-delivery-statistics-2012-2019/> [perma.cc/846X-4FDS]), Canada only tracks Supreme Court of Canada delivery times. See "Average Time Lapses" (last modified 28 February 2018), online: *Supreme Court of Canada* <www.scc-csc.ca/case-dossier/stat/cat5-eng.aspx> [perma.cc/CKL9-JHRJ]. Some courts likely track these times, but their data is frequently not public. The British Columbia Court of Appeal releases its average delivery time. See *Annual Report: Court of Appeal for British Columbia: 2018* (Vancouver: BCCA, 2018) at 21, online (pdf): <www.courts.gov.bc.ca/Court_of_Appeal/about_the_court_of_appeal/annual_report/2018_CA_Annual_Report.pdf> [perma.cc/69QY-4BES].)

and the United States¹²—where data and scholarship on wait times and their effect is significantly richer.

Canadian courts do not fare better than governments. They also mostly do not seem to know (or if they do know, they do not publicly share their knowledge, data, or analysis). Few Canadian courts issue annual reports about court and judicial behaviour, but most reports do not track the average time parties wait for decisions.¹³ And most Canadian courts—likely because of budgetary constraints—do not even issue public annual reports.¹⁴ The Supreme Court of Canada and the British Columbia Court

12. See Jonathan Petkun, “Can (and Should) Judges be Shamed? Evidence from the ‘Six-Month List’” (2018) [draft paper], online: <ssrn.com/abstract=3205398> [perma.cc/RMJ5-AA29]; Miguel de Figueiredo, Alexandra D Lahav & Peter Siegelman, “Against Judicial Accountability: Evidence from the Six Month List” (2018) at 1, 5 [draft paper], DOI: <10.2139/ssrn.2989777>.

13. Two courts release some information about decisions’ delivery time: the BC Court of Appeal notes that 89 per cent of its decisions are released under six months, and the Supreme Court of Canada notes a mean delivery time of 4.6 months. See BCCA, *supra* note 11; *Annual Report 2018: Supreme Court of British Columbia* (Victoria: BCSC, 2019), online (pdf): <www.courts.gov.bc.ca/supreme_court/about_the_supreme_court/annual_reports/2018_SC_Annual_Report.pdf> [perma.cc/3G7L-F3AS]; *The Court of Queen’s Bench of Alberta: Annual Report: 2016 to 2017* (Edmonton: ABQB, 2017), online (pdf): <albertacourts.ca/docs/default-source/qb/2016-2017-annual-report-with-appendix-jan-19-2018.pdf?sfvrsn=593aac80_0> [perma.cc/UG83-DHMG]; *The Superior Court of Justice: Realizing Our Vision: Report for 2015 and 2016* (Toronto: Ontario Superior Court of Justice, 2016), online (pdf): <www.ontariocourts.ca/scj/files/annualreport/2015-2016.pdf> [perma.cc/6HGG-PF3L]; *Court of Appeal for Ontario: Annual Report 2013* (Toronto: ONCA, 2013), online (pdf): <www.ontariocourts.ca/coa/en/ps/annualreport/2013.pdf> [perma.cc/NTG5-V6BC]; *Superior Court of Quebec: 2010-2014 Activity Report* (Montreal: QCCS, 2015), online (pdf): <numerique.banq.qc.ca/patrimoine/details/52327/2482342> [perma.cc/TSPJ-UGML]; *The Nunavut Court of Justice: A Statistical and Comparative Review of Court Operations in Nunavut: 2017* (Iqaluit: NUCJ, 2018), online (pdf): <www.nunavutcourts.ca/index.php/annualreports?download=442:2017-annual-reports-en> [perma.cc/MFH2-L9ZQ]; *Courts Administration Service: Supporting the Delivery of Justice for All Canadians, Annual Report, 2017-2018*, Catalogue No JU16E-PDF (Ottawa: CAS, 2018), online (pdf): <www.cas-satj.gc.ca/en/publications/ar/2017-18/pdf/Annual_Report_Final_EN.pdf> [perma.cc/RG38-8PH3]; *2018 Year in Review: Supreme Court of Canada* (Ottawa: SCC, 2019), online (pdf): <www.scc-csc.ca/review-revue/2018/yr-ra2018-eng.pdf> [perma.cc/9NPG-DVML].

14. The following courts do not appear to publish annual reports: the Court of Appeal for Alberta; the Saskatchewan Court of Queen’s Bench & the Saskatchewan Court of Appeal; the Manitoba Court of Queen’s Bench & Manitoba Court of Appeal of Manitoba; the Court of Appeal of Quebec (although it does publish some statistics—see “Statistics and Publications” (last visited 14 July 2022), online: *Court of Appeal of Quebec* <courdappelluquebec.ca/en/about-the-court/statistics-and-publications/#c363> [perma.cc/7VV8-V8W2] >); the Supreme Court of Newfoundland and Labrador & the Court of Appeal of Newfoundland and Labrador; the Supreme Court of Prince Edward Island & the Prince Edward Island Court of Appeal; the Nova Scotia Supreme Court & the Nova Scotia Court of Appeal; the Nunavut Court of Appeal; the Northwest Territories & the Court of Appeal for the Northwest Territories; and the Supreme Court of Yukon & the Court of Appeal of Yukon. Notably, however, some provincial courts publish annual reports in provinces where superior courts do not. So provincial courts seem to have the capability to do so. See e.g. the Provincial Court of Newfoundland and Labrador, “Publications: Annual Reports” (last visited 14 July 2022), online: *Provincial Court of Newfoundland and Labrador* <court.nl.ca/provincial/publications/index.html> [perma.cc/KS4N-HR3M].

of Appeal are the only Canadian courts to transparently offer data on wait times.¹⁵

Scholars' efforts have not filled these gaps. Most empirical Canadian research focuses on Canada's apex and appeal courts,¹⁶ but that research and data is often dated and sparse.¹⁷ And while that research is essential, it is not especially helpful for generating empirical evidence about delay in Canada's busiest courts—trial courts—and the thousands of people in civil,¹⁸ family,¹⁹ and criminal²⁰ disputes. Put plainly, “public reporting on the work load of [Canada's courts]...is poor and relatively inaccessible”²¹

Because of this data deficit, we cannot easily analyze how best to help Canadian judges or litigants have a better experience in Canadian courts

15. *Supra* note 11.

16. See Peter McCormick, “Structures of Judgment: How the Modern Supreme Court of Canada Organizes its Reasons” (2009) 32:1 Dal LJ 35 at 37-41; Ian Greene et al, *Final Appeal: Decision-Making in Canadian Courts of Appeal* (Lorimer: Toronto, 1998) at 131-157.

17. Evan Rosevear & Andrew McDougall, “Cut to the Case Counsellor: Patterns of Judicial Writing at the Supreme Court of Canada 1970-2015” (Paper delivered at the Canadian Political Science Association Annual Conference, University of Regina, 31 May 2018) at 1, 2 [unpublished], online (pdf): <www.cpsa-acsp.ca/documents/conference/2018/856.McDougall.Rosevear.pdf> [perma.cc/HF46-YSVR]; For a recent notable example, see Xavier Beauchamp-Tremblay & Antoine Dusséaux, “Not Your Grandparents’ Civil Law: Decisions Are Getting Longer. Why and What Does It Mean in France and Québec?” (20 June 2019), online (blog): *Slaw* <www.slaw.ca/2019/06/20/not-your-grandparents-civil-law-decisions-are-getting-longer-why-and-what-does-it-mean-in-france-and-quebec/> [perma.cc/RD2E-LLHE] (“[f]rom 2003 to 2017, the length of lower court decisions (CQ and CS) has increased by about 40 per cent and the length of the CA decisions has increased by 20 per cent”). For recent notable exceptions, see Kevin LaRoche, M Laurentius Marais & David Salter, “The Length of Civil Trials and Time to Judgment in Canada: A Case for Time-Limited Trials” (2021) 99:2 Can Bar Rev 286, online: <cbr.cba.org/index.php/cbr/article/view/4696/4502> [perma.cc/XX9G-UETL]; Mathew P Good, “Justice Delayed? A Quantitative Assessment of Trends in Length of Hearing, Length of Written Judgment and Time to Judgment: British Columbia, 1970 to 2015” (2021) 70:6 Advocate 831, online: <the-advocate.ca/emag/issues/2021/Nov/page_32.html> [perma.cc/M3H3-TNB2].

18. As of 2017/2018, 337,495 civil cases were active in Canada's superior courts. See “Active civil court cases, by elapsed time from case initiation to first disposition, Canada and selected provinces and territories,” (10 March 2022), online: *Statistics Canada* <www150.statcan.gc.ca/t1/tb1/en/tv.action?pid=3510011601&pickMembers%5B0%5D=1.1&pickMembers%5B1%5D=2.2&pickMembers%5B2%5D=3.2> [perma.cc/QV2R-7YU7].

19. As of 2017/2018, 142,000 family cases were active in Canada's superior courts. See “Active civil court cases, by elapsed time from case initiation to first disposition, Canada and selected provinces and territories” (10 March 2022), online: *Statistics Canada* <www150.statcan.gc.ca/t1/tb1/en/tv.action?pid=3510011601&pickMembers%5B0%5D=1.1&pickMembers%5B1%5D=2.2&pickMembers%5B2%5D=3.3> [perma.cc/S27L-EVZZ].

20. No such data appears to be available. See “Integrated Criminal Court Survey” (24 January 2019), online: *Statistics Canada* <www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=1209304> [perma.cc/35CN-A32S].

21. Lori Hausegger, Matthew Hennigar & Troy Riddell, *Canadian Courts: Law, Politics, and Process* (Oxford: Oxford University Press, 2009) at 40; see also at 333. This accessibility is likely due to funding and resources. In contrast, the Federal Court releases data four times a year. See “Statistics” (last modified 20 May 2022), online: Federal Court <www.fct-cf.gc.ca/en/pages/about-the-court/reports-and-statistics/statistics> [perma.cc/P8D4-QHPT].>

when writing or receiving judicial decisions. For example, we have no idea if Justice Brown's writing time (75 hours) or the parties' seven-month wait time for the decision are typical;²² if the 1.5-day hearing required a lengthy (276 paragraphs/95 pages) and complex decision (24 cases and three statutes were cited);²³ if judges are generally taking longer to issue decisions;²⁴ or if the decision satisfied the parties. The implication of such "lousy data is lousy knowledge."²⁵ And, with such lousy knowledge, any guidance to Canadian judges about what is an acceptable or desirable length, delivery time, or complexity for decisions rests almost exclusively on anecdote. It also stymies the ability to generate solutions to improve decisions' length, delivery time, and complexity.

To fill a very small part of Canada's data deficit, this article presents a novel empirical study of the evolution of judicial decisions in one court—the British Columbia Supreme Court (BCSC). Its focus is on the evolution of decisions' length and delivery time from 1980-2018. This analysis demonstrates concerning trends, including the continual average increase in decision length and delivery time over the last 40 years. Either in isolation or tandem, these trends suggest the current process for Canadian decision-writing may lead to slower decision-making and longer, slower, more complex judicial decisions. They also suggest that the current process for writing Canadian judicial decisions likely does not further the goals of access to justice and may even hinder them.

Extensive, transparent data and further study of Canadian judicial decision-making and -writing is urgently needed—e.g. pan-Canadian data on decision length and delivery time. And targeted reforms are likely needed to address delay, timeliness, accessibility, and consistency in

22. No Canadian database tracks this concept. Loom Analytics, a private company, is tracking how long judges take to issue reasons. See generally *Loom Analytics* <www.loomanalytics.com/> [perma.cc/TLK5-7BWX]. But Loom doesn't track how judges use their time e.g. reading evidence/case law versus drafting decisions. See also Cristin Schmitz, "Ontario Superior Court to review management of judicial tardiness," *The Lawyer's Daily* (29 May 2017), online: <www.thelawyersdaily.ca/articles/3830/ontario-superior-court-to-review-management-of-judicial-tardiness> [perma.cc/Z6RV-MVLB] ("[w]hether the judiciary ... has dropped the ball in addressing an apparently persistent problem [(delay in issuing decisions)] is 'hard to know without more context'"). For one of the only historical discussions on delay in issuing Canadian decisions, see Robert J Sharpe & Kent Roach, *Brian Dickson: A Judge's Journey* (Toronto: University of Toronto Press, 2004) at 370-375 (discussing the SCC backlog in issuing decisions when Chief Justice Dickson was appointed and how he resolved it).

23. See Khan, *supra* note 5. Lengthy opinions are likely infrequently useful. See Gerald Lebovits, Alifya V Curtin & Lisa Solomon, "Ethical Judicial Opinion Writing" (2008) 21:2 *Geo J Leg Ethics* 237 at 252-258.

24. See *supra* note 11.

25. Gillian K Hadfield, *Rules for Flat World: Why Humans Invented Law and How to Reinvent It for A Complex Global Economy* (Oxford: Oxford University Press, 2017) at 218.

Canadian judicial decisions. But reforms must not be based on anecdote, intuition, one-off examples (like Justice Brown’s experience), or single empirical studies (like this one). Instead, proposed reforms—e.g. re-allocating writing time—should be based on more deliberate design strategies such as those that human-centred design employs.²⁶ For example, courts could and should generate extensive, transparent data on judicial decision-writing, judicial decisions, and the judicial process; rely on interdisciplinary methods to better understand current problems; ideate new ways of writing decisions that respond to that research; prototype and iterate those new ideas; and finally, extensively consult users about writing and reading decisions.

Such study might reveal the answer to Justice Brown’s question: more writing time is one way to improve the health of Canada’s litigation system. Or it might highlight other important solutions that might have a bigger impact—e.g. providing more guidance to judges on decisions’ content and structure (perhaps through standardization); reducing the need for “mega” and longer trials; reducing the need for long, complex judicial decisions; optimizing scheduling for trial time and judicial decision-making; increasing the number of judges, clerks, legal officers, support

26. During my LLM research, I mocked up an initial reform plan for decisions that relied on principles of human-centered design. See Khan, *supra* note 5 at 64-83. On the background of human-centred design generally, see “What is Human-Centered Design” (last visited 14 July 2022), online: www.designkit.org/human-centered-design [perma.cc/2G3Y-63PH]; Dave Thomsen, “Why Human-Centered Design Matters,” *Wired* (last visited 14 July 2022), online: www.wired.com/insights/2013/12/human-centered-design-matters/ [perma.cc/DBN2-RVV9]; Joseph Giacomini, “What is Human Centred Design?” (2014) 17:4 *Design J* 606 DOI: <10.2752/175630614X14056185480186>. For human-centred design in the legal system specifically, see Jonathon Rose, “A User Centred Judicial System” (15 January 2017) *Medium*, online (blog): medium.com/@jonathenrose/a-user-centred-judicial-system-71fd6353d450 [perma.cc/UV8Q-JA6W]; Margaret Hagan, “Law by Design” (last visited 14 July 2022) online: [Law by Design](http://www.lawbydesign.co) <www.lawbydesign.co> [perma.cc/QEF4-CX47]; Victor D Quintanilla, “Human-Centered Civil Justice Design” (2017) 121:3 *Penn St L Rev* 745, DOI: <10.2139/ssrn.2655818>; “Design Process” online: *Open Law Lab* <www.openlawlab.com/approach-process/design-process/> [perma.cc/3JP4-J6LZ]; Susan Ursel, “Building Better Law: How Design Thinking Can Help Us Be Better Lawyers, Meet New Challenges, and Create the Future of Law” (2017) 34:1 *Windsor YB Access Just* 28, DOI: <10.22329/wyaj.v34i1.4999>; W David Ball, “Redesigning Sentencing” (2014) 46:4 *McGeorge L Rev* 817, online: scholarlycommons.pacific.edu/mlr/vol46/iss4/6/ [perma.cc/5BTY-XJ3U]; Margaret Hagan, “A Human-Centered Design Approach to Access to Justice: Generating New Prototypes and Hypotheses for Intervention to Make Courts User-Friendly” (2018) 6:2 *Ind J Law & Soc Equality* 199, online: www.repository.law.indiana.edu/ijlse/vol6/iss2/2/ [perma.cc/56UE-XHVM]; Victor D Quintanilla & Michael A Yontz, “Human-Centered Civil Justice Design: Procedural Justice and Process Value Pluralism” (2018) 54:1 *Tulsa L Rev* 113, online: digitalcommons.law.utulsa.edu/tlr/vol54/iss1/7/ [perma.cc/B8WA-LY9S]; Alyson Miller, “What Human-Centered Design Can Tell Us About the State of Dispute Systems Design” (MA Thesis, University of Victoria, 2019) [unpublished], online (pdf): dspace.library.uvic.ca/bitstream/handle/1828/10882/Miller_Alyson_MADR_2019.pdf [perma.cc/8UW7-ZKKN].

staff; increasing data collection about the judicial process (and ongoing analysis of that data); testing out new technological solutions; etc.

This article explains that analysis and thesis in four parts:

- Part I explains the empirical study's design and the machine learning methodology this article avails;
- Part II explores descriptive trends and predictions about what situations may lead to longer and slower judicial decisions;
- Part III analyzes granular results about how BCSC judicial workloads and decisions' word count and delivery time have evolved over the last 40 years and offers some reforms to explore and test that could improve decisions' length, delivery time, and complexity; and
- Part IV addresses the limitations of this article, dataset, and method and calls for further expansive research of judicial decisions and decision-making.

I. *Explaining the empirical study's design and the machine learning methodology*

The dataset was designed to facilitate both descriptive (e.g. how have decisions' delivery time evolved over the last 40 years) and predictive analysis (e.g. what factors best explain that evolution from a statistically significant perspective). The descriptive analysis goal is straightforward: describe the evolution of decisions' length and delivery time over the last 40 years as well as any other notable trends. The predictive analysis goal is also straightforward: identify which dataset features are most likely to predict longer decisions (i.e. word count) and slower delivery time (i.e. the number of days it takes to issue decisions).

To facilitate that analysis, I reviewed all reported BCSC decisions in Quicklaw from 1980, 2000, and 2018²⁷ and hand-coded 16 features:²⁸ (1) case name; (2) Quicklaw citation; (3) registry; (4) whether the decision or Quicklaw provided hearing information; (5) the date the hearing concluded; (6) how many court days were required for the decision; (7) the date the decision was released; (8) how many days lapsed before the decision was released; (9) judge; (10) moving party-type; (11) responding party-type; (12) subject; (13) word count; (14) use of headings; (15) self-

27. I relied on all reported cases from Quicklaw as Quicklaw offered more reported cases than Westlaw or CanLII. For an explanation of how I gathered the cases and why I chose Quicklaw, see Khan, *supra* note 5, Appendix A, Figure 18. And see Appendix E, Figures 10 & 11 for a graphical representation of the overall sample and breakdown by subject matter.

28. Bulk access to the decisions in my dataset was impossible when I began this study. So, I could not use any automated or NLP techniques.

reported information; and (16) appeal information.²⁹ Before removing cases with invalid data points, I gathered 4,993 decisions (i.e. observations) using filtered search terms.³⁰ After removing cases with invalid data points,³¹ 4,988 decisions remained.³² 1980 provided the smallest sample with 1,186 decisions; 2000 provided the largest sample with 1,925 decisions, and 2018 provided the second largest sample with 1,877 decisions.³³ To ensure that Quicklaw’s reported decisions in 1980, 2000, and 2018 were representative of the surrounding years in Quicklaw’s database, I searched the surrounding years to observe how many cases Quicklaw reported. The observations in 1980, 2000, and 2018 were largely consistent with their surrounding years when using the same search parameters.³⁴

Based on the dataset’s descriptive trends, I had three hypotheses about longer and slower decisions:

1. Longer trials (i.e. more court days), particular judges, and particular subjects lead to longer decisions. In other words, “court days,” “judge,” and “subject” are “significant predictors”³⁵ of longer decisions;

29. See Khan, *supra* note 5, Appendix B for the coding manual.

30. *Ibid* (see Appendix A for search parameters).

31. See e.g. *Wuerch v Hamilton* (1980), 6 ACWS (2d) 362, [1980] BCJ No 1332 (QL) (BCSC) (no hearing data given; only the month was given); *Rod McCallum Mercury Sales Ltd v Zurich Insurance Co*, [1980] BCJ No 76 (QL) (BCSC) (oral judgment given in 1978 and decision filed in 1980); *Royal Bank of Canada v Telecommunications Workers Union of British Columbia*, [1980] BCJ No 2179 (QL) (BCSC) (duplicate); *First Bauxite Corp (Re)*, 2019 BCSC 89, [2019] BCJ No 94 (unclear as to when the decision was actually given); *Young v Young*, 2018 BCSC 1396, [2018] BCJ No 3012 (QL) (BCSC) (judge delivered reasons in 2015 but did not sign the transcript, so they were not delivered until 2018).

32. Some decisions did not include hearing information. I could assume that when hearing information was not provided, the decision was an oral decision. This assumption may be true. The dataset included 527 observations where no hearing information was given. The dataset included 392 observations where no hearing information was given but where the decisions were listed as oral decisions. Nonetheless, the dataset also included 135 observations where no hearing information was given, and the decisions were not listed as oral. So, for greater accuracy, I tested removing the decisions where no hearing information was given when I measured delivery time and word count. In some years, the increases were less substantial but still significant. In the following analysis, I have tested with and without these “no hearing information” observations and noted any change in the footnotes.

33. Very few times did multiple decisions arise from the same case with the same hearing dates (once in 1980, 12 times in 2000, and 10 times in 2018). The coding does not, however, classify those rare situations as duplicates. Rather, it treats each decision separately as Quicklaw records since the decisions dealt with different issues, e.g. motions. Accordingly, each case is a separate entry in the court days count. In most situations, however, where multiple decisions arose from the same case, they had their own separate dates. If a decision was a clear duplicate, I did not include it in the dataset.

34. For example, in 1979 and 1981, Quicklaw reported 1,138 and 1,103 cases, respectively; in 1999 and 2001, Quicklaw reported 2,181 cases and 1,855 cases, respectively; and in 2016 and 2017, Quicklaw reported 2,073 and 1,974 cases, respectively.

35. By “significant predictor,” I mean a statistically significant relationship between the two

2. Longer trials (i.e. more court days), particular judges, and particular subjects lead to slower decisions. In other words, “court days,” “judge,” and “subject” are significant predictors of slower decisions; and
3. The relationship between “word count” and delivery time (“days to issue”) is bi-directional. In other words, they are significant predictors of each other (more words lead to longer delivery time and vice versa).

With a data scientist’s assistance,³⁶ two methods were used to test these hypotheses and to examine what dataset features are significant predictors of word count and delivery time: (1) classification and regression trees (also known as CART models); and (2) random forest models.³⁷ In short, both methods avail algorithmic machine learning and decision trees to analyze data and to offer predictions on significant predictor features.³⁸ One of the CART models that was run on the 1980 sample is included in Figure 1 on the next page to help explain and visualize part of the machine learning process.

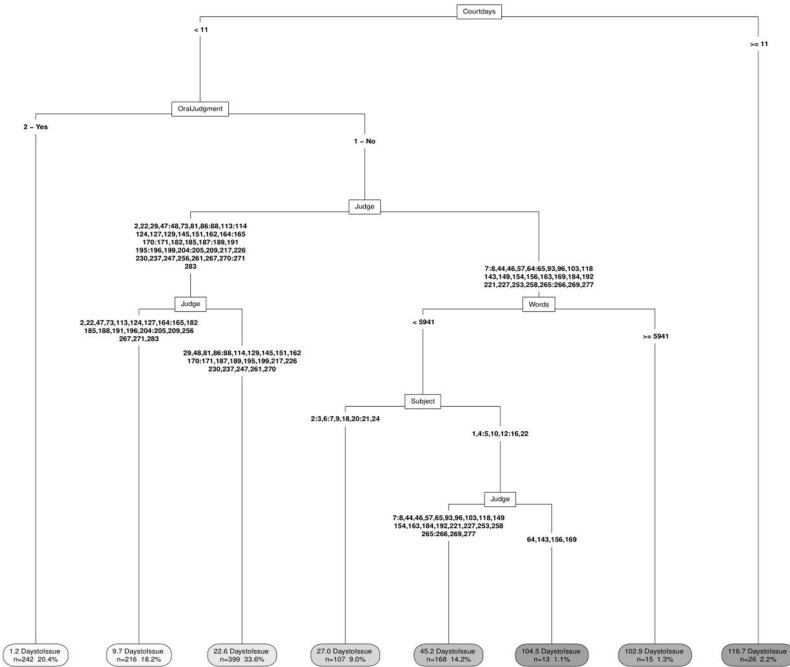
features. In contrast to some statistics or economics literature, data science uses the terms “predictor and response features” more than the terms “independent and dependent features.” As I am relying on random forest models—a data science tool—I have used the term “predictor” in the analysis rather than “independent feature.”

36. Thanks to Jonathan Kroening for your keen attention to detail, and for your interest in this project. You truly helped me understand my data and drastically improved my vision. And you have a great ability to help people improve themselves and their work. All errors are my own.

37. For a further explanation of how these models were designed and their metrics, see Khan, *supra* note 5, Appendix C. For a breakdown of the ranger coding package the model availed, see Marvin N Wright, Stefan Wagner & Philipp Probst, “Package ‘ranger’” (18 June 2022), online (pdf): <cran.r-project.org/web/packages/ranger/ranger.pdf> [perma.cc/8TX8-N7X3].

38. For an analysis and detailed explanation of classification and regression trees, see Leo Breiman et al, *Classification and Regression Trees* (New York: Chapman & Hall, 1993); see also Dan Steinberg, “CART: Classification and Regression Trees” in Xidong Wu & Vipin Kumar, eds, *The Top Ten Algorithms in Data Mining* (Boca Raton: CRC Press, 2009) 179. For an analysis and detailed explanation of random forest models, see Leo Breiman, “Random Forests” (January 2001), online (pdf): *Statistics Department, University of California Berkeley* <www.stat.berkeley.edu/~breiman/randomforest2001.pdf> [perma.cc/UD98-VGMB]; see also Andy Liaw & Matthew Wiener, “Classification and Regression by randomForest” (2002) 2:3 R News 18, online: <www.researchgate.net/publication/228451484> [perma.cc/U9SX-BK9D].

Figure 1: Sample CART Model



As you can see, CART models usefully visualize the hypothesis or problem you seek to understand by outputting a decision tree. Because CART models only build a single decision tree, however, they can be biased. For example, one dominant feature with many subcategories or levels (like the dataset’s judge feature)³⁹ may appear a biased number of times in the single decision tree.⁴⁰ In the above tree, judge appears at three of the seven nodes. Due to the potential for this bias, single CART models are less robust than random forest models. Accordingly, I did not rely on the single CART models to draw conclusions.

This article’s analysis relied exclusively on random forest models. Rather than relying on a single decision tree—like a single CART model does—random forest models create multiple, uncorrelated decision trees to increase analytical robustness and to avoid one predictor dominating the model. The article could have relied on classic regression analysis.⁴¹ Random forest models, however, can consistently outperform some

39. See A Hapfelmeier & K Ulm, “A New Variable Selection Approach Using Random Forest” (2013) 60 Computational Statistics & Data Analysis 50 at 51, DOI: <10.1016/j.csda.2012.09.020>.

40. See Khan, *supra* note 5, Appendix D for another CART model in a larger size.

41. For context, regression analysis is a form of predictive modeling that investigates relationships between dependent and independent features.

classic regression models.⁴² As Katz et al note, random forest models are “‘unreasonably effective’ in a wide array of supervised learning contexts,” and they consistently outperform common approaches to predictive analysis.⁴³ Similarly, as Breiman notes, random forests are extremely accurate and can supply more and better information than existing regression models.⁴⁴ Reliance on random forests models, therefore, enabled robust predictions about the statistically significant relationship between the examined features and delivery time and word count. That said, multiple modelling techniques and statistical methods must be run before governments, courts, or judges act upon statistical or machine learning conclusions. Redundancy and replication must guide reform.

Effectively, the random forest models employed this machine learning prediction sequence:

1. Random samples were drawn from my dataset (these samples are called bootstraps and are equivalent to the size of the original dataset but sampled with replacement—i.e. a single observation can present itself multiple times while others may be missing in the sample). This random sampling (with replacement) simulated collecting more “new data” and added variability so that predictions are robust when this simulated “new data” is introduced;
2. Those samples grew multiple decision trees (i.e. they created multiple CART models similar to the one you just observed);
3. At each split in one of those trees (commonly known as a node), the model randomly limited the number of predictors the tree could choose. So, the tree had to choose the best split from only *some* of the predictors. Coupled with what Point 1 noted, such limiting improves the model’s generalizability and predictive power;

42. See e.g. Raphael Couronné, Philipp Probst & Anne-Laure Boulesteix, “Random Forest Versus Logistic Regression: A Large-Scale Benchmark Experiment” (2018) 19:art 270 BMC Bioinformatics 1, DOI: <10.1186/s12859-018-2264-5>.

43. Daniel Martin Katz, Michael J Bommarito II & Josh Blackman, “A General Approach for Predicting the Behavior of the Supreme Court of the United States” (2017) 12:4 PLOS ONE 1 at 7, DOI: <10.1371/journal.pone.0174698>.

44. Leo Breiman, “RF/tools: A Class of Two-Eyed Algorithms” (May 2003) Siam Workshop, online (pdf): *Statistics Department, University of California Berkeley* <www.stat.berkeley.edu/~breiman/siamtalk2003.pdf> [perma.cc/H6YD-S9TL] [Breiman, “RF/tools”]. The upside of regression analysis over random forest analysis is that regression analysis can provide more interpretability. The criticism of random forest models is the lack of interpretability, i.e. the so-called “black box” problem. As Breiman notes, however, “[f]raming the question as the choice between accuracy and interpretability is an incorrect interpretation of what the goal of a statistical analysis is. The point of a model is to get useful information about the relation between the response and predictor features as well as other information about the data structure” (*ibid* at 29).

4. Multiple trees continued to grow, form nodes, and split until 1000 trees were created;
5. Those trees formed a “forest” of randomly created trees, hence the name random forest model;
6. This process of creating a random forest produced a ranking of predictors. Predictions were made by averaging all decision trees’ votes.⁴⁵ To validate and explain those predictions, the model relied on multiple evaluation metrics that are explained further in Appendix A,⁴⁶ including a cut-off p-value of <0.05 that avails the PIMP method;⁴⁷ and
7. The random forest models tested those predictions by permuting the predictors.⁴⁸ It measured any decrease in accuracy with the permuted predictor on a holdout data set (known as the out-of-bag set). For example, when the random forest model for delivery time permuted the “word count” feature and tested it on the out-of-bag sample, the prediction worsened by 36 days (in other words, the prediction was off by 36 days). Accordingly, that model indicated that “word count” is the most significant predictor of delivery time.

The models did not consider the impact of the registry feature or the five information features as predictors of word count or delivery time.⁴⁹ Registry was excluded because over 50 per cent of decisions occurred in one city, and the information features were excluded because they did not provide sufficient predictive power.⁵⁰ The models did, however, consider the impact of 11 of the earlier mentioned features—(6) how many court days were required for the decision; (7) the date the decision was released (the dates were amalgamated into “hearing year”); (8) how many days

45. For an accessible breakdown of random forests, see Leo Breiman & Adele Cutler, “Random Forests” (last visited 14 July 2022), online: *Statistics Department*, University of California Berkeley <www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#intro> [perma.cc/5UQC-TH4Y]; Breiman, “RF/tools,” *supra* note 44. See also Gérard Biau & Erwan Scornet, “A Random Forest Guided Tour” (2016) 25:2 TEST 197, DOI: <10.1007/s11749-016-0481-7>.

46. See Khan, *supra* note 5, Appendix C.

47. Andre Altmann et al, “Permutation Importance: A Corrected Feature Importance Measure” (2010) 26:10 *Bioinformatics* 1340 at 1340, DOI: <10.1093/bioinformatics/btq134> (“[t]he PIMP p-values are easier to interpret and provide a common measure that can be used to compare feature relevance among different models” at 1347).

48. Effectively, the predictor is scrambled to become noisy and kept in the model to observe the impact.

49. These information features are (1) case name; (2) Quicklaw citation; (4) whether the decision provided hearing information; (5) the day the hearing concluded; and (7) the date the decision was released.

50. One could arguably consider the month in which a decision is delivered.

lapsed before the decision was released; (9) judge; (10) moving party-type; (11) responding party-type; (12) subject; (13) word count; (14) use of headings; (15) self-reported information; and (16) appeal information—and the models were instructed to use those features as predictors of word count and delivery time.

Despite not relying on the single CART models to draw conclusions, they were invaluable in identifying the potential dominance of the “hearing year” (1980, 2000, and 2018) and the “judge” features. The “hearing year” was a dominant feature because decisions in 2018 are clearly longer and slower than decisions in 1980. As discussed with the CART example, the “judge” feature was also dominant (but less so) because the feature included over 290 subcategories of individual judges, so it naturally could appear more than other features despite the random forest’s sampling and limiting. Accordingly, nine random forest models were run to ensure robust predictions and to facilitate comparing/contrasting the evolution of significant predictors and the relationship of the “judge” feature with other features:

1. *entire dataset model*: including the “hearing year” and “judge” features provided a holistic analysis, but one that is potentially biased and correlated;
2. *entire dataset model with the “judge” feature excluded*: removing the “judge” feature facilitated the chance of less bias and correlation;
3. *entire dataset model with “hearing year” and “judge” features excluded*: removing the “hearing year” and “judge” features facilitated the chance of less bias and correlation;
4. *1980 models with the “judge” feature included and excluded*: including the “judge” feature facilitated observing the feature’s impact on other features, and removing the “judge” feature facilitated the chance of less bias and correlation;
5. *2000 models with “judge” feature included and excluded*: including the “judge” feature facilitated observing the feature’s impact on other features, and removing the “judge” feature facilitated the chance of less bias and correlation; and
6. *2018 models with “judge” feature included and excluded*: including the “judge” feature facilitated observing the feature’s impact on other features, and removing the “judge” feature facilitated the chance of less bias and correlation.

Despite the metrics on which this article relies and running multiple models, random forest models are just models. Notwithstanding the thoroughness of any study design, random forests are fallible, and they

do not indicate influence, causation, or correlation. They only indicate statistically significant relationships. As Breiman and Cutler aptly note, “[t]ake the output of random forests not as absolute truth, but as smart computer generated guesses that may be helpful in leading to a deeper understanding of the problem.”⁵¹ This point must be remembered in all empirical legal research that avails random forest prediction models (or really, any prediction model).

II. *Exploring descriptive trends and predictions about decision length and delivery time*

1. *Descriptive trends: on average, BCSC trial decisions’ current length rivals that of US Supreme Court decisions, and parties wait much longer than they used to*

a. *BCSC court time per matter: on average, no real increase; but now, judges encounter more long and “mega trials”*

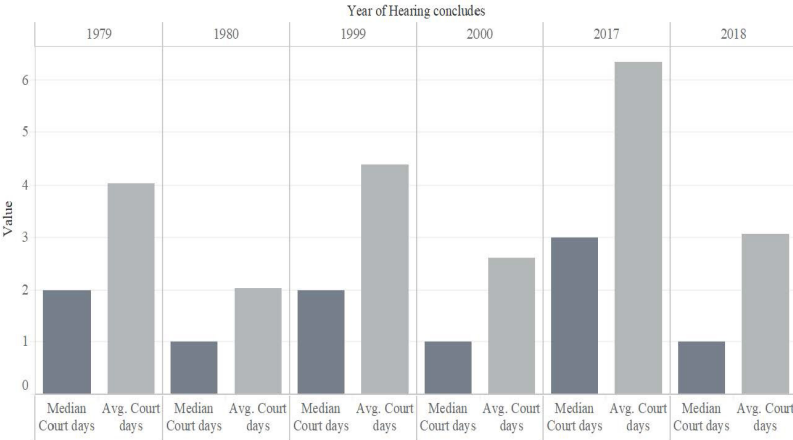
This dataset does not demonstrate a marked mean or median increase in the amount of court days (trial time) per decision like one might expect, as Figure 2 demonstrates:⁵²

- In 1979, the median court days was two, and the mean court days was four;
- In 1980, the median court days was one, and the mean court days was two;
- In 1999, the median court days was two, and the mean court days was four;
- In 2000, the median court days was one, and the mean court days was 2.6;
- In 2017, the median court days was three, and the mean court days was six; and
- In 2018, the median court days was one, and the mean court days was 3.8.

51. Leo Breiman & Adele Cutler, “Random Forests: A Philosophical Note” (last visited 14 July 2022), online: *Statistics Department, University of California Berkeley* <www.stat.berkeley.edu/~breiman/RandomForests/cc_philosophy.htm> [perma.cc/NBE9-YHQ7].

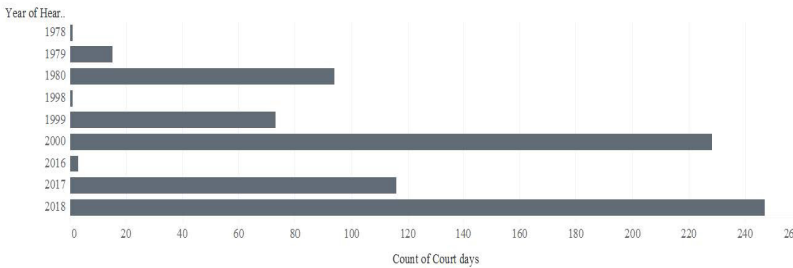
52. But this outcome could be because I did not parse out chambers matters from trial matters. On this topic, see a recent paper by LaRoche, Marais & Salter, *supra* note 17.

Figure 2: Evolution of mean and median court days



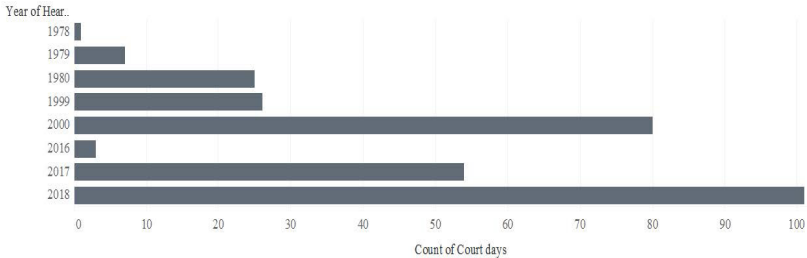
But the assumption, which I formed from anecdotal conversations with lawyers and judges, that 2018 would include more “mega” and longer trials than 1980 is true.⁵³ That said, despite more “mega” and longer trials, the median and mean of court days per decision did not increase.⁵⁴ These averages, however, should be approached with caution since the number of trials over five and ten days has markedly increased since 1980 as Figures 3 and 4 demonstrate—e.g. well over 200 five-day trials in 2018 and just over 100 ten-day trials in 2018.

Figure 3: Amount of trials over 5 days



53. In 1979, the longest trial was 45 days; in 1980, it was 50 days; in 1999, it was 40 days; in 2000, it was 91 days; in 2017, it was 73 days; and in 2018, it was 125 days. I thank Geoff Cowper, QC for his fantastic insight on this point and the issue of longer and “mega” trials. See Khan, *supra* note 5, Appendix E, Figure 12 for the court days distribution broken down in bins of five for 1980, 2000, and 2018.

54. *Ibid*, Appendix E, Figure 13 for the breakdown of the entire dataset’s court days by subject matter. Family, criminal, civil procedure, personal injury, and tort are the top five and occupied the most days.

Figure 4: Amount of trials over ten days

Recognizing this increase of trials over five and ten days is important. The random forest models I will shortly discuss suggest the “court days” feature is the most significant predictor of word count, and one of the most significant predictors of delivery time. Put another way, “mega” and longer trials can have big effects because they often require grappling with significant amounts of oral and written evidence. The longer judges spend in trial necessarily means that judges spend less time on other decisions or that they must altogether delay writing other decisions.⁵⁵ So when judges are assigned one (or more) “mega” or long trials, they may have no or far less time to write decisions for shorter trials or simple matters.

b. *BCSC decision length and delivery time: on average, judges in 2018 take three times longer to issue decisions three times longer than their 1980 colleagues*

This dataset demonstrates a clear increase in the mean and median of decision length and delivery time. Sometimes, a picture is worth a thousand words. In this case, a picture is worth three times as many words and three times as many days. On median, BCSC judges in 2018 take three times longer to issue decisions that are three times longer than their 1980 colleagues, as Figures 5 and 6 demonstrate:

- In 1980, median word count was 1,483 words, and the median delivery time was eight days.⁵⁶
- From 1980 to 2000, median word count increased almost 1000 words to 2,434 words, and the median delivery time increased

55. See e.g. BV Harris, “The Continuing Struggle with the Nuanced Obligation on Judges to Provide Reasons for Their Decisions” (2016) 132 Law Q Rev 216 at 224-225, 233.

56. The mean word count was 2,017 words, and the mean delivery time was 23 days with all observations included. Considering the greatest number of observations without hearing information occurred in 1980, I tested the median word count and delivery time without those observations: the median word count increased slightly to 1,575 words, and the median delivery time increased to 15 days. The exclusion also increased the mean word count to 2,171 days and the mean delivery time to 29.98 days.

12 days to 20 days.⁵⁷ When comparing decisions by subject—
from 1980 to 2000—the highest median word count more than
doubled,⁵⁸ and the highest mean word count almost doubled.⁵⁹
The slowest median delivery time increased four-fold,⁶⁰ and the
slowest mean delivery time doubled.⁶¹

- From 2000 to 2018, median word count increased almost 2,300
words to 4,740 words, and the median delivery time increased
eight days to 28 days.⁶² When comparing decisions by subject—
from 1980 to 2018—the highest median word count increased
almost six-fold,⁶³ and the highest mean word count increased
four-fold.⁶⁴ The slowest median delivery time increased more
than eight-fold,⁶⁵ and the slowest mean delivery time increased
almost four-fold.⁶⁶

57. The mean word count was 3,376 words, and the mean delivery time was 42 days. Excluding the observations without hearing information increased the median word count to 2,578 words and the median delivery time to 25 days. The exclusion also increased the mean word count to 3,566 words and the mean delivery time to 48 days.

58. The 1980 sample for contracts decisions relied on 108 observations, and the 2000 sample for tort decisions relied on 92 observations.

59. The 1980 sample for tort decisions relied on 56 observations, and the 2000 sample for tort decisions relied on 92 observations.

60. The 1980 sample for administrative decisions relied on 59 observations, and the 2000 sample for tort decisions relied on 92 observations.

61. The 1980 sample for contracts decisions relied on 108 observations, and the 2000 sample for tort decisions relied on 92 observations.

62. The mean word count was 6,777 words, and the mean delivery time was 56.75 days. Excluding the observations without hearing information didn't increase the median hearing time, but it increased the median word count by 22 words to 4,762 words. The exclusion increased the mean word count to 6,799 words and the mean delivery time to 57 days.

63. The 1980 sample for contracts decisions relied on 108 observations, and the 2018 sample for tort decisions relied on 44 decisions.

64. The 1980 sample for tort decisions relied on 56 observations, and the 2018 sample for tort decisions relied on 44 observations.

65. The 1980 sample for administrative decisions relied on 59 observations, and the 2018 sample for tort decisions relied on 44 observations.

66. The 1980 sample for contracts decisions relied on 108 observations, and the 2018 sample for tort decisions relied on 44 observations.

Figure 5: Evolution of decisions' delivery mean and median time

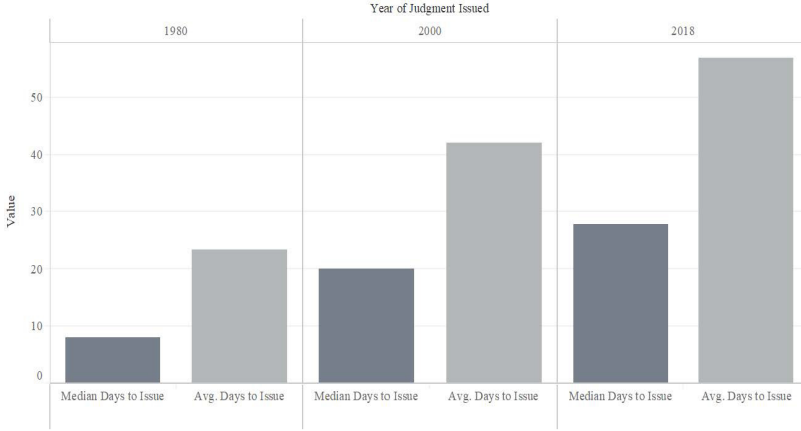
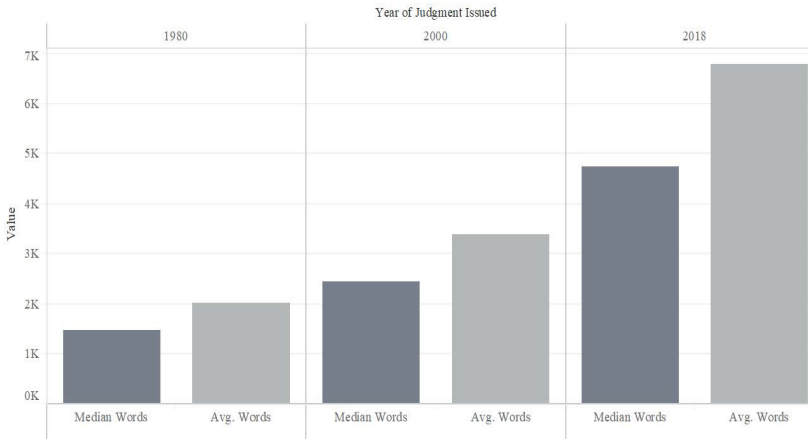


Figure 6: Evolution of decisions' mean and median length



The current length of BCSC trial decisions rivals that of US Supreme Court decisions (these averages are actual decisions versus certiorari decisions).⁶⁷ Increased length likely increases decisions' complexity and may have a significant economic impact. Such complexity does not increase access to justice and may even decrease it. Increased length is likely related to increased delivery times. The random forest models suggest as much, a point I will shortly discuss.

67. While the median word count for 2018 in the dataset is 4,740 words, the mean word count is 6,777 words. That mean rivals the mean of US Supreme Court decisions in early 2018. See Adam Feldman, "Empirical SCOTUS: An opinion is worth at least a thousand words (Corrected)" (3 April 2018), online (blog): *SCOTUSblog* <www.scotusblog.com/2018/04/empirical-scotus-an-opinion-is-worth-at-least-a-thousand-words/> [perma.cc/FLH3-LT9D].

Many reasons could explain the trend of longer and slower judicial decisions.⁶⁸ I will discuss four.

First, perhaps the increased length and complexity (and thereby slower delivery times) is a response to legal realism's criticisms. As Simon argues, the judicial process has remained in daunting disarray ever since "American realism's onslaught on formalist jurisprudence."⁶⁹

Second, perhaps the academicization of law has also contributed—e.g. maybe more judges are reading and responding to academics' claims and critiques and perhaps lawyers are making more "academic" or "scholarly" legal arguments.⁷⁰ The increasing number of written Canadian decisions and their increased length over the last 40 years coincides with the development of the academic institution of law in Canada, which exploded over the last 40 years.⁷¹ Canadian legal discourse and commentary has become far more scholarly since the 1983 Arthurs Committee on Legal Education, Law, and Learning.⁷² An example is the proliferation of Canadian law professors with a PhD since that report.⁷³ Having a PhD is now the expected standard for entering the legal academy.⁷⁴ Increasingly, highly-educated academics and scholars are educating law students instead of the previous historical reality where many academics were "just" lawyers with practical, practice experience and specialized expertise (or an LLM or other post-graduate degree). This academization of law could

68. See Khan, *supra* note 5 at 25-26 for a further discussion of these points—e.g. appeal aversion, the advent of the Charter, robust standard for sufficient reasons, etc.

69. Dan Simon, "A Psychological Model of Judicial Decision Making" (1998) 30:1 Rutgers LJ 1 at 3; John N Drobak & Douglass C North, "Understanding Judicial Decision-Making: The Importance of Constraints on Non-Rational Deliberations" (2008) 26 Wash UJL & Pol'y 131, online: <journals.library.wustl.edu/lawpolicy/article/id/1388/> [perma.cc/98QK-58VX] ("the Legal Realists made a valuable contribution to jurisprudence by their emphasis on the importance of the hidden factors in judicial decision-making. Their legacy remains influential today" at 136).

70. See e.g. Michel Bastarache, "The Role of Academics and Legal Theory in Judicial Decision-Making" 1999 37:3 Alta L Rev 739, DOI: <10.29173/alr1470>.

71. See Wilbur Fee Bowker & Timothy J Christian, "Legal Education" (last modified 4 March 2015), online: *The Canadian Encyclopedia* <www.thecanadianencyclopedia.ca/en/article/legal-education> [perma.cc/PZC8-MLWV].

72. See e.g. *Law and Learning: Report to the Social Sciences and Humanities Research Council of Canada by the Consultative Group on Research and Education in Law*, Catalogue No CR22-6/1983 (Ottawa: Minister of Supply and Services Canada, 1983); *Task Force on the Canadian Common Law Degree: Final Report* (Ottawa: Federation of Law Societies of Canada, 2009), online (pdf): <docs.flsc.ca/APPRTaskForceReportOct2009.pdf> [perma.cc/G5VS-UT46].

73. Dia Dabby, Bethany Hastie & Jocelyn Stacey, "Doctoral Studies in Law: From the Inside Out" (2016) 39:1 Dalhousie LJ 221 at 223, online: <digitalcommons.schulichlaw.dal.ca/dlj/vol39/iss1/7/> [perma.cc/YZ93-EVWT].

74. Craig Forcese, "Want to be a Law Prof? Data on Whether You Should Do a Doctorate" (9 July 2014), online (blog): *Craig Forcese* <www.craigforcese.com/blog/2014/7/9/want-to-be-a-law-prof-data-on-whether-you-should-do-a-doctor.html> [perma.cc/NM4W-X2J2].

contribute to a perceived academization of litigation and decisions.⁷⁵ Either consciously or unconsciously, judges may feel increased pressure to write academically worthy decisions versus terse, rote, legally sufficient decisions notwithstanding judges being very different writers than academics. And lawyers may equally feel the pressure to put forward more “academic” arguments that judges then must respond to.

Third, perhaps judges write longer decisions and take longer to issue them because Canadian law has become complicated and parties run longer trials.⁷⁶ Further testing on the relationship must occur, including building models with pan-Canadian data to predict which decisions are most likely to be lengthy and succumb to slower delivery times. Trials becoming longer could be because the law is more complicated. But if the law is more complicated today than in the 1980s, such complication could be compromising Canada’s rule of law.⁷⁷ Unless you disagree with the continual pursuit of the rule of law’s optimistic promise for clarity,⁷⁸ Canada’s common law should not be getting more complicated. The question then becomes who is perhaps responsible and how might we make it less complicated.⁷⁹ But perhaps the common law has not become more complicated. Perhaps some trials and decisions have become longer because of the proliferation of published decisions and decisions’ increased length. So even if actual common law legal tests are not more complicated, the proliferation of jurisprudence nonetheless likely increases complexity, inconsistency, and delay.⁸⁰ Computers’ impact here appears clear. They enable judges to do two things more easily: issue and publish more written

75. I thank Martin Taylor, QC for a fantastic discussion on this idea and for causing me to think about how education may have impacted what judges do now versus what they did 40 years ago. I also thank Colin Campbell, QC. For academic commentary, see e.g. Daphne Barak-Erez, “Writing Law: Reflections on Judicial Decisions and Academic Scholarship” (2015) 41:1 Queen’s LJ 255, online: <journal.queenslaw.ca/sites/qljwww/files/Issues/Vol%2041%20i1/7.%20Barak-Erez.pdf> [perma.cc/N6CH-K96U].

76. I have repeatedly heard these comments from academics and practitioners.

77. Of course, edge cases may always exist even if the law is extremely accessible, clear, and uncomplicated. And such cases may include copious amounts of evidence and long trials (e.g. constitutional issues or novel factual and legal situations). But many, if not most, trials are likely not edge cases (this point and claim should also be subjected to data and analysis). Yet—perhaps because of the power of the false-uniqueness effect (or similar “specialness” biases)—lawyers and their clients believe their cases are more special or unique than they are. In other words, they believe their average cases are edge cases and run them as such, likely with lots of evidence and long trials.

78. See e.g. Jeremy Waldron, “The Rule of Law” (22 June 2016), online: The Stanford Encyclopedia of Philosophy <plato.stanford.edu/archives/sum2020/entries/rule-of-law/> [perma.cc/U6F3-3YPN].

79. See e.g. Jamie Murray, Thomas E Webb & Steven Wheatley, “Encountering Law’s Complexity” in Jamie Murray, Thomas E Webb & Steven Wheatley, eds, *Complexity Theory and Law: Mapping An Emergent Experience* (London, UK: Routledge, 2019) 3 at 12.

80. See e.g. Martin Teplitsky, “Excessive Cost and Delay: Is There a Solution?” (2000) 19:2 *Advocates’ Society J* 5.

decisions; and execute the court-reporter phenomenon that ultimately leads to longer, more complicated decisions where judges include more facts and law.⁸¹ With surplusage of decisions, facts, and law, the potential for error, delay, and inconsistency likely increases. Coupled together, these phenomena can contribute to the increased length of decisions, delivery time, and “mega” trials. If decisions’ content and structure aren’t simplified—e.g. standardized or streamlined—these increases may simply continue to perpetuate, especially as judges may be seeking to conform to standards of longer decisions because they believe that long, detailed decisions are indicia of good work.⁸²

Fourth, perhaps the adage “if I had more time, I would have written a shorter letter” applies to judicial decision-writing. But the idea that more time would help is a hypothesis that courts must test. For example, perhaps more writing time could worsen the current situation. The adage “work expands so as to fill the time available for completion”⁸³ may equally apply to judicial decision-making and decision-writing. More writing time may simply encourage judges to write even longer, more complicated, slower decisions. Giving judges more time does not mean they would use it to write shorter, clearer, more accessible decisions. Put simply, more writing time will not guarantee shorter or speedier decision-making. In other words, the question “if I had more time, would I have written a shorter and faster decision?” remains a live and important question.

Regardless of the reason for longer, slower decisions, more judges are likely spending more time writing decisions versus sitting in court—recall Justice Brown’s experience from the introduction.⁸⁴ And all decisions’ readers are likely spending far longer reading decisions (if they actually

81. Edward Berry, *Writing Reasons: A Handbook for Judges* (LexisNexis, Toronto, 2015) (the “temptation to tell the whole story (much of which may be irrelevant) proves sometimes irresistible” at 36).

82. See Lee Epstein, William M Landes & Richard A Posner, *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice* (Cambridge, MA: Harvard University Press, 2013) (governments likely want judges who will do a good job and who subscribe to the “high commitment” culture: they want people who will “identify with the mission” at 34-35). See also Robert J Sharpe, *Good Judgment: Making Judicial Decisions* (Toronto: University of Toronto Press, 2018) (writing extrajudicially, Sharpe notes that “my scholarly work was largely aimed at helping lawyers and judges solve practical and immediate day-to-day problems ... it seemed better to me that whatever knowledge and skills I had acquired would be better deployed if I were a judge. So when the chance came, I opted for a judicial career” at 12). I assume other judges may hold this view and similarly want to do a good job. Further research should be done on reviewing judicial applications about their motivations for becoming a judge and how they have played out practically.

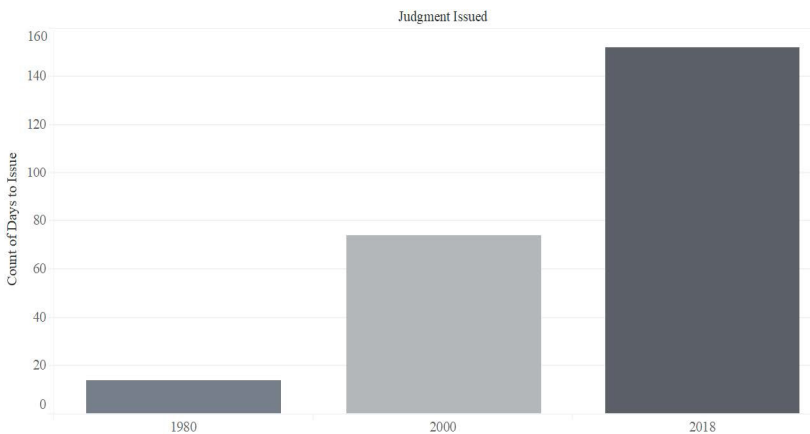
83. Cyril Northcote Parkinson, “Parkinson’s Law,” *The Economist* (19 November 1955), online: <www.economist.com/news/1955/11/19/parkinsons-law> [perma.cc/S8W4-SDQL].

84. No current appropriate ratio exists as far as I am aware. Perhaps judges should spend more time writing and thinking versus sitting in court. Considering the current extent of delay in Canadian courts, however, such a ratio would be curious.

have time to even read them). The possible effect is at least two-fold: parties likely wait longer for their decisions to be issued, and judges likely issue longer, more complex decisions. Both effects may lead to longer preparation time for lawyers and more protracted and expensive litigation. Lawyers then download these costs to their clients who pay more money for slower, more complex results. Theoretically, lawyers should read longer decisions when preparing for litigation. Practically, however, the length and opportunity cost likely force lawyers to skim decisions. Other users must also expend time trying to digest lengthy, unstandardized decisions. Finally, the issue is not just an economic one. Slow, lengthy, complex decisions likely do not further the aims of Canada’s rule of law—i.e. “certainty, accessibility, intelligibility, clarity and predictability.”⁸⁵

c. *BCSC decisions taking longer than 180 days: 1.1 per cent of decisions used to take 180 days or more; now, eight per cent of decisions do*
Another related and concerning upward trend is the increase of BCSC decisions taking 180 days⁸⁶ or longer to be released that Figure 7 demonstrates:

- In 1980, 1.1 per cent of decisions were released after 180 days or more;
- In 2000, 3.8 per cent of decisions were released after 180 days or more; and
- In 2018, 8.0 per cent of decisions were released after 180 days or more.



85. *R v Ferguson*, 2008 SCC 6 at para 69.

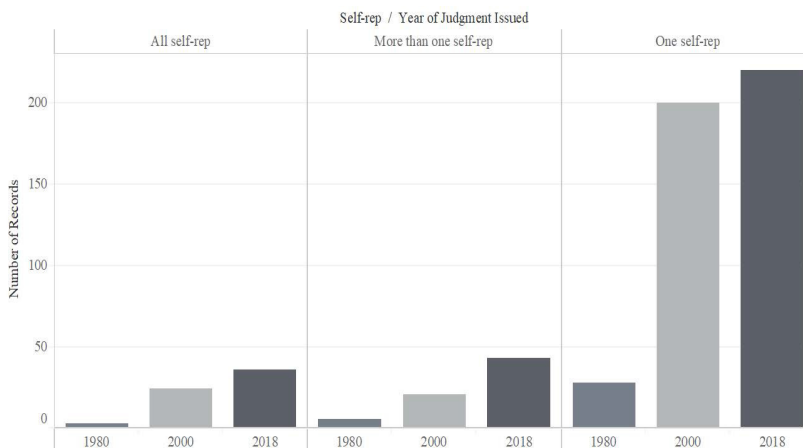
86. See *Ethical Principles for Judges* (Ottawa: Canadian Judicial Council, 2021) at 21, online (pdf): <cjc-ccm.ca/sites/default/files/documents/2021/CJC_20-301_Ethical-Principles_Bilingual_Final.pdf> [perma.cc/35RH-XCRZ] (the Canadian Judicial Council recognizes that decision-writing is a time-consuming and difficult exercise. Accordingly, it prescribes a six-month timeline to deliver all decisions, except in special circumstances).

I cannot say more than increased justice delayed is increased justice denied.

- d. *BCSC self-representation: three per cent of litigants used to appear without lawyers; now, at least 16 per cent of reported decisions involve some type of self-representation*

Another concerning upward trend is the increase of self-representation over the past 40 years that Figure 8 demonstrates:

- In 1980, only 3.0 per cent of decisions had some type of self-representation;
- In 2000, 12.7 per cent of decisions had some type of self-representation; and
- In 2018, 15.9 per cent of decisions had some type of self-



Increased self-representation may be related to the increase of longer decisions and more complex decisions. Further research on this relationship is needed. As just alluded, longer and more complex decisions may increase legal services’ price tag,⁸⁷ a large part of Canada’s access to

87. See e.g. Teplitzky, *supra* note 80 (“[a]nother factor contributing to increased cost is both the large volume of reported decisions and the availability on the Internet of all written decisions, whether reported or not. The task of comprehensive legal research is much more difficult than it was in 1966. I recall Dean Wright, the long-time editor of the D.L.R.s and the O.R.s, telling our class that he would not publish cases which were ‘wrong.’ The ‘law’ was found in the D.L.R.s. Of course, back then there were few, if any, specialized reports. Today, publishers want to sell more volumes and publish many decisions that do not seem to add anything to the existing jurisprudence. I am not sure what, if anything, can be done about this problem. I suspect nothing. We should at least realize that, in the final analysis, the client pays the tab directly for additional research and indirectly through the higher overhead costs of maintaining law libraries” at 19 [emphasis added]).

justice crisis.⁸⁸ And for the individuals that proceed to litigation without a lawyer, longer decisions and more complex decisions likely make the litigation experience harder for them,⁸⁹ and by extension, the judges who must resolve their disputes.

e. *BCSC decisions' structure: historically, almost no decisions used headings; now, 80 per cent of decisions do*

Decisions' structure over the last 40 years has also changed. Paragraphs were introduced and are now the norm, and headings have also become ubiquitous as Figure 9 demonstrates:

- In 1980, only 8.0 per cent of decisions availed headings;
- In 2000, 45 per cent of decisions availed headings; and
- In 2018, 81 per cent of decisions availed headings.

Such use of headings, however, is not standardized. And without standardization, their use may not increase readability, clarity, or length. In fact, decisions with headings were considerably longer than decisions without headings in 1980, 2000, and 2018 (but this reality likely exists because decisions generally became longer),⁹⁰ and they also had considerably slower delivery times.⁹¹

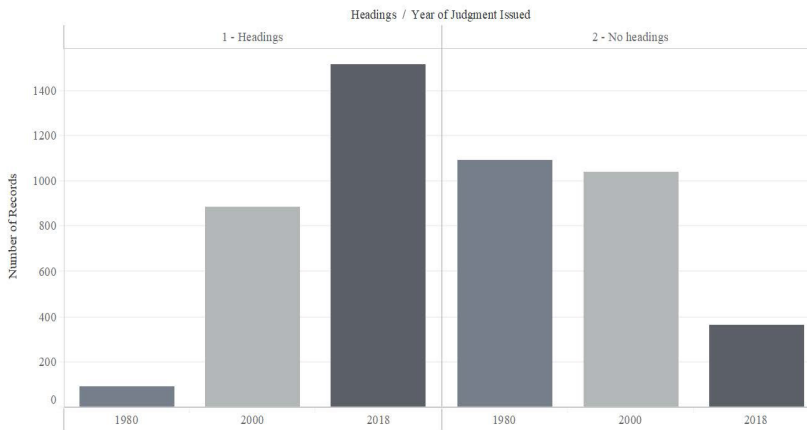
88. Trevor Farrow et al, "Everyday Legal Problems and the Cost of Justice in Canada: Overview Report" (2016) 12:12 Osgoode Leg Studies Research Paper No 57, Can Forum on Civ Justice 1 at 3, DOI: <10.2139/ssrn.2795672>.

89. Margarita Dvorkina & Julie Macfarlane, "Reading and Understanding Case Reports: A Guide for Self-Represented Litigants" (2017) at 4, online (pdf): *The National Self-Represented Litigants Project* <representingyourselfcanada.com/wp-content/uploads/2017/09/Reading-A-Case-Primer.pdf> [perma.cc/254E-5Z4T].

90. Decisions' median word count with headings in 1980 was 2,945 words, in 2000 was 3,598 words, and in 2018 was 5,705 words. Decisions' median word count without headings in 1980 was 1,357 words, in 2000 was 1,629 words, and in 2018 was 2,047 words. The mean word counts were also all higher for decisions with headings.

91. Decisions' median delivery time with headings in 1980 was 26 days, in 2000 was 32 days, and in 2018 was 35 days. Decisions' median delivery time without headings in 1980 was seven days, in 2000 was 11 days, and in 2018 was two days. The mean delivery times were also all higher for decisions with headings.

Figure 9: Increased use of headings



f. *Final takeaways: we have much to learn about oral decisions, judges' writing practices, and finding ways to solve simple errors in trial decisions*

Finally, in creating this dataset, I noted four noteworthy points:

- Oral decisions still occur at the end of some trials.⁹² But oral decisions are not issued like they were in 1980. In fact, the median delivery time for oral decisions went from zero days in 1980 to two days in 2018, and the mean delivery time went from 1.25 days in 1980 to 13 days in 2018.⁹³ The suggestion of more oral decisions to reduce delay has been proposed,⁹⁴ and the random forest model conclusions we are about to discuss indicate the suggestion is meritorious. But oral decisions delivered 13 days after trial are no longer true oral decisions. Rather, they are just written decisions delivered orally in court. Nonetheless, those decisions are still faster than traditional written decisions.
- Per incuriam errors still occur despite technology that should preclude them.⁹⁵

92. See e.g. *Bains v Chatakanonda*, 2018 BCSC 2412.

93. See Khan, *supra* note 5, Appendix E, Figure 14.

94. See e.g. Michael Kirby, "Ex Tempore Judgments: Reasons on the Run" (1995) 25:2 UWA L Rev 213 at 213-14, 230-231, online: <www5.austlii.edu.au/au/journals/UWALawRw/1995/18.html> [perma.cc/VUU7-CKKJ]; JE Côté, "The Oral Judgment Practice in Canadian Appellate Courts" (2003) 5:2 J App Pr & Pro 435, online: <journals.librarypublishing.arizona.edu/appellate/article/id/2811/> [perma.cc/N5EJ-ZCDV].

95. See e.g. *Saltman v Sharples Contracting Ltd*, 2018 BCSC 883.

- Some judges appear to be writing decisions as longer civil and criminal trials proceed as they issue decisions at the end of a trial or very soon after.⁹⁶ Whether this practice is efficient versus deciding too quickly is unknown.
- Some judges batch issue decisions. This practice makes sense considering that some judges get writing weeks.⁹⁷ One must wonder, however, if or how the practice affects reasoning—e.g. does some cross-contamination of reasoning occur (e.g. some recency effects or bias)? And if such contamination is occurring, how might we prevent or alleviate it?

2. *Predictive analysis: if your litigation requires more than a few court days, you will likely have to wait a long time for a lengthy decision*

The random forest models of the entire dataset showcase useful overall predictive conclusions for word count and delivery time. They demonstrate that only facets of my initial three hypotheses about longer and slower decisions were valid (using this dataset and model):

- *Hypothesis 1*: only one of the three hypothesized features for longer decisions—“court days”—significantly predicted word count. But the other two features—“judge” and “subject”—did not.
- *Hypothesis 2*: only two of the three hypothesized features for slower decisions—“court days” and “judge”—significantly predicted delivery time. But the other feature—“subject”—did not.
- *Hypothesis 3*: the relationship between “word count” and “days to issue” is not bi-directional: “word count” significantly predicts “days to issue” (i.e. “delivery time”), but not the converse.

a. *Length: the longer your trial, the longer your decision*

Regardless of whether the earlier discussed dominant features (“judge” and “hearing year”) were included or excluded, the result was the same in all three models of the entire dataset: the “court days” feature was the only significant predictor of word count.⁹⁸ Accordingly, the relationship between “court days” and “word count” appears robust. When the random forest models of this entire dataset permuted the “court days” feature, the

96. See e.g. *Dinnissen v Lee*, 2018 BCSC 2233; *R v Dosanjh*, 2018 BCSC 1153; *R v Seepersad*, 2018 BCSC 2252; *R v Rigo*, 2018 BCSC 2015.

97. See e.g. “Callaghan Scheduling Convention” (October 1992), online: *Superior Court of Justice* <www.ontariocourts.ca/scj/news/callaghan/> [perma.cc/DC5D-D28H].

98. See Khan, *supra* note 5, Appendix E, Figures 15-17.

three models demonstrate that the prediction was off by 4,390 words (the mean of all three models).

In sum, court days appear to be a significant predictor of word count—the more court days, the longer the decision. While court days have not markedly increased, on mean or median, from 1980 to 2018, the number of longer trials clearly has (2018 has far more trials with more than five court days), and this increase likely explains this feature’s significance.

b. *Delivery time: the longer your trial and the longer your decision, the longer your wait*

Regardless of whether the earlier discussed dominant features (“judge” and “hearing year”) were included or excluded, one result was the same in all three models of the entire dataset: the “word count” and “court days” features were significant predictors of delivery time.⁹⁹ Accordingly, the relationship between “word count,” “court days,” and delivery time appears robust. When the models permuted the “word count” feature, the three models demonstrate that the delivery time prediction was off by 33 days (the mean of all three models). Similarly, when the models permuted “court days,” the prediction was off by 28 days (the mean of all the three models).

The results for the other features become slightly less clear in the models of the entire dataset that included and excluded the “judge” and “hearing year” features:

- *“Judge” and “hearing year” included:* “oral judgment,” “hearing year,” and “judge” are significant predictors of delivery time. When the “oral judgment” feature was permuted, the prediction was off by 24 days. Considering the potential dominance of the “hearing year” and “judge” features, one must approach those predictors with some caution. Decisions’ delivery time has markedly increased over the past 40 years, so the “hearing year” feature’s classification as a significant predictor makes sense. But due to the historical evolution, the prediction is not overly informative. The “judge” feature can be approached with slightly less caution since it was not overly dominant in the individual 1980 and 2000 models, which I will discuss in a moment. When the “judge” feature was permuted, the prediction was off by 16 days.¹⁰⁰

99. *Ibid* (see Appendix E, Figures 18-20).

100. *Ibid* (see Appendix E, Figure 18).

- “*Judge*” excluded: the result is the same as the “judges” and “hearing year” included model except permuting “oral judgment” made the prediction worsen by 22 days (versus 24 days when judges were not excluded).¹⁰¹ “Judge” is clearly no longer a predictor and “hearing year” is also no longer a predictor.
- “*Judge*” and “*hearing year*” excluded: only “word count” and “court days” are significant. “Subject” almost becomes statistically significant,¹⁰² but it did not pass the requisite statistical threshold under the PIMP method.¹⁰³

In sum, “word count” and “court days” are significant predictors of delivery time—the more words and court days, the slower the delivery time. The phenomenon of increased words over the last 40 years explains, at least partly, the evolution of slower delivered decisions.

III. *Analyzing the granular evolution of BCSC decisions’ length and delivery time and judges’ workload over the last 40 years, and suggesting reforms to test*

To further understand the evolution of BCSC decisions, Part III examines the specific trends of word count, delivery time, and judges’ workload. When examining these trends and how they relate to subject matters (e.g. tort disputes), I relied only on median and mean word counts and delivery times with more than ten decisions. In some cases, particular subjects had higher medians and means than the ones I ranked as highest, but those averages were based on small sample sizes (so I did not classify them as the highest).¹⁰⁴

1. *Length: decisions are no longer relatively short*

1980:¹⁰⁵ Contract decisions had the highest median word count (2,208 words).¹⁰⁶ Civil procedure decisions (1,001 words) had the lowest median word count.¹⁰⁷ Tort decisions had the highest mean word count (3,656 words). Appeal decisions had the lowest mean word count (1,192 words).¹⁰⁸

101. *Ibid* (see Appendix E, Figure 19).

102. *Ibid* (see Appendix E, Figure 20).

103. See *supra* note 45 (I set a cut off p-value of <.05 to avail the PIMP method). If a feature was outside of the cut-off threshold, I did not consider it in the analysis. But Khan, *supra* note 5, Appendix E includes all features and their ranking.

104. To see all figures, see Khan, *supra* note 5, Appendix E.

105. *Ibid* (see Appendix E, Figure 21 for the mean and median breakdowns of word count in 1980).

106. Technically, IP decisions had the highest median word count (2,742 words), but that median is only based on two observations. So, this result was disregarded.

107. Technically, tax decisions had the lowest median word count (722 words), but that median is only based on one observation. So, this result was disregarded.

108. Technically, tax decisions had the lowest mean word count (722 words), but that mean is only

The random forest models for 1980 indicate that “court days” are the only significant predictor of word count.¹⁰⁹ Including versus excluding judges made no difference to the predictions.

2000:¹¹⁰ Tort decisions had the highest median word count (4,885 words).¹¹¹ Civil procedure decisions had the lowest median word count (1,907 words). Tort decisions had the highest mean word count (6,556 words).¹¹² Debtor/creditor decisions had the lowest mean word count (2,156 words). The random forest models for 2000 indicate that “court days” and “headings” are significant predictors of word count (in 1980, only “court days” was a significant predictor).¹¹³ The addition of “headings” as a significant predictor makes sense in comparison to its absence in 1980 since hardly any judges used headings in 1980. Including versus excluding judges made no difference to the predictions.

2018:¹¹⁴ Tort decisions had the highest median word count (12,475 words).¹¹⁵ Civil procedure had the lowest median word count (3,096 words).¹¹⁶ Tort decisions had the highest mean word count (14,981 words).¹¹⁷ Civil procedure decisions had the lowest mean word count (4,415 words).¹¹⁸ The random forest model for 2018 with judges included indicates that “court days” and “headings” are significant predictors of word count. With judges excluded, court days are the only significant predictor of word count (in 2000, headings remained significant even with judges removed).¹¹⁹ The relationship between the “headings” and “judge” features makes sense as some judges use headings more frequently than others. As discussed, all

based on one observation. So, this result was disregarded.

109. Khan, *supra* note 5, Appendix E, Figures 22-23.

110. *Ibid* (see Appendix E, Figure 24 for the mean and median breakdowns of word count in 2000).

111. Technically, Aboriginal decisions had the highest median word count (19,487 words), but that median is only based on two observations. So this result was disregarded.

112. Technically, Aboriginal decisions had the highest mean word count (19,487 words), but that mean is only based on two observations, so this result was disregarded. Constitutional (12,342 words) and public law (6,835 words) had higher means than tort decisions, but those results are based on six and three decisions, respectively. So these results were disregarded.

113. Khan, *supra* note 5, Appendix E, Figures 25-26.

114. *Ibid* (see Appendix E, Figure 27 for the mean and median breakdowns of word count in 2018).

115. Technically, Aboriginal decisions (130,321 words) and constitutional decisions (13,491) had higher median word counts, but those medians are only based on one and five observation(s), respectively. So these results were disregarded.

116. Technically, public law (2,993 words) and OCPA decisions (449 words) had the lowest median word counts, but the medians are each based on one observation. So, these results were disregarded.

117. Aboriginal decisions (130,321 words) and constitutional decisions (13,917) had higher mean word counts, but those means are only based on one and five observation(s), respectively. So these results were disregarded.

118. Technically, public law (2,993 words) and OPCA decisions (449 words) had the lowest mean word count, but these means are each based on one observation. So these results were disregarded.

119. Khan, *supra* note 5, Appendix E, Figures 28-29.

decisions with headings in 1980, 2000, and 2018 are longer than those without headings, and headings were increasingly used in 2000 and 2018. The significance of “headings” in both 2000 and 2018, therefore, makes sense. Notably, the “court days” predictions have remained constant since the 1980 sample. Accordingly, the prediction appears robust. It also again demonstrates that Hypothesis 1 was partially valid (using this dataset and model).¹²⁰

Summary and reform suggestion for testing: The dominance of tort decisions as the longest decisions in all sample years is curious. Either tort law is unclear, or parties are generating large evidentiary records (the importance of “court days” as a significant predictor of length in the random forest models suggests the latter is likely true). Either way, judges are increasingly issuing lengthy decisions about private law disputes between two parties that did not involve accidental physical injury, so these decisions should include less expert evidence (unless it was medical malpractice).¹²¹ That civil procedure decisions were the shortest is promising. It suggests that such issues are perhaps clearer and that parties put forward less evidence. Both tort and civil procedure are likely amenable to reforms: judges and parties need more guidance on content and structure in tort disputes and tort decisions, and civil procedure decisions may be open to more standardization. Finally, the relationship of “court days” to decision length is likely the dominant consideration in improving decisions’ delivery time. Reducing the need for “mega” and longer trials could be one way to reduce decisions’ length. But this intuition must be rigorously tested.

2. *Delivery time: decisions are no longer delivered relatively fast*

1980:¹²² Administrative decisions had the slowest median delivery time (18 days).¹²³ Criminal decisions had the quickest median delivery time (zero days). Contract decisions had the slowest mean delivery time (41 days).¹²⁴

120. Similar to the random forest models of the entire data set, only one of the three hypothesized features for longer decisions—“court days”—consistently predicted word count. The “judge” and “subject” features did not.

121. Personal injury was coded separately from tort. Some of these tort decisions included medical malpractice where multiple expert opinions would likely be present.

122. Khan, *supra* note 5, Appendix E, Figure 30 for the mean and median breakdowns of delivery time in 1980.

123. Municipal also had the same median delivery time (18 days). Technically, IP (117 days), insolvency (52 days), and constitutional (38.5 days) decisions had slower median delivery times. But these medians are based on two, seven, and eight observations, respectively. So these results were disregarded.

124. Technically, IP (117 days), insolvency (52 days), and constitutional (44.9 days) decisions had slower mean delivery times. But these means are based on two, seven, and eight observations,

Criminal decisions had the fastest mean delivery time (5.5 days).¹²⁵ The random forest model with judges included for 1980 indicates that “oral judgment” is the only significant predictor of delivery time. Considering how fast oral decisions were released in 1980 explains this significance. The random forest model with judges excluded changes that prediction: “court days,” “oral judgment,” and “word count” are all significant predictors of delivery time.¹²⁶ This change demonstrates the earlier discussed power of the “judge” feature and its clear relationship to other features, especially in the delivery time models. Nonetheless, the “judge” feature still was not a significant predictor of delivery time.

2000:¹²⁷ Tort decisions had the slowest median delivery time (60 days).¹²⁸ Criminal decisions had the quickest median delivery time (four days). Tort decisions had the slowest mean delivery time (84 days).¹²⁹ Criminal decisions had the fastest mean delivery time (17 days). Both random forest models indicate that “court days,” “word count,” and “oral judgment” are all significant predictors of delivery time.¹³⁰

2018:¹³¹ Tort decisions again had the slowest median delivery time (151 days).¹³² Equitable decisions had the fastest median delivery time (6.5 days).¹³³ Tort decisions had the slowest mean delivery time (151.7 days).¹³⁴ Equitable decisions had the fastest mean delivery time (18.7 days).¹³⁵ The random forest model for 2018 with judges included indicates that “oral judgment,” “word count,” and “judge” are all significant predictors of delivery time. Notably, this 2018 model was the only model where the

respectively. So these results were disregarded.

125. Technically, tax decisions (0 days) had the fastest mean delivery time, but that mean is based on one observation. So this result was disregarded.

126. Khan, *supra* note 5, Appendix E, Figures 31-32.

127. *Ibid* (see Appendix E, Figure 33 for the mean and median breakdowns of delivery time in 2000).

128. Technically, Aboriginal (207 days), constitutional (112 days), and tax (110 days) decisions had slower median delivery times, but these medians are based on two, six, and one observation(s), respectively. So these results were disregarded.

129. Technically, Aboriginal (207 days), constitutional (125 days), and tax (110 days) decisions had slower mean delivery times, but these means are based on two, six, and one observation(s), respectively. So these results were disregarded.

130. Khan, *supra* note 5, Appendix E, Figures 34-35.

131. *Ibid* (see Appendix E, Figure 36 for the mean and median breakdowns of delivery time in 1980).

132. Technically, Aboriginal (508 days) and IP (220 days) decisions had the slowest median delivery time, but those medians are each based on one observation. So these results were disregarded.

133. Technically, OPCA decisions (0 days) had the fastest median delivery time, but that median is based on one observation. So this result was disregarded.

134. Technically, Aboriginal (508 days) and IP (220 days) had the slowest mean delivery time, but those means are each based on one observation. So these results were disregarded.

135. Technically, OPCA decisions (0 days) had the fastest mean delivery time, but that result is based on one observation. So these results were disregarded.

judge became significant in the year-by-year random forest models. The random forest model with judges excluded maintains the “oral judgment” and “word count” features as significant predictors, and “court days” now also become a significant predictor (the same change happened in 1980 but not 2000).¹³⁶ This change is intriguing: it suggests some relationship between the “judge” and “court days” features. I currently have no intuition for this relationship. It could be because some judges in 2018 sat more court days than their colleagues. Notably, the “court days” and “word count” predictions have remained largely constant since the 1980 sample. Accordingly, the prediction appears robust. It also again demonstrates that Hypothesis 2 was partially valid (using this dataset and model).¹³⁷

Summary and reform suggestion for testing: Considering tort decisions’ length, their dominance as the slowest decision in 2000 and 2018 is unsurprising. The same points about tort decisions’ length apply equally to their slow delivery time, including the random forest conclusions. That criminal decisions are typically one of the fastest decisions is an important consideration for reducing delay in criminal proceedings.¹³⁸ This fact suggests that delay in the criminal process is likely due to other factors, e.g. the trial process itself and administrative delay. The relationship of “court days,” “word count,” and “oral judgment” to decisions’ length is a dominant consideration in improving decisions’ delivery time.¹³⁹ Reducing the need for “mega” and longer trials and increasing the use of oral decisions could improve delivery time. Reducing decisions’ length—as well as testing and addressing the underlying reasons for greater length that we just discussed—is another. But these intuitions must be rigorously tested.

136. See Khan, *supra* note 5, Appendix E, Figures 37-38.

137. Similar to the random forest models of the entire data set, only one of the three hypothesized features for longer decisions—“court days”—consistently predicted delivery time. The “judge” and “subject” features did not.

138. While not noted in the analysis for 2018, criminal decisions were also some of the fastest issued decisions (the median was seven days, and the mean was 28 days).

139. Similar to the random forest models of the entire dataset, the individual year random forest models for 1980, 2000, and 2018 did not demonstrate a bi-directional relationship between “word count” and “days to issue”.

3. *Judges' workload: a clear subjectivity exists in judges' outputs—some judges take longer and are more expressive than their colleagues (either in their words or what they decide to publish or both)*

1980: The dataset captured 70 judges issuing decisions in 1980. The median output of decisions per judge was 12.5 decisions,¹⁴⁰ and the median court days per judge was 23.5 days.¹⁴¹ The number of days judges sat in court varied broadly (from 121 days to one day):

- The judge with the most sitting days (121 days) was also the fifth most prolific decision writer (issued 46 decisions); was above the median word count (1,888 words) and the mean word count (2,228 words); and above the median delivery time (21 days) and well above the mean delivery time (52 days).
- The judge with the second most sitting days (110 days) was the eighth most prolific writer (issued 39 decisions); was above the median word count (1,800 words) and mean word count (2,219 words); and was just above the median delivery time (11 days) and mean delivery time (27 days).
- The judge with the most decisions (50 decisions) sat for 96 days.

2000: The dataset captured 115 judges issuing decisions in 2000. The median output of decisions per judge was 16 decisions,¹⁴² and the median court days per judge was 43 days.¹⁴³ The number of days judges sat in court again varied broadly (from 175 days to one day):

- The judge with the most sitting days (175 days) was the fifth most prolific decision writer (issued 44 decisions); was just above the median word count (2,456 words) but below the mean word count (3,012 words); and below the median delivery time (11 days) and well below the mean delivery time (15 days).
- The judge with the second most sitting days (153 days) was not a prolific decision writer (issued 15 decisions); was above the median word count (3,580 words) and the mean word count (4,811 words); and was well below the median delivery time (zero days) and mean delivery time (15 days).
- The judge with the most decisions (54 decisions) sat for 89 days.

140. The mean was 16.9 decisions.

141. The mean was 36.8 days.

142. The mean was 16.77 decisions.

143. The mean was 47 days.

2018: The dataset captures 127 judges issuing decisions in 2018. The median output of decisions per judge was 15 decisions,¹⁴⁴ and the median court days per judge was 50 days.¹⁴⁵ The number of days judges sat in court again varied broadly (from 224 days to one day). The 224 days involved multiple observations of the same matter. The better range is 147 days to one:

- The judge with the third most¹⁴⁶ sitting days (147 days) was not a prolific writer (issued 19 decisions); was below the median word count (4,155 words) but above the mean word count (9,382 words); and below the median days to issue (50 days) but above the mean days to issue (122 days).
- The next most sitting judge (146 days) was not a prolific writer (18 decisions); was above the median word count (7,414 words) and the mean word count (9,382 words); and was above the median days to issue (30 days) and the mean days to issue (73.9 days).
- The judge with the most decisions (40 decisions) sat for 85 days.

Summary and reform suggestion for testing: Overall, the 1980, 2000, and 2018 samples demonstrate a clear subjectivity in judges' outputs: some judges took longer and were more expressive than their colleagues. On median, judges' outputs in 2018 are like their 1980 colleagues (15 versus 12.5 decisions). But, on mean, judges in 1980 issued more decisions than their 2018 colleagues (16.9 versus 14.77 decisions). Importantly, while judges in 1980 issued more decisions in this dataset than their 2018 colleagues, those decisions were much shorter and faster on median and mean. The number of decisions they outputted does not mean they were more productive or efficient than their 2018 colleagues. On median and mean, judges in 2018 sat in court more days than their 1980 colleagues (50 versus 23.5 days and 53.88 versus 36.8 days, respectively). This jump in sitting time is consistent with the dominance of court days as an important predictive feature. These numbers—as well as other conclusions—would change if the dataset and model included the court days for judges sitting in jury trials and chambers (currently, this bulk data is not publicly posted for BCSC judges).¹⁴⁷ Variance and subjectivity in workload and output likely occur for four reasons: some judges are likely assigned more cases than others based on their past performance (as well as their expertise);

144. The mean was 14.77 decisions.

145. The mean was 53.88 days.

146. The judge at 224 days and 149 days mainly sat on the same matter.

147. Presumably, one could download the chambers list every day, scrape those data, and create a robust list for just how many days judges sit in court.

some judges likely want to write more or less than their colleagues and make more or less of an impact on Canadian law; some judges likely want to work less or more than their colleagues; and finally, some judges likely follow the trends of longer decisions more than some of their colleagues. Many judges are frankly also likely overworked: “unless most judges routinely worked nights, many weekends and part of their vacations, our court system would collapse.”¹⁴⁸ Governments and courts should work to address this issue either through better scheduling (or updating scheduling conventions);¹⁴⁹ increasing the number of judges, clerks, legal officers, support staff, data collection for analysis; and exploring the use of more technological solutions.¹⁵⁰ Again, the use of these solutions must be rigorously tested.

IV. *Limitations (this research has limited application) and future research suggestions (we need more analysis of trial-level judicial decisions and decision-making)*

These results are limited to the dataset I created and my coding protocol. I have not checked them against other reported databases,¹⁵¹ historical court data (if it exists), or unreported decisions.¹⁵² Many other features likely impact decision length and delivery time—such as how many expert opinions are in a particular case or the amount of evidence that is tendered in 2021 versus 1980. Gathering such data would allow for more granular and powerful conclusions. Similar research with tall and wide data should be conducted on all trial and appeal decisions in Canadian jurisdictions, e.g. comparing length, delivery time, structure, etc.¹⁵³

148. See e.g. Western Larch, *supra* note 1 at para 275. No current data on this subject exists. From conversations with judges, it remains a very live concern. See also Kathy Mack & Sharyn Roach Anleau, “Time for Judgment Writing: Final Report” (Judicial Research Project, Flinders University, 17 May 2010) [unpublished], online (pdf): *Australian Judicial Officers Association* <www.ajoa.asn.au/wp-content/uploads/2013/11/P41_02_08-Final-report-14-May-2010.pdf> [perma.cc/RG64-9LKY] (Australian data on this subject evidence that most judges and masters were overworked and lacked adequate time).

149. Superior Court of Justice, *supra* note 97.

150. As noted earlier, I mocked up a reform plan for judicial decisions and decision-making employing principles of human-centred design during my LLM research. See Khan, *supra* note 5, Part IV.

151. I did check Quicklaw’s reported appeal rate in 1980 using Westlaw. It was largely consistent.

152. The dataset will be missing dozens to hundreds of unreported decisions. Accordingly, these conclusions are incomplete, and my descriptive or predictive conclusions cannot fully account for the missing data.

153. See Greene et al, *supra* note 16 at 138, 168. Greene et al conducted a cross-Canada-wide survey of all appeal courts. Their data showed a wide divergence of average page lengths from 10.4 pages in British Columbia to 3.1 pages in New Brunswick and a wide divergence of delivery times with New Brunswick, PEI, and Newfoundland only having 11 decisions over 30 days versus Quebec having over 50 decisions and Nova Scotia having 35 decisions.

In the late 1990s, Greene et al noted that Canadian appeal decisions' average length hardly changed in the 20th century.¹⁵⁴ I anticipate the data would now show such changes, e.g. the SCC's average page length spiked from below 30 pages to above 40 pages in the early 2000s,¹⁵⁵ and the SCC's average page length has only increased since then.¹⁵⁶ This dataset, along with a broader dataset from other Canadian provinces, should also be used to explore predicting future trends and to test the predictors the random forest models identified as significant in other jurisdictions. The study of alleviating judicial delay in issuing decisions is occurring elsewhere,¹⁵⁷ as is the study of reducing length.¹⁵⁸ Yet these topics are hardly discussed in Canada, at least not publicly. Reducing decisions' length and delivery time need far more Canadian and comparative study. In sum, this research is the beginning. It must be replicated and subjected to far more analysis before proceeding or relying on it. Further data, research, and analysis are fundamental to help Canadian judges write and issue better judicial decisions.

Conclusion

In 1980, former Chief Justice Wilson of the BCSC opined that "some judgments are too brief, *many more are too long*."¹⁵⁹ This article questions what he would say now. BCSC judges are increasingly taking longer to issue decisions and issuing longer decisions. The same is likely true for other provinces and courts. Increased delivery times could be related to many factors. A key one that should be explored is the relationship of decision length to delivery time that the random forest models identified as significant. Shorter decisions likely lead to faster delivery times. The models also suggested a strong relationship between longer decisions,

154. *Ibid* at 132.

155. Donald R Songer, *The Transformation of the Supreme Court of Canada: An Empirical Examination* (Toronto: University of Toronto Press, 2008) at 153.

156. Rosevear & McDougall, *supra* note 17 at 4.

157. See Petkun, *supra* note 12 (Petkun discusses the merits of social incentivizes for increasing judicial speed, including the statutory six-month list in the US that discloses the number and name of motions that have been pending for longer than six months); see Figueiredo, Lahav & Siegelman, *supra* note 12 (Figueiredo, Lahav & Siegelman argue that judges may be making more errors to comply with the US statutory six-month list and that incentivizing judges to process cases faster is likely a mistake); Alessandro Melcarne, Giovanni B Ramello & Paige Marta Skiba, "The Role of Adjudication Procedures in Judicial Timeliness: Quasi-Experimental" (Vanderbilt University Law School Working Paper, 25 November 2020) [unpublished], DOI: <10.2139/ssrn.3195922> (Melcarne, Ramello & Skiba argue that Italian judges issue decisions faster when they do not provide legal justification, e.g. reasons).

158. See Meg Penrose, "To Say What the Law Is Succinctly" (2017) 50 Loy LA L Rev 101 (Penrose proposes a voluntary word count for majority, concurring, and dissenting US Supreme Court opinions).

159. The Honourable JO Wilson, *A Book for Judges* (Ottawa: Minister of Supply and Services Canada: 1980) at 79 [emphasis added].

slower delivery times, and more court days. Further research should be done on why longer trials require longer decisions or why parties require such long trials.

History suggests that long trials may not be truly necessary for complex issues; that long trials might not be necessarily more complex; and that long trials might not require complex, lengthy decisions. A prime example from this article's dataset is the seminal trial level decision, *Auton v AGBC*. It addressed complex equality section 15 *Charter* issues. Yet the trial judge issued a faster and shorter decision in 2000 than most tort decisions in 2018:

- The initial trial decision took 89 days to issue and was 13,111 words.¹⁶⁰
- The trial decision dealing with remedy took 84 days to issue and was 6,642 words.¹⁶¹
- The appeal decision took 231 days to issue and was 19,301 words (with a dissent) and 14,677 words (without the dissent).¹⁶²
- The SCC decision took 164 days to issue and was only 6,636 words.¹⁶³

An even more striking historical example is the “Persons Case” of 1929, *Edwards v Canada (Attorney General)*.¹⁶⁴ It is one of the most impactful constitutional cases in Canada's common law. Yet, it too, is quite short with only 6,365 words.

As the individual random forests for 1980, 2000, and 2018 demonstrated, aside from the “word count,” “oral judgment,” and “court days” features, the other features have not remained constant as significant predictors over a 40-year period. Accordingly, future prediction studies should focus on these features as well as other possible explanatory features—including more granular data on judges, lawyers, and other potential predictor features. That individual judges have become one of delivery time's significant predictors in 2018 is concerning. One explanation is that the judicial process is becoming more subjective over time, including the impact of judges on delivery time.

My intuition is that many judges would answer Justice Brown's question by saying that “if I had more time, I would have written a shorter and faster decision.” They may simply say that they have the time or

160. *Auton et al v AGBC*, 2000 BCSC 1142.

161. *Auton v AGBC*, 2001 BCSC 220.

162. *Auton v British Columbia (Attorney General)*, 2002 BCCA 538.

163. *Auton (Guardian ad litem of) v British Columbia (Attorney General)*, 2004 SCC 78.

164. *Edwards v Canada (Attorney General)* (1929), [1930] 1 DLR 98, [1929] 3 WWR 479 (UKPC).

confidence to write shorter decisions that would lead to quicker delivery times and more digestible decisions. But his call to action and question must not be treated as more than a hypothesis (indeed, he might be wrong). And it also must not be ignored. Instead of relying on it as the final answer or overlooking his plea for help, governments, courts, and academics should finally answer his question and call for help. In Justice Brown's words, "No sacred cows, judicial or otherwise, should stand immune from scrutiny.... Everything must be on the table for examination and reform."¹⁶⁵

Appendix A: Design of the Random Forest Models & Explanation of Metrics

Models were first run to evaluate possible significant predictors and for discussion and evaluation purposes. The models deemed for further exploration were then trained on a grid of hyperparameters using 1-fold cross validation to best tune the following inputs:

- ***mtry***: the number of features randomly evaluated in the determination of each split (values of 1 to the total number of features were evaluated); and
- ***min.node.size***: a cut-off ensuring that a node is only split further if it meets this minimum number of observations (a sequence from 5 to 100 by 5 were evaluated).

The models had these specifications:

- tuning models were run with 100 trees;
- all models employed the "variance" splitrule that chooses splits based on what best reduces variance in the dependent feature;
- all models were coded in the R language (code available on request);
- all models availed the ranger package;
- all final models were run with 1000 trees; the models for tuning hyperparameters were run with 100 trees (availed the caret package);
- *mtry* and *min.node.size* were tuned to the best performance using 10-fold cross validation on a grid search of those parameters; and
- 50 permutations (forests) were used in the Altmann method for computing feature importance with p-values, where the null hypothesis is that there is no association of the independent features on the dependent feature.

165. *Western Larch*, *supra* note 1 at para 277.

The models relied on four explanation metrics:

1. **RF RMSE (root mean squared error):** For the purpose of the word count and delivery time models, the RF RMSE provides the amount the prediction would worsen if you ran the model with this feature permuted (i.e. scrambled)—e.g. scramble the words feature in the entire random forest model for delivery time, and the prediction would be off by 36 days, etc.¹⁶⁶ Similar to RF-Squared, no absolute criterion exists for a good RMSE value.¹⁶⁷
2. **RF Prediction Error:** The RF Prediction Error is the RMSE of the full model (with all predictors). This calculation is the amount of error, on average, one can expect the random forest to have on the predictions it makes. This value is calculated from the predictions on a holdout (out-of-bag) sample of the data.
3. **RF R-Squared:** It indicates how well the model explains the variability in the dependent feature (e.g. word count or delivery time). For example, if the number is .50, then approximately half the model's features can explain the observed variation. Lower R-squared does not, however, indicate poor prediction power.¹⁶⁸

Altmann p-value: P-values provide one way of measuring uncertainty and error probability.¹⁶⁹ Part II's random forest model availed the Altmann approach to estimating p-values of random forest predictors: "non-informative predictors do not receive significant p-value."¹⁷⁰ Part II employs a cut-off p-value of <0.05. Anything higher (e.g. 0.1) is not treated as a significant predictor of word count or delivery time.

166. See Matthew Baumer, "randomForest" (16 December 2015), online: *RPubs* <rpubs.com/mbaumer/randomForest> [perma.cc/8Q2A-7G56].

167. See Robert Nau, "What's the bottom line? How to compare models" (last visited 14 July 2022), online: *Duke University's Fuqua School of Business* <people.duke.edu/~rnau/compare.htm> [perma.cc/2VSE-7EPV]. See also Hapfelmeier & Ulm, *supra* note 39 at 51; Carolin Strobl, James Malley & Gerhard Tutz, (2009) 14:4 *Psychological Methods* 323, DOI: <10.1037/a0016973> ("[a]n introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests" at 343. Larger values demonstrate predictive power between the predictor variable and the dependent variable).

168. See Robert Nau, "What's a good value for r-squared" (last visited 14 July 2022), online: *Duke University's Fuqua School of Business* <people.duke.edu/~rnau/rsquared.htm> [perma.cc/B9Y3-DGEJ]. See also Clay Ford, "Is R-Squared Useless?" (17 October 2015), online: *University of Virginia Library* <data.library.virginia.edu/is-r-squared-useless/> [perma.cc/9VLL-F4GB].

169. Alan S Rigby, "Getting Past the Statistical Referee: Moving Away from P-Values and Towards Interval Estimation" (1999) 14:6 *Health Education Research* 713 at 713, DOI: <10.1093/her/14.6.713>.

170. Andre Altmann et al, "Permutation Importance: A Corrected Feature Importance Measure" (2010) 26:10 *Bioinformatics* 1340 at 1340, DOI: <10.1093/bioinformatics/btq134> ("[t]he PIMP P-values are easier to interpret and provide a common measure that can be used to compare feature relevance among different models" at 1347).

