Studying Nonobviousness

JASON RANTANEN, LINDSAY KRIZ & ABIGAIL A. MATTHEWS[†]

Many scholars have observed that an empirical study is only valid to the extent it is reliable. Yet assessments of the reliability of empirical legal studies are rare. The closest most scholarship comes is to compare the results of their studies to those of others. As a result, in many legal fields, including intellectual property law, scholars lack a grounded understanding of how valid or reliable empirical legal studies really are.

This Article examines the reliability of empirical studies of judicial decisions by closely comparing two recent studies of the patent law doctrine of nonobviousness. We find these studies provide robust results despite differences in the cases selected to include in each dataset. However, the amount of agreement varied for some data fields more than others. Particularly, there was more inter-study variability for fields that examined judicial reasoning than fields for decision outcomes. This finding provides some validation for the use of macro-level studies of judicial decision-making. To the best of our knowledge, this is the first analysis to directly compare the actual coding (as opposed to just the outcomes) of two different studies examining the same patent law doctrine.

Building on the existing data, we also make an original contribution to the literature on nonobviousness by extending the time studied to the present. In contrast with studies examining the immediate period after the Supreme Court's decision in KSR v. Teleflex, we find (1) a substantial decline in the number of 35 U.S.C. § 103 district court cases appealed to the Federal Circuit, (2) a higher rate at which courts deem the patent nonobvious, and (3) a high affirmance rate for district court determinations of both "obvious" and "nonobvious."

TABLE OF CONTENTS

INTRODUCTION

A study is only as valuable as the data, methodology, and analysis that it contains. While scholars and policymakers presume academic research—particularly when it is based on empirical studies—is reliable, scholars have done relatively little work to assess that assumption for legal studies. This is especially significant given the lack of systematic peer assessment of most empirical legal studies due to the nature of how journals publish legal scholarship.[1] With the increasing popularity of empirical studies in intellectual property law,[2] questions about the reliability of these studies are paramount.

Reliability is the extent to which results are consistent and an accurate representation of the population.[3] One concept at the heart of reliability is replicability. Replicability is the ability of subsequent researchers to redo a study or experiment and achieve the same results.[4] And conversely, irreplicability shows an experiment does not work, and "tells us something about the world that is not true."[5] While other fields have experienced a "replicability crisis," as Janet Freilich explains in *The Replicability Crisis in Patent Law*, law has been slow to follow.[6] Indeed, published assessments of the replicability of empirical legal studies—and reliability, more generally—are more elusive than the empirical studies themselves.

The lack of published reliability assessments is as true in empirical intellectual property law studies as in other areas of the law. For example,

---

1. It is well known that most legal scholarship is published in student-edited journals with relatively little involvement by non-students. *See, e.g.*, Bernard J. Hibbitts, *Last Writes? Reassessing the Law Review in the Age of Cyberspace,* 71 N.Y.U. L. REV. 615, 630–31 (1996); Howard A. Denemark, *The Death of Law Reviews Has Been Predicted: What Might Be Lost When the Last Law Review Shuts Down?*, 27 SETON HALL L. REV. 1, 7 (1996). With the exception of the *Northwestern Law Review*'s annual empirical studies issue, and perhaps a few others, pre-publication assessment of empirical legal studies is not conducted by scholars with familiarity with empirical methods. A common response is that legal scholarship's true peer assessment occurs after publication as legal scholars, lawyers, judges, and policymakers read and engage with it. *See, e.g.*, Richard A. Wise, Lucy S. McGough, James W. Bowers, Douglas P. Peters, Joseph C. Miller, Heather K. Terrell, Brett Holfeld & Joe H. Neal, *Do Law Reviews Need Reform? A Survey of Law Professors, Student Editors, Attorneys, and Judges*, 59 LOY. L. REV. 1, 30 (2013)*.* While this may be a good point for argument-based legal scholarship, we think that it has less weight when it comes to empirical legal studies, which often require substantial time, effort, and expertise to assess their credibility and reliability.

2. There are so many studies in intellectual property law that a literature review of them comprises an entire Research Handbook volume. *See generally* RESEARCH HANDBOOK ON THE ECONOMICS OF INTELLECTUAL PROPERTY LAW: VOL. 2: ANALYTICAL METHODS (Peter S. Menell & David L. Schwartz eds., 2019).

3. Nahid Golafshani, *Understanding Reliability and Validity in Qualitative Research*, 8 QUALITATIVE REP. 597, 598 (2003).

4. Janet Freilich, *The Replicability Crisis in Patent Law*, 95 IND. L.J. 431, 438–40 (2020).

5. *Id.* at 440.

6. *Id.* at 433. There is an annual *Empirical Legal Studies Replication Conference*, but participation is limited due to the fact that so few legal scholars undertake replication studies. A list of the 2019 participants can be found at *Empirical Legal Studies Replication Conference, 2019*, SCIENCEDIRECT: INT'L REV. L. & ECON., https://www.sciencedirect.com/journal/international-review-of-law-and-economics/special-issue/ 10TQ91RK6PC (last visited Mar. 18, 2022). Another obstacle legal scholars face is the limited access to data, since it is not the norm for legal scholars to publish datasets and therefore very few do so. Legal scholars wishing to execute a replication study therefore can either request access to data used in a prior study or collect the data themselves using the limited number of available sources.

despite the many studies of nonobviousness—the "ultimate condition" for a patent[7]—there has been relatively little work done to assess the reliability of these studies. While articles typically discuss whether their results are consistent or inconsistent with past studies,[8] and there has been limited inter-study analysis focused on comparing outcomes,[9] no prior nonobviousness scholarship has systematically compared two studies for the same time period and the same selection criteria to assess study replicability. This is problematic because studies rarely overlap in the time they encompass, rendering comparisons of limited value for assessing reliability.[10] Nor has anyone directly compared the coding for these studies to determine the reliability of particular types of data and its variables. More broadly, to the best of our knowledge, there are no data-level comparisons for any of the studies of intellectual property law doctrines.[11]

To gain insight into the reliability question, this article compares two different research groups' datasets of patent law's nonobviousness requirement: (1) Ryan Holte and Ted Sichelman's *Cycles of Obviousness* ("Holte & Sichelman (2019)")[12] and (2) Jason Rantanen's *The Federal Circuit's New Obviousness Jurisprudence* ("Rantanen (2013)").[13] We selected these two studies because they are the most comprehensive studies of the Federal Circuit's

---

7. *See generally* Nonobviousness—The Ultimate Condition of Patentability (John F. Witherspoon ed., 1980) (collecting Judge Rich's essays on the topic).

8. *See, e.g.*, Jason Rantanen, *The Federal Circuit's New Obviousness Jurisprudence: An Empirical Study*, 16 Stan. Tech. L. Rev. 709 (2013) [hereinafter Rantanen, *Obviousness Jurisprudence*]; Ryan T. Holte & Ted Sichelman, *Cycles of Obviousness*, 105 Iowa L. Rev. 107 (2019); Glynn S. Lunney, Jr. & Christian T. Johnson, *Not So Obvious after All: Patent Law's Nonobviousness Requirement, KSR, and the Fear of Hindsight Bias*, 47 Ga. L. Rev. 41 (2012); Brendan Seth O'Brien O'Shea, Note, *What Is Obvious: Empirical Assessment of KSR's Impact*, 45 AIPLA Q.J. 517 (2017); Gregory N. Mandel, *A Nonobvious Comparison: Nonobviousness Decisions at the PTAB and in the Federal Courts*, 24 Tex. Intell. Prop. L.J. 403 (2016).

9. Jason Rantanen, *Empirical Analyses of Judicial Opinions: Methodology, Metrics, and the Federal Circuit*, 49 Conn. L. Rev. 227, 242–59 (2016) [hereinafter Rantanen, *Judicial Opinions*].

10. Accurate measurement includes three different but related concepts. Reliability examines the consistency, whether we get the same answer looking at different samples from the same population. Replicability also refers to consistency, but to the exact same sample. It is the notion that if a different set of individuals repeat a specific study, they should reach the same result. And validity is the extent to which we are measuring what we say we are measuring. In this article, we focus on reliability generally, and inter-rater reliability, specifically. Inter-rater reliability is the extent to which different individuals are consistent in their judgments. *See generally* Paul C. Price, Rajiv S. Jhangiani, I-Chant A. Chiang, Dana C. Leighton & Carrie Cuttler, Research Methods in Psychology (3d ed., 2017); Anol Bhattacherjee, Social Science Research: Principles, Methods, and Practices (2d ed., 2012).

11. In contrast, there are a few data-level reliability assessments in patent law outside of studies of judicial opinions. *See, e.g.*, Lisa Larrimore Ouellette & Andrew Tutt, *How Do Patent Incentives Affect University Researchers?*, 61 Int'l Rev. L. & Econ. art. 105883 (2020); David L. Schwartz & Jay P. Kesan, *Analyzing the Role of Non-Practicing Entities in the Patent System*, 99 Cornell L. Rev. 425 (2014); Germán Poo-Caamaño & Daniel M. German, *Software Patents: A Replication Study*, Proc. of the 11th Int'l Symp. on Open Collaboration (2015).

12. Holte & Sichelman, *supra* note 8.

13. Rantanen, *Obviousness Jurisprudence*, *supra* note 8. Note that one of the authors of this paper is the author of Rantanen, *Obviousness Jurisprudence*.

jurisprudence before and after the Supreme Court's opinion in *KSR v. Teleflex*,[14] collectively encompassing the period from 1997 to 2013. Indeed, both confidently assert they collected the entire population of district courts' nonobviousness decisions.[15] While there are other excellent studies of nonobviousness, those either cover relatively short time periods or have only limited overlap with other studies.[16] Both Holte & Sichelman (2019) and Rantanen (2013) examined changes in Federal Circuit decisions involving nonobviousness before and after *KSR*, the first Supreme Court opinion addressing obviousness in the Federal Circuit era.[17] The central goal of these studies was to test whether *KSR* produced a change in this core element of patent law.[18]

We draw on two approaches rooted in the concept of replicability to assess the reliability of this type of study. First, we compare the results of the two studies while controlling for major methodological differences by limiting the comparison to decisions from the same time period that contained an actual determination of obviousness based on the studies' methodologies. In addition, we examine the record-level agreement between the studies' coding of individual decisions. This approach draws on an element of high-quality studies involving human coding: inter-rater agreement assessments, which are used to show the degree to which individual coders agree on an outcome.[19] Rather than simply apply this tool within our own study, however, we instead leverage it to measure the amount of agreement between the final datasets for these two studies.

Our reliability assessment reveals a complex picture. Somewhat reassuringly, the results from the two studies are largely consistent for decisions from the same time period. Surprisingly, however, fewer than two-thirds of the decisions analyzed in both studies were the same—even when limited to the

---

14. The Court held that rigid and formal application of the teaching-suggestion-motivation test ("TSM" test) is contrary to existing patent law precedent, and courts should instead take a wholistic and functional approach to the obviousness inquiry by considering a variety of factors—including TSM—that are indicative of obviousness. KSR Int'l Co. v. Teleflex Inc., 550 U.S. 398, 418–19 (2007).

15. *See* Holte & Sichelman, *supra* note 8, at 136 ("[W]e are fairly confident that we identified and collected all district court and Federal Circuit obviousness decisions—including so-called Rule 36 summary affirmances with no opinion—issued between January 1, 2003 and December 31, 2013."); Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 730 ("As a starting point, the coded cases represent essentially the entire universe of opinions during the time period.").

16. *See* studies cited *infra* note 26.

17. *See generally KSR Int'l Co.,* 550 U.S. 398. Other studies examining the before/after of *KSR* include Lunney & Johnson, *supra* note 8; Jennifer Nock & Sreekar Gadde, *Raising the Bar for Nonobviousness: An Empirical Study of Federal Circuit Case Law Following* KSR, 20 FED. CIR. BAR J. 369 (2010); and Ali Mojibi, *An Empirical Study of the Effect of* KSR v. Teleflex *on the Federal Circuit's Patent Validity Jurisprudence,* 20 ALB. L.J. SCI. & TECH. 101 (2010).

18. We note that Holte & Sichelman (2019) indicate that their data collection was part of a larger project to examine obviousness-related changes and the 2019 article reports on only a portion of their dataset. *See* Holte & Sichelman, *supra* note 8, at 138.

19. *See* Mark A. Hall & Ronald F. Wright, *Systematic Content Analysis of Judicial Opinions*, 96 CALIF. L. REV. 63, 112–16 (2008) (describing the use of inter-rater agreement measures in empirical legal studies).

identical time period and using the same criteria. Within that set of cases, however, the core data coding was generally identical, with a few notable exceptions. Specifically, we find differences in the coding for procedural postures and in some coding related to judicial reasoning.

This comparison has significant implications for understanding the reliability of patent law studies specifically and doctrinal assessments of judicial decisions more generally. To the best of our knowledge, this is the first analysis to directly compare the coding from two separate research groups studying judicial decisions involving patent law doctrines that used the same criteria to record data about the cases. Our findings suggest that these types of studies are reasonably replicable and can have robust results—although contextualized interpretation of their results remains critical.[20] Along those lines, we suggest that while the results reported in articles can be important, scholars should not overlook details on methodology, coding instruction, and the replicability of the data.

<div align="center">***</div>

Because they focused specifically on the effects of *KSR*, the datasets for Rantanen (2013) and Holte & Sichelman (2019) stop a few years after that decision. In Part II, we[21] extend the empirical data on obviousness to December 2019. Using this more recent data, we find that:

- The inter-rater reliability analysis indicated at least "substantial agreement" for all recorded variables.

- The number of Federal Circuit decisions in appeals arising from the district courts that involved a 35 U.S.C. § 103 issue peaked between 2010 and 2015, and in recent years has declined to half of that peak.

- The percentage of Federal Circuit decisions with an outcome of "obvious" in appeals arising from the district courts remained high between 2006 and 2014, but since 2015 the number of "obvious" outcomes has fallen dramatically.

- The Federal Circuit continues to affirm district courts on the issue of obviousness at a high rate (around 80% of the time since 2013), and—in contrast with the immediate post-*KSR* period studied by

---

20. In particular, as has been discussed in the literature, even when care is taken to identify the entire population of decisions, selective forces resulting in population biases can affect the composition of appellate decisions. *See, e.g.*, Rantanen, *Judicial Opinions*, *supra* note 9, at 243–44 (discussing population biases in appellate decisions).

21. Rantanen, Kriz, and Matthews.

Holte & Sichelman (2019) and Rantanen (2013)—since 2013, affirmances have been equally high for district court outcomes of "nonobvious."

- The number of grants of summary judgment involving § 103 that were reviewed by the Federal Circuit in its decisions has fallen substantially in recent years. Nearly all § 103 decisions arising from the district court that were reviewed by the Federal Circuit between 2016 and 2019 involved a bench or jury trial.

The results from the data we collected for this new period are striking and stand in sharp contrast with the period immediately after *KSR*. They show a decline in appeals from district court decisions involving § 103 and a Federal Circuit that is more deferential to the determinations that are appealed. They also suggest that there has been a shift in the composition of the nonobviousness issues arriving at the Federal Circuit or a Federal Circuit that is normatively more skeptical of obviousness challenges to issued patents. There are reasons to think there is some truth to both.

Finally, in order to maximize data transparency, the data that we used, the codebook, and the Stata code for constructing and analyzing the data are being published contemporaneously with this article.[22]

I. RELIABILITY

A. BACKGROUND

Patent law's nonobviousness requirement has long fascinated scholars, judges, and attorneys alike. This requirement, codified at 35 U.S.C. § 103, prohibits patents on claimed inventions "if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious" as of the patent application's effective filing date.[23] Many people have offered doctrinal and theoretical perspectives on the requirement, from Judge Giles Rich's commentaries[24] to Dmitry Karshtedt's recent *Nonobviousness—Before and After*.[25] Scholars have closely examined the nonobviousness requirement using quantitative techniques, with studies

---

22. Jason Rantanen, Lindsay Kriz & Abigail Matthews, *Replication Data for "Studying Nonobviousness,"* HARVARD DATAVERSE, V1 (2022), https://doi.org/10.7910/DVN/A1UTYC.

23. *See* 35 U.S.C. § 103 (2018). While § 103 was amended in 2011, those amendments were relatively minor. *See* JOHN F. DUFFY & ROBERT MERGES, PATENT LAW AND POLICY: CASES AND MATERIALS 511 (2017).

24. *See* NONOBVIOUSNESS, *supra* note 7 (collecting Judge Rich's essays on the topic); *see also* Michael Abramowicz & John F. Duffy, *The Inducement Standard of Patentability*, 120 YALE L.J. 1590, 1605 (2011); Rebecca S. Eisenberg, *Obvious to Whom? Evaluating Inventions from the Perspective of PHOSITA*, 19 BERKELEY TECH. L.J. 885, 891–92 (2004).

25. *See* Dmitry Karshtedt, *Nonobviousness: Before and After*, 106 IOWA L. REV. 1609 (2021).

focusing on frequency, outcomes, and doctrine,[26] or using it as a proxy to assess judicial decisionmaking more broadly.[27] These studies illustrate both the interest in empirical examinations of the doctrine and the need to understand how reliable this data is.[28] After all, anyone can create some numbers; the real question is whether they are worth anything.

While there are many empirical studies of nonobviousness, we focus on *The Federal Circuit's New Obviousness Jurisprudence* ("Rantanen (2013)")[29] and *Cycles of Obviousness* ("Holte & Sichelman (2019)")[30] for the reasons discussed in the Introduction. Both studies examined written opinions and Rule 36 summary affirmances involving nonobviousness during the first decade of the twenty-first century.[31]

Rantanen (2013) studied whether the Federal Circuit changed its approach to nonobviousness after *KSR* with respect to both the analysis of the issue and outcome of the case.[32] The study period ranged from ten years before the grant of certiorari in *KSR* (June 26, 2006) to five years after the Court's decision (April 30, 2007).[33] It found that following *KSR*, the Federal Circuit reached a conclusion that patents[34] were obvious more often than it did before *KSR* in

---

26. *See, e.g.,* Glynn S. Lunney, Jr., *E-Obviousness*, 7 MICH. TELECOMM. & TECH. L. REV. 363 (2001) [Lunney, *E-Obviousness*]; Christopher A. Cotropia, *Nonobviousness and the Federal Circuit: An Empirical Analysis of Recent Case Law,* 82 NOTRE DAME L. REV. 911 (2007); Lee Petherbridge & R. Polk Wagner, *The Federal Circuit and Patentability: An Empirical Assessment of the Law of Obviousness*, 85 TEX. L. REV. 2051 (2007); O'Brien O'Shea, *supra* note 8; Calvin M. Brien, *An Empirical Analysis of Patent Validity in Inter Partes Reviews through the Lens of* KSR, 46 AIPLA Q.J. 413 (2018); Nock & Gadde, *supra* note 17; Michelle Friedman Murray, *Nonobviousness Standards for Hardware and Software Before and After* KSR*: What Is the Difference*, 93 J. PAT. & TRADEMARK OFF. SOC'Y 259 (2011); Mandel, *supra* note 8. For additional doctrinal studies of nonobviousness, see 2 RESEARCH HANDBOOK ON THE ECONOMICS OF INTELLECTUAL PROPERTY LAW, *supra* note 2, at 281–309.

27. *See generally* Banks Miller & Brett Curry, *Expertise, Experience, and Ideology on Specialized Courts: The Case of the Court of Appeals for the Federal Circuit*, 43 LAW & SOC'Y REV. 839 (2009); Banks Miller & Brett Curry, *Experts Judging Experts: The Role of Expertise in Reviewing Agency Decision Making*, 38 LAW & SOC. INQUIRY 55 (2013).

28. A few studies have also used experimental human techniques in human studies research to examine obviousness. Lunney & Johnson**,** *supra* note 8; Gregory N. Mandel, *Patently Non-Obvious: Empirical Demonstration that the Hindsight Bias Renders Patent Decisions Irrational*, 67 OHIO STATE L.J. 1391 (2006).

29. Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 732.

30. Holte & Sichelman, *supra* note 8.

31. Rule 36 summary affirmances allow the Federal Circuit to affirm the district court ruling without an opinion. Kimberly A. Moore, Markman *Eight Years Later: Is Claim Construction More Predictable?*, 9 LEWIS & CLARK L. REV. 231, 234 (2005). Thus, the Federal Circuit is able to efficiently rule on cases where the determination under review was correct and there are no new legal issues to explain or discuss. *Id.*

32. Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 732. In addition to examining appeals arising from the district courts and International Trade Commission, Rantanen (2013) also examined appeals arising from the USPTO. This data was reported separately from appeals arising from the district courts and ITC, and is not discussed in this article. Note that there are many limitations of these types of macro-level studies of judicial decisions, not the least of which are selection effects and the consequences of procedural postures. For an in-depth discussion of these issues, see *id*. and Rantanen, *Judicial Opinions*, *supra* note 9.

33. In other words, from June 26, 1996 to April 30, 2012.

34. To be more accurate, the issue is whether patent *claims* are nonobvious. For ease of reading, however, we use the colloquial "patents."

appeals arising from the district courts and International Trade Commission ("ITC").[35] Before the Supreme Court's grant of certiorari in *KSR*, the Federal Circuit held patents to be obvious 46% of the time, while after *KSR* it reached a conclusion that the patents were obvious 57% of the time.[36] Success for patent challengers was in part attributed to the fact that the Federal Circuit affirmed lower court findings of obviousness at a greater rate post-*KSR*.[37] Before *KSR*, the Federal Circuit affirmed 60% of district court determinations that the patents were obvious, while after *KSR* it affirmed 81% of those decisions.[38] One explanation offered by the study was that the Federal Circuit was being more deferential to district courts following *KSR*; another was that the Federal Circuit's jurisprudence itself reflected a raised bar for patents to be nonobvious.[39]

Recently, former professor (now judge) Ryan Holte and Professor Ted Sichelman undertook another study of nonobviousness decisions during the period around *KSR*. In contrast with the Rantanen (2013) study, which focused entirely on the Federal Circuit, Holte and Sichelman studied both appellate decisions and district court decisions, providing a deeper picture of § 103[40] decisions in patent infringement litigation.[41] Their study analyzed Federal Circuit and district court cases between January 1, 2003 and December 31, 2013,[42] and found that prior to *KSR*, the Federal Circuit held patents obvious 49% of the time, while after *KSR* the Federal Circuit found patents obvious 57% of the time.[43] As in Rantanen (2013), Holte & Sichelman (2019) examined affirmance rates, finding that prior to *KSR,* the Federal Circuit affirmed obviousness outcomes 81% of the time, while after *KSR* it affirmed them 90% of the time.[44] Nonobviousness dispositions reflected the opposite pattern, with the Federal Circuit affirming district court determinations of nonobviousness 85% of the time prior to *KSR* but 68% of the time following *KSR*.[45]

Both studies mostly used the same data coding instructions.[46] They typically contain coding instructions in "codebooks," which are designed to

---

35. Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 736. In hindsight, the data in Rantanen (2013) would have been more clearly reported if it had not included appeals from the ITC with appeals from the district courts—especially given that there were very few decisions arising from the ITC (only 6), so they do not really add anything to the study.

36. *Id.* at 737.

37. *Id.* at 741.

38. *Id.* at 744.

39. *Id.* at 746–47.

40. 35 U.S.C. § 103 (2018).

41. *See generally* Holte & Sichelman, *supra* note 8.

42. *Id.* at 145.

43. *Id.* at 142. The authors reported a "mixed" outcome rate of 4% prior to *KSR* and 7% after *KSR*. *Id.*

44. *Id.* at 146.

45. *Id.*

46. Indeed, portions of the Holte & Sichelman (2019) codebook were verbatim of the Rantanen (2013) codebook—which, to be clear, is a good thing for assessing replicability! *See* Rantanen, *Judicial Opinions*, *supra* note 9, at 281–82.

serve as a guide while collecting and reporting the data.[47] These guides "should be sufficiently rich so that [they] not only enable[] the researchers to code their data reliably but also allows others to replicate, reproduce, update, or build on the variables housed in the database . . ."[48] As a preliminary matter, Holte and Sichelman indicate they relied heavily on the techniques described in Rantanen (2013) to identify all obviousness decisions during their period of analysis.[49] Thus, both studies collected relevant opinions by first executing at least a broad Westlaw search, then manually refined the search results to eliminate any opinion that did not actually contain an obviousness determination.[50] Beyond that, comparing the codebooks[51] shows that nearly all of the coding instructions for Federal Circuit decisions mirror one another.[52]

With the addition of the district court data, Holte and Sichelman were also able to directly examine the outcomes at the district courts. They found that, whereas prior to *KSR* 69% of outcomes at the district courts were that the patents were nonobvious, after *KSR* 48% of outcomes were that the patents were nonobvious.[53] However, the appeals actually decided by the Federal Circuit indicated about an even split of district court findings of obvious and nonobvious.[54] From this Holte and Sichelman reached the conclusion that "a much higher absolute percentage of litigants began to appeal nonobviousness findings post-*KSR*."[55] This adds support to the idea that not only were there fewer nonobviousness outcomes at the district courts but those (fewer) nonobvious outcomes were being appealed more often—thus suggesting a shift in the underlying spectrum of cases being reviewed by the Federal Circuit.[56]

---

47. LEE EPSTEIN & ANDREW D. MARTIN, AN INTRODUCTION TO EMPIRICAL LEGAL RESEARCH 106, 112 (2014).

48. *Id.* at 106.

49. Holte & Sichelman, *supra* note 8, at 136 n.219.

50. Rantanen (2013) used a variety of searches on Westlaw to identify cases. Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 727. In contrast, Holte & Sichelman (2019) first identified cases by relying upon PatStats and Docket Navigator, then used a Westlaw search to supplement those sources. Holte & Sichelman, *supra* note 8, at 136 n.221. In addition, Holte & Sichelman (2019) drew upon their identification of district court decisions in locating appellate decisions. *Id.* The more expansive search techniques used by Holte & Sichelman (2019) may explain some of the differences in the resulting data sets discussed below.

51. *See* JASON RANTANEN, CODEBOOK FOR EMPIRICAL STUDY OF FEDERAL CIRCUIT OBVIOUSNESS JURISPRUDENCE (2013), https://empirical.law.uiowa.edu/sites/empirical.law.uiowa.edu/files/wysiwyg_uploads/ Obviousness%20Codebook%20Final%202013-07-05.pdf [hereinafter RANTANEN CODEBOOK]; TED SICHELMAN & RYAN HOLTE, CODEBOOKS FOR CYCLES OF OBVIOUSNESS (2020), https://papers.ssrn.com/sol3/ papers.cfm?abstract_id=3719135 [hereinafter HOLTE & SICHELMAN CODEBOOK].

52. *See* RANTANEN CODEBOOK, *supra* note 51; HOLTE & SICHELMAN CODEBOOK, *supra* note 51.

53. Holte & Sichelman, *supra* note 8, at 141 fig.1.

54. The Rantanen (2013) study referred to this as the "substrate" of Federal Circuit decisions and also reported about an even split of obvious and nonobvious decisions by the lower tribunal. *See* Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 740.

55. Holte & Sichelman, *supra* note 8, at 145.

56. *See* Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 746–47.

B. Reliability Assessment

This study assesses the reliability of empirical studies of judicial decisions by comparing the results and data of Rantanen (2013) and Holte & Sichelman (2019). To do so, we first compare the results of the two studies within identical time periods and with similar record selection criteria. Next, we draw on methodologies traditionally used to assess inter-rater agreement within studies and apply those techniques to assess the agreement between the coding of the individual records that are included in both studies.

One point we want to stress at the outset: this approach is only possible because data and methodological information was available for both studies.[57] The data and Codebook for Rantanen (2013) are publicly available on the internet.[58] The Holte & Sichelman (2019) Codebook is also available on the internet and the authors provided us with a copy of their data.[59] Without access to their data and codebook, our best alternative would be to compare the overall numbers from their article. This would have provided limited ability to conduct a reliability assessment, given that the time periods reported in the two studies overlap only in part. Lack of available data can pose a substantial barrier to conducting reliability assessments.[60] Fortunately, that was not the case here.

C. Comparison of Study Results

The temporal overlap, coding of identical fields, and use of essentially the same coding criteria for the Rantanen (2013) and Holte & Sichelman (2019) studies provide an opportunity for something that has conventionally been quite rare in the literature of empirical studies of patent law decisions: the ability to directly compare the results of two studies of the same subject during the same time period conducted by different research groups.[61] In addition, advancements in data collection and management capabilities, along with interdisciplinary expertise, allow for direct comparison at the individual record level. These comparisons provide a better understanding of the reliability of this type of research.

1. Methodology for Comparison of Results

A major barrier to assessing reliability by comparing results is that case-based studies frequently differ in their record selection criteria. This is true here.

---

57. *Cf.* Todd E. Pettys, *Free Expression, In-Group Bias, and the Court's Conservatives: A Critique of the Epstein-Parker-Segal Study*, 63 Buff. L. Rev. 1 (2015) (replicating a portion of the Epstein, Parker, and Segal study and identifying coding issues in the Epstein, Parker, and Segal methodologies).

58. The data and codebook are available at *Studies and Data*, Fed. Cir. Data Project, https://empirical.law.uiowa.edu/studies-and-data (last visited Mar. 18, 2022).

59. *See* Holte & Sichelman Codebook, *supra* note 51.

60. *See, e.g.*, Jason Rantanen, *The Future of Empirical Legal Studies: A Response to Holte & Sichelman's Cycles of Obviousness*, 105 Iowa L. Rev. Online 15 (2020) [hereinafter Rantanen, *Response to Holte & Sichelman*]; Schwartz & Kesan, *supra* note 11.

61. *See generally* Rantanen, *Judicial Opinions*, *supra* note 9.

With respect to differences in time period, while Rantanen (2013) contained decisions for about a six-year period before the starting point of the period in Holte & Sichelman (2019), Holte & Sichelman (2019) contained decisions for about a year and a half after Rantanen (2013). Rantanen (2013) also distinguished between Federal Circuit decisions before the date of certiorari and after the Court's decision, whereas Holte & Sichelman (2019) used the date of the Court's decision as the breakpoint.[62] In addition, Rantanen (2013) included decisions arising from the ITC and decisions in which the Federal Circuit did not reach a final conclusion on obviousness; Holte & Sichelman (2019) did not.[63] As a result, even though the results of both studies are consistent, and support the overall conclusion that the Federal Circuit's nonobviousness jurisprudence appears to have changed after *KSR*, the statistics reported in the two papers cannot just be pasted next to one another.

In order to directly compare the two studies, we began with their complete datasets of Federal Circuit decisions and created subsets that met the comparison criteria.[64] We limited both datasets to Federal Circuit decisions that arose from the district courts and that were decided between January 1, 2003 and April 30, 2012 (the overlapping time period). Because they were not addressed in Rantanen (2013) or Holte and Sichelman (2019), we excluded interference proceedings and appeals involving a preliminary injunction.[65] In addition, the unit of analysis was a Federal Circuit decision.[66] Because Holte and Sichelman (2019) only counted decisions in which there was an "actual obviousness determination,"[67] we excluded from the Rantanen (2013) dataset decisions that

---

62. *See* Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 726–27; Holte & Sichelman, *supra* note 8, at 140; *see also* Holte & Sichelman Codebook, *supra* note 51.

63. *See* Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 737–38; *see also* Rantanen, *Judicial Opinions*, *supra* note 9; Holte & Sichelman, *supra* note 8, at 161. An example of this would be a Federal Circuit decision vacating summary judgment of obviousness and remanding to the district court.

64. The complete Stata code for the construction of the comparison dataset and the analyses contained in this article will be made publicly available at empirical.law.uiowa.edu and archived on the Harvard Dataverse. Rantanen, et al., *Replication Data, supra* note 22.

65. The reason for treating these decisions separately is discussed in Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 727 n.89. One decision was coded as a preliminary injunction in one dataset and a bench trial in another; we kept that decision in the dataset.

66. *See generally* Rantanen, *Judicial Opinions*, *supra* note 9 (discussing the different levels at which judicial analyses in patent opinions can be analyzed). Differences in the unit of analysis for nonobviousness would only have a de minimis effect due to the rarity of inconsistent outcomes in these analyses. *See generally id*. In order to compare decisions at the decision-level, decisions with multiple analyses that were coded differently from the Rantanen (2013) dataset were labeled "Mixed." As the data in this article show, these were rare. "Mixed" decisions are discussed in the footnotes in Part I.A. and in the text in Part I.B.

67. *See* Holte & Sichelman, *supra* note 8, at 136. For example, Holte & Sichelman (2019) did not include decisions to vacate in their data set. *See id.* at 146 n.251.

had been coded as "2-No Final Determination."[68] This resulted in 179 decisions from Rantanen (2013) and 153 decisions from Holte and Sichelman (2019).[69]

The remainder of this section reports descriptive statistics comparing the results for the two subsets from Rantanen (2013) and Holte and Sichelman (2019) meeting the criteria described above. We focus on those elements of the studies that are most comparable, recognizing that there are other facets of each study that go beyond what we can directly compare.

2. Numbers of Pre- and Post-decision Records in Each Dataset

As shown in the table below, while the two datasets contained almost the same number of decisions for the pre-decision time period, the set from Rantanen (2013) contained 115 decisions for the period between April 30, 2007 and April 30, 2012 while Holte and Sichelman (2019) contained 89.

TABLE 1: RECORDS IN DATASETS

	Pre-certiorari Records in Dataset	Records Between Certiorari & Decision	Post-decision Records in Dataset
Rantanen (2013)	44	20	115
Holte & Sichelman (2019)	47	17	89

This difference is surprising: while both studies used the same general decision selection criteria, there were substantially more decisions included in Rantanen (2013) for the post-decision time period.

3. Procedural Postures of Decisions in Each Dataset

To examine possible reasons for the difference in cases—and to see whether there were differences in other results between the studies—we next compared the results for the procedural posture of the appeal, the § 103 determination at the district court level, the outcome on § 103 at the appellate level, and whether the appellate court affirmed or reversed the district court. Typically, an appeal involving obviousness arising from patent infringement litigation in the district court will result from a grant of summary judgment, a

---

68. *See id.* at 146 n.252. This resulted in the exclusion of 33 decisions from Rantanen (2013).

69. We debated whether to keep decisions that met the "type" criteria in one study but not the other. Ultimately, we decided to not include a decision in a particular dataset if it did not meet the criteria, even if it met the criteria in the other dataset. This meant that a small number of decisions that were in both datasets were treated as if they were in just one dataset. The consequence of this was to remove four decisions from the Rantanen (2013) dataset that were present in Holte & Sichelman (2019) and four decisions from the Holte & Sichelman (2019) dataset that were present in Rantanen (2013). Additional details are available in the Stata code file "2021-02-05 Construction of Comparison Dataset.do."

bench trial, or a jury verdict.[70] Generally, jury verdicts are appealed in a grant or denial of a renewed motion for Judgment as a Matter of Law ("JMOL").[71]

As the below table shows, while the number of records in each dataset involving an appeal from a grant of summary judgment and bench trials was about the same, Rantanen (2013) contained many more decisions coded as appeals arising from jury verdicts while Holte and Sichelman (2019) contained more decisions coded as arising from JMOLs. The difference for jury verdicts appears especially sharp in the post-*KSR* time period: Rantanen (2013) had forty-six records coded as arising from jury verdicts in the post-*KSR* time period (40% of the decisions for that period) while Holte and Sichelman (2019) had twenty-one (23% of the decisions for that time period). We explore likely explanations for this difference in more depth below in Part I.G.

TABLE 2: PROCEDURAL POSTURE OF DISTRICT COURT DETERMINATIONS REVIEWED BY FEDERAL CIRCUIT

	Rantanen (2013)	Holte & Sichelman (2019)[72]
Jury	70	37
Bench	47	44
JMOL	11	26
Summary Judgment	51	43

Setting aside the differences in how researchers coded Jury versus JMOL, however, the procedural posture results are relatively close: 45% of Rantanen (2013) involved an appeal from a jury decision (either coded as Jury or JMOL) versus 42% of Holte and Sichelman (2019); 26% versus 30%, respectively, involved an appeal from a bench trial, and in both datasets 28% involved an appeal from a grant of summary judgment. This is also true for the post-*KSR* period: if one adds jury determinations to JMOLs post-*KSR*, this results in fifty-2 decisions (45%) in Rantanen (2013) and 34 decisions (37%) in Holte and Sichelman (2019)—still a notable difference, but not as large as the difference when examining the Jury versus JMOL data individually.

---

70. *See* Holte & Sichelman, *supra* note 8, at 147.
71. 1 STEVEN ALAN CHILDRESS & MARTHA S. DAVIS, FEDERAL STANDARDS OF REVIEW § 5.07 (4th ed., 2019).
72. Holte & Sichelman (2019) also had three decisions with multiple procedural postures; we refer to these as "Multiple Postures."

4. Obviousness Outcomes of Decisions in Each Dataset

Both Holte and Sichelman (2019) and Rantanen (2013) report data on the district court determinations reviewed by the Federal Circuit.[73] Rantanen (2013) refers to these as the "substrate" of the appeal: the determinations on obviousness at the district court that are actually being reviewed by the appellate court.[74] The below table compares the rates at which each study reported that the determination being reviewed by the Federal Circuit had reached a conclusion of "obvious."

TABLE 3: DISTRICT COURT DETERMINATIONS REVIEWED
BY THE FEDERAL CIRCUIT

	Pre-decision in *KSR*	Post-decision in *KSR*
Rantanen (2013)[75]	46% (29/63)	47% (54/114)
Holte & Sichelman (2019)[76]	48% (29/60)	55% (48/86)

Overall, this table shows consistent results for both studies for the pre-*KSR* period but modestly different results for the post-*KSR* period. We discuss potential explanations for this difference in more detail below in Part I.D.

Outcomes at the appellate level reflect a similar pattern. As shown in Table 4, the descriptive statistical comparison for the pre-decision period (January 1, 2003, to April 30, 2011) is very similar for the obviousness outcomes. However, for the post-*KSR* period, the number of "nonobvious" outcomes reported by Rantanen (2013) was higher. The consequence of this difference is that the Holte and Sichelman (2019) dataset indicates a higher rate of obviousness than the Rantanen (2013) dataset. However, both datasets provide evidence of a shift in § 103 outcomes at the Federal Circuit in appeals arising from the district courts between the pre- and post-*KSR* periods.[77]

---

73. Holte & Sichelman, *supra* note 8, at 146; Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 740–44.

74. Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 740.

75. One decision in Rantanen (2013) did not have the determination under review coded and two decisions had "Both" outcomes; these are not included in the tables or percentages.

76. In addition, there were seven decisions in Holte & Sichelman (2019) with "Both" outcomes as the determination being reviewed. These are not included in the table or percentages.

77. As shown in Rantanen, *Obviousness Jurisprudence*, *supra* note 8 and Mojibi, *supra* note 17, the difference is even greater when the period between grant of certiorari and the Court's announcement of its decision is excluded. Excluding this period from the obviousness analysis makes sense because the Court may have been signaling that the Federal Circuit's obviousness standard was too low. *See* Mojibi *supra* note 17, at 585. In fact, running the analysis for the portion of the comparison period (starting January 1, 2003) prior to certiorari results in an "obvious" outcome of just 42% (Rantanen (2013) data) and 41% (Holte & Sichelman (2019) data).

TABLE 4: FEDERAL CIRCUIT DETERMINATIONS OF "OBVIOUS"

	Pre-decision in *KSR*	Post-decision in *KSR*
Rantanen (2013)[78]	50% Obvious (31/62)	57% Obvious (63/110)
Holte & Sichelman (2019)	52% Obvious (33/64)	66% Obvious (59/89)

5. Appellate Dispositions in Each Dataset

As shown in Table 5, the overall affirmance rates were about the same for both before and after *KSR*, and were approximately the same between the two studies. Because Holte and Sichelman (2019) did not include decisions to vacate, it was not possible to compare those outcomes.[79]

TABLE 5: FEDERAL CIRCUIT DISPOSITIONS WHEN MAKING FINAL DETERMINATIONS ON OBVIOUSNESS

	Pre-decision in *KSR*	Post-decision in *KSR*
Rantanen (2013)[80]	87% Affirmance (55/63)	86% Affirmance (95/110)
Holte & Sichelman (2019)[81]	86% Affirmance (54/63)	79% Affirmance (70/89)

The results between the two studies were generally consistent, except for the coding for Jury versus JMOL and Holte and Sichelman (2019)'s modestly higher comparative obviousness rates for the five years after *KSR*. This is good! It suggests that two different research groups that were trying to study the same phenomenon saw basically the same thing. The results from both datasets consistently show that about half of the district court determinations that the Federal Circuit reviewed arose from a jury verdict (either in a grant or denial of a JMOL), a higher obviousness rate at the Federal Circuit during the five years following *KSR*, and a very high affirmance rate for district court determinations involving § 103. And yet, there was still a substantial difference between the two datasets, both in terms of the number of cases studied and the reported values for some metrics.

---

78. In addition, there were seven "Both" outcomes at the Federal Circuit (2 pre-*KSR* decision and 5 post-*KSR* decision) in the Rantanen (2013) dataset. These are not included in the table or percentages.

79. For a discussion of why decisions to vacate can matter, see Rantanen, *Judicial Opinions*, *supra* note 9, at 263–65.

80. In addition, Rantanen (2013) had five "Mixed Dispositions" and one decision that did not have disposition coded. These are not included in the table or percentages.

81. In addition, Holte & Sichelman (2019) had one "Mixed Disposition." This is not included in the table or percentages.

D. DATASET COMPARISON

Looking at summary statistics only provides a general comparison of the two datasets, however. A more thorough analysis requires comparing the actual records within the two datasets. In particular, the difference in the number of cases studied, the obviousness rates for the post-*KSR* period, and the difference in the number of decisions in the two datasets, suggests the need for additional analysis.

Direct comparison analysis reveals much greater differences in the composition of the two datasets than the above summary statistics suggest. Fewer than two thirds of the cases were the same across the two studies. However, for the decisions that matched (i.e., were contained in both datasets), the coding was highly consistent—except for the Jury/JMOL coding described above and some of the coding relating to judicial reasoning.

1. Methodology For Direct Record Comparison

Historically, direct record comparison in empirical studies of patent law decisions has not been a simple task. To the best of our knowledge, this Article is the first to report this type of analysis. The ability to compare studies at the record level has been frustrated by a lack of publication of (or even access to) study data, different mechanisms for identifying and recording individual decisions, and the lack of a universal identifier to allow records to be matched across studies.[82] This has restricted the ability to use automated matching and merge techniques. Comparing data for large datasets is especially challenging because of the number of individual determinations and matches that must be made.

Consequently, the first task for the direct record comparison was to identify all matches between records in Holte and Sichelman (2019) and Rantanen (2013).[83] We initially matched using an algorithmic analysis—essentially using automated text comparison techniques to find as many matches as possible between citations and case names—followed by human review. The end goal was to find every match between the two datasets through an iterative process. The final dataset contains all records from the Rantanen (2013) and Holte and Sichelman (2019) sets generated in Part I, along with the corresponding data from the other dataset, if a match was made.[84]

To evaluate how much agreement there is between the Rantanen (2013) and Holte and Sichelman (2019) data, we analyze the inter-rater reliability

---

82. *See* Rantanen, *Judicial Opinions*, *supra* note 9, at 12. Different studies use different case name conventions, and while official Reporter citations can be used for published, precedential opinions, variations in how this data is recorded pose challenges for automated matching. In addition, some datasets do not include official reporter citations for some or all decisions. Furthermore, some decisions—particularly Rule 36 summary affirmances—may share the same reporter citation.

83. We also matched the records to the *Compendium of Federal Circuit Decisions* to allow for improved future analysis of the data.

84. The final matched dataset is available at *Studies and Data*, *supra* note 58.

between the two studies. Inter-rater reliability refers to the level of agreement between two or more raters. Our goal is to quantify the level of agreement between Holte and Sichelman (2019) and Rantanen (2013). While percentages can provide some information about agreement, relying on percentages alone is insufficient because there may be instances in which two individuals agree purely by chance.

For all the inter-rater reliability measures we calculate below, we make a few key assumptions. First, we treat Holte and Sichelman as one "coder" and Rantanen as the second coder. In other words, rather than treating an individual human as the rater, we treat the study's final dataset—the coding that Holte and Sichelman agreed on—as the coder. This makes sense because our goal is to assess the amount of agreement between these two studies. The remaining assumptions are necessary to calculate the inter-rater agreement statistics. For our second assumption, we assume the two raters coded the data independently, i.e., Holte and Sichelman's coding did not affect Rantanen's rating and vice versa.[85] We also assume there were no additional coders and only Holte and Sichelman and Rantanen made the final determination on all the observations. Our final assumption is that each variable had the same number of categories. For example, when coding the appellate court's disposition of the case, there were only two categories: obvious or nonobvious.

To assess inter-rater reliability, we calculate kappa statistics.[86] Kappa (κ) ranges from –1 to +1, with –1 meaning less agreement than chance and +1 representing complete agreement.[87] A score of zero indicates the coding was reached by pure chance. Since we have a fixed number of raters (two) and all nominal data, we primarily rely on Cohen's kappa statistic.[88]

We do not expect to discover perfect reliability. We interpret kappa using Landis and Koch's[89] benchmark scale calculated using Gwet's probabilistic-

---

85. While Rantanen did provide his dataset to Holte and Sichelman early in their project, Holte and Sichelman confirmed that their fields were coded independently of the Rantanen (2013) dataset. *See* E-mail from Ryan Holte & Ted Sichelman, Professor of L., Univ. of San Diego Sch. of L., to Jason Rantanen, Professor of L., Univ. of Iowa Coll. of L. (Feb. 4, 2021) (on file with corresponding author).

86. The kappa coefficient is measured by the following formula:

$$k = \frac{p_o - p_e}{1 - p_e}$$

where p_o is the overall proportion of observed agreement and p_e is the overall proportion of agreement expected by chance. Jacob Cohen, *A Coefficient of Agreement for Nominal Scales*, 20 EDUC. & PSYCH. MEASUREMENT 37, 39–40 (1960).

87. *Id.* at 41–42.

88. *See id.* at 39–43. There are many kappa alternatives available, but scholars created the alternative formulas to accommodate multiple raters and/or ordinal data, which are unnecessary for our data.

89. J. Richard Landis & Gary G. Koch, *The Measurement of Observer Agreement for Categorical Data*, 33 BIOMETRICS 159 (1977).

based method.[90] Landis and Koch qualify their agreement categories as "Poor," "Slight," "Fair," "Moderate," "Substantial," and "Almost Perfect" depending on the magnitude of kappa.[91] Additional data on the kappa analysis is provided in Appendix A.

2. Decisions Included in the Studies

Comparison of the records in the two datasets revealed 125 records that were found in both datasets, 28 that were found only in Holte and Sichelman (2019), and 54 that were found only in Rantanen (2013).[92] Of the 28 records found only in Holte and Sichelman (2019), 15 were pre-decision and 13 were post-decision. Of the 54 records found only in Rantanen (2013), 15 were pre-decision and 39 were post-decision. In fact, even though the numbers of records from the pre-cert period were about the same, the actual overlap for that period was only 49 out of 79 records. And the difference for the post-cert period was even more than the summary statistics in Part I.B suggested: only 76 records were in both studies out of 128 records.

TABLE 6: COMPARISON OF DECISIONS INCLUDED IN EACH DATASET

Case Included in Holte & Sichelman (2019)	Case Included in Rantanen (2013)		
	No	Yes	Total
No	0	54	54
Yes	28	125	153
Total	28	179	207

This is a remarkably poor match! Despite both sets of coders being highly sophisticated in patent law, and using the same criteria, the two sets of coders concluded they should include a given Federal Circuit decision in the dataset only 60% of the time; yet based on chance alone we would expect that the coders would agree 67% of time. While there are an unknown number of decisions that

---

90. Daniel Klein, *Implementing a General Framework for Assessing Interrater Agreement in Stata*, 18 STATA J. 871, 879–80 (2018); KILEM L. GWET, HANDBOOK OF INTER-RATER RELIABILITY: THE DEFINITIVE GUIDE TO MEASURING THE EXTENT OF AGREEMENT AMONG RATERS (4th ed. 2014).

91. To interpret the kappa values, we use the following benchmark guidelines: 0.0–0.20: slight; 0.21–0.40: fair; 0.41–0.60: moderate; 0.61–0.80: substantial; and 0.81–1: almost perfect. Note that because we use Gwet's probabilistic-based method, a given kappa value does not directly translate into a benchmark range. We report the benchmark agreement. Additional details are in Appendix A.

92. As noted in Part I.C., there were a few decisions that were contained in both studies, but which were coded by one set of authors or the other in a way that did not meet our inclusion criteria. For example, even if a decision appeared in Holte & Sichelman (2019) and was coded as having a final determination, if that decision was coded as "No Final Determination" in Rantanen (2013), it was not considered to be part of the Rantanen (2013) dataset that we used for the comparison. *See supra* note 72. In total, there were four decisions from Rantanen (2013) that were dropped from the Rantanen (2013) dataset used in this comparison for this reason and four decisions from Holte & Sichelman (2019) that were dropped from the Holte & Sichelman (2019) dataset used in this comparison.

both studies did not include ("No-No's"), the lack of agreement on marginal cases is surprising—particularly when viewed in relation to the number of decisions that both studies included ("Yes-Yes's"). And this is just the threshold determination of whether the decision involved a final § 103 analysis.

To further examine the differences, one of us[93] blind-coded[94] the eighty-two records found in only one of the two datasets.[95] The results are below. The bottom line is that most of the cases found in only one of the two datasets probably should have been included in both studies.

TABLE 7: ANALYSIS OF DECISIONS INCLUDED IN ONLY ONE DATASET

Dataset	Determination of Whether Decision Involved § 103				
	Definitely	Not Sure	Definitely Not	Rule 36	Total
Holte & Sichelman (2019) only	13	4	5	7	28
Rantanen (2013) only	29	3	1	21	54
Total	42	7	6	28	82

This observation, coupled with further review of the opinions, suggests that most of the difference appears to be due to type II errors, false negatives. Both studies missed opinions that should probably have been included but did not include many opinions that probably should not have been included.[96]

---

93. Kriz blind-coded the records.

94. By "blind-coded" we mean that Rantanen provided Kriz with a set of case names and citations, organized chronologically, with instructions to apply a three-point scale: (1) Decision clearly contains an obviousness determination, (2) it is a judgment call as to whether decision contains an obviousness determination, or (3) decision clearly does not contain an obviousness determination. The specific criteria was that:

> An obviousness determination is defined as whether or not the Federal Circuit addressed the obviousness of a utility patent in the context of a validity determination. As a result, obviousness determinations involving design patents were excluded, as were obviousness determinations for the purpose of determining whether there was an interference-in-fact in the context of an interference proceeding. Obviousness-type double-patenting analyses were also excluded.

Decisions from each study were interspersed and there was no indication of which decision was from which study. It was preferable that Kriz did this as Rantanen may have been biased due to having conducted the prior study.

95. The Rule 36 decisions were not included in this analysis.

96. In response to our observation, the authors of Holte & Sichelman (2019) provided additional detail on their decision selection methodology beyond what was described in Holte & Sichelman (2019). E-mail from Ted Sichelman, Professor of L., Univ. of San Diego Sch. of L., to Jason Rantanen, Professor of L., Univ. of Iowa Coll. of L. (Jan. 29, 2021) (on file with author) [hereinafter Jan. 2021 E-mail from Ted Sichelman to Jason Rantanen]. Specifically, they noted that because Holte & Sichelman (2019) focuses on judicial reasoning, they excluded district court opinions resting solely on jury determinations unless they could locate the JMOL order on Westlaw. Because they significantly relied on district court cases to identify associated Federal Circuit

In addition, we examined the characteristics of the decisions that were only in one dataset or the other. The following tables summarize the characteristics of records that were only in Rantanen (2013) or Holte and Sichelman (2019) for the pre- and post-*KSR* time period.

TABLE 8: VARIABLE COMPARISONS OF DECISIONS IN ONLY ONE DATASET

Procedural Posture	Dataset		
	Holte & Sichelman (2019) only	Rantanen (2013) only	Total
Jury	7	27	34
Bench	9	14	23
JMOL	5	1	6
SJ	7	12	19
Total	28	54	82

Posture Outcome	Dataset		
	Holte & Sichelman (2019) only	Rantanen (2013) only	Total
Obvious	14	19	33
Nonobvious	13	35	48
Both	1	0	1
Total	28	54	82

CAFC Result	Dataset		
	Holte & Sichelman (2019) only	Rantanen (2013) only	Total
Obvious	12	18	30
Nonobvious	16	34	50
Mixed	0	2	2
Total	28	54	82

appeals, this methodology effectively reduced the number of appeals that they identified that arose from jury verdicts in the context of a denial of a renewed motion for JMOL. In addition, they indicated that they "focused their coding on Federal Circuit decisions with some reasoning, either in the district court opinion, or in the Federal Circuit opinion itself," and that the consequence of this was to not include Federal Circuit opinions that merely affirmed a jury determination with no reasoning. *Id.* Judge Holte and Professor Sichelman indicated that 17 of the decisions that were in Rantanen (2013) but not in Holte & Sichelman (2019) were reviewed in an earlier round of case coding but they decided not to include them in the final dataset for this reason. *Id.* Because these decisions were not contained in the dataset for *Cycles of Obviousness*, however, we did not include them in our analysis here. Nor have we (the authors of this paper) tested whether there is a significant difference in the amount of judicial reasoning in the opinions that were included versus those that were not, and so we cannot say one way or another whether this is distinction is reflected in the data.

Disposition	Dataset		
	Holte & Sichelman (2019) only	Rantanen (2013) only	Total
Affirmed	21	50	71
Reversed	7	2	9
Mixed	0	2	2
Total	28	54	82

This comparison suggests that the differences in which cases were included in the two studies affected some of the results the studies reported. Particularly, for outcomes at the district court: while two thirds of the decisions that were only in Rantanen (2013) involved a district court outcome of nonobvious, only about half of the decisions that were only in Holte and Sichelman (2019) did. In addition, half of the decisions that were only in Rantanen (2013) arose from denials of JMOLs, whereas only 25% of the decisions that were only in Holte and Sichelman (2019) did. The differences for the other two fields (outcomes at the Federal Circuit and affirmances) were present, but not as substantial.

The bottom line is that despite using what appears to be the same inclusion criteria based on the methodological descriptions within the articles and the codebooks, the two research teams made different decisions about case inclusion. At the composite level, out of the 207 decisions in the comparison set that were contained in one or both studies, only 125 were in both studies; with at least another 41 that probably should have been included. In addition, these differences affected results: the cases that were only in Rantanen (2013) were characterized by more appeals from denials of JMOLs and fewer outcomes of "obvious" both in the cases being reviewed and in the Federal Circuit's opinions. Before exploring likely reasons for this difference, however, we also wanted to examine whether there were differences in how the two teams coded information at an individual record level for those that both included in their datasets.

E. INDIVIDUAL RECORD COMPARISON

Besides examining the characteristics of decisions that were contained in only one of the two datasets, we also looked at how the sets of authors coded decisions once they were in the dataset. This section compares the coding for the 125 decisions that met the comparison set requirements. As with including cases, they used the same written criteria for both studies.[97] Overall, agreement between the two studies was high, with some notable exceptions.

---

97. *See* RANTANEN CODEBOOK, *supra* note 51; Holte & Sichelman Codebook, *supra* note 51.

1. Procedural Posture

Overall, there was high agreement in the coding for these cases. The kappa coefficient is 0.72 (standard error of 0.05) and is statistically significant at the 0.05 level. This suggests that there is substantial agreement between Holte and Sichelman (2019) and Rantanen (2013) on the procedural posture variable.[98]

TABLE 7: COMPARISON OF PROCEDURAL POSTURES FOR
DECISIONS IN BOTH DATASETS

Holte & Sichelman (2019) Procedural Posture	Rantanen (2013) Procedural Posture					
	Jury	Bench	JMOL	SJ	Multiple Postures	Total
Jury	26	1	1	2	0	30
Bench	1	31	0	3	0	35
JMOL	13	0	8	0	0	21
SJ	1	1	0	34	0	36
Multiple Postures	2	0	1	0	0	3
Total	**43**	**33**	**10**	**39**	**0**	**125**

The biggest area where the Rantanen (2013) data differed from that of Holte and Sichelman (2019) is in classifying a case as a jury trial versus a JMOL. Of the fifteen disagreements on "JMOL," for example, fourteen were because Rantanen (2013) classified the decision as arising from a jury verdict, while Holte and Sichelman (2019) classified it as arising from a JMOL. This difference appears to be a result of different coding methodologies: while Rantanen (2013) consistently coded denial of a JMOL as "Jury,"[99] Holte and Sichelman (2019) sometimes recorded denials of a JMOL in their dataset of Federal Circuit decisions as "Jury" and sometimes as "JMOL."[100] As noted above, this is an area where coding instructions may result in more consistent coding.

2. Coding of Obviousness Outcomes

There was high agreement on the coding of the district court determinations being reviewed by the Federal Circuit, with 111 agreements out of 125 records. The kappa coefficient is 0.79 (standard error 0.05), representing substantial agreement.

---

98.   The percentage agreement between the datasets for the Procedural Posture field was 79%.
99.   *See* RANTANEN CODEBOOK, *supra* note 51.
100.   Jan. 2021 E-mail from Ted Sichelman to Jason Rantanen, *supra* note 96.

TABLE 8: COMPARISON OF DISTRICT COURT OUTCOMES FOR
DECISIONS IN BOTH DATASETS

Holte & Sichelman (2019) Lower Court Result	Rantanen (2013) Lower Court Result			
	Obvious	Nonobvious	Both	Total
Obvious	57	4	2	63
Nonobvious	2	54	0	56
Both	5	1	0	6
Total	64	59	2	125

There was also very high agreement on coding of the outcomes at the Federal Circuit; there were 114 agreements on § 103 and 11 disagreements.[101] Overall agreement was 91%.[102] The Cohen's kappa coefficient was 0.82 (standard error 0.05), representing almost perfect inter-rater reliability.[103]

TABLE 9: COMPARISON OF FEDERAL CIRCUIT OUTCOMES FOR
DECISIONS IN BOTH DATASETS

Holte & Sichelman (2019) Federal Circuit Result	Rantanen (2013) Federal Circuit Result			
	Obvious	Nonobvious	Mixed Outcomes	Total
Obvious	73	3	4	80
Nonobvious	3	41	1	45
Mixed Outcomes	0	0	0	0
Total	76	44	5	125

---

101. In addition, there were three decisions that Rantanen (2013) coded as being "No Final Decision" and Holte & Sichelman (2019) coded as reaching a final result on obviousness.

102. Taking into account the three cases that Rantanen (2013) coded as being "No Final Decision."

103. Standard error of 0.05 is statistically significant at the 0.05 level. Overall agreement was 91%. For comparison, Rantanen (2013) reported a Cohen's kappa for the CAFC_Result category of 0.96. Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 766.

3. Coding of Appellate Disposition

TABLE 12: COMPARISON OF APPELLATE DISPOSITIONS FOR
DECISIONS IN BOTH DATASETS

Holte & Sichelman (2019) Federal Circuit Disposition	Rantanen (2013) Federal Circuit Disposition			
	Affirmed	Reversed	Mixed Dispositions	Total
Affirmed	100	1	2	103
Reversed	0	19	1	20
Mixed Dispositions	0	1	0	1
Total	100	21	3	124

As Table 12 shows, there was significant agreement on the coding of appellate disposition. Cohen's kappa agrees, producing a coefficient of 0.87 (standard error 0.05)—almost perfect agreement.[104]

F. CODING OF JUDICIAL REASONING

One important component of both studies is the examination of the court's actual reasoning. Both studies looked at the language of the court's opinions relating to their § 103 analysis, examining some core hypotheses in the literature. In particular, the two studies examined the changes in which the Federal Circuit opinions (1) used some recognizable form of the teaching-suggestion-motivation ("TSM") test, (2) used the "formal" version of the TSM tests, and (3) invoked the concept of "common sense."[105] Both studies found that there had been a decline in the Federal Circuit's use of the TSM test post-*KSR*, particularly in the court's use of a formal version of the test, and an increase in the court's invocation of "common sense."[106] Neither study considered Rule 36 summary affirmances in this analysis.[107]

Comparing the coding for the set of opinions included in both studies revealed mixed results. For this analysis, entries coded as "Blank" are treated as the absence of one of these elements of judicial reasoning. Inter-rater reliability was highest for whether the court applied common sense; the kappa was 0.80 (standard error 0.10), representing substantial agreement.

---

104. The percentage agreement between the two datasets was 96%. One record was not coded for appellate disposition in Rantanen and was not included in the analysis.

105. For additional details on these categories, *see* Rantanen *Obviousness Jurisprudence*, *supra* note 8, at 752–57.

106. *See* Rantanen, *Obviousness Jurisprudence*, supra note 8, at 759–60; *see also* Holte & Sichelman, *supra* note 8, at 129–30.

107. *See* Rantanen, *Obviousness Jurisprudence*, *supra* note 8; *see also* Holte & Sichelman, *supra* note 8, at 155.

The other two methods of judicial reasoning, however, did not fare as well. The kappa coefficient for the "Formal" version of the TSM test was 0.67 (standard error 0.09), representing substantial agreement. Inter-rater reliability agreement for whether there was *some* use of the TSM test was even lower. The kappa coefficient was 0.38 (standard error 0.09) indicating only fair agreement, two steps below the substantial category on the benchmark scale.

TABLE 10: COMPARISON OF ANALYSIS CODING FOR DECISIONS
CONTAINED IN BOTH DATASETS

Holte & Sichelman (2019) Court Used Common Sense	Rantanen (2013) Court Used Common Sense		
	Yes	No	Total
Yes	9	3	12
No	1	97	98
Total	10	100	110

Holte & Sichelman (2019) TSM Formal	Rantanen (2013) TSM Formal		
	Yes	No	Total
Yes	15	3	18
No	8	84	92
Total	23	87	110

Holte & Sichelman (2019) Some Use of TSM	Rantanen (2013) Some Use of TSM		
	Yes	No	Total
Yes	33	16	49
No	18	43	61
Total	51	59	110

This makes sense: both the "Formal" version of the TSM test and the use of "common sense" are relatively objective determinations based on the presence of certain language in the opinion, whereas whether there was *some use* of the TSM test would naturally be more subjective. For this reason, when referencing data from these studies, it may be better to rely on the more objective—and replicable—metrics.

G. REPLICATION OF REGRESSIONS

Besides comparing the coding for the two studies, we also replicated two of the regressions from Table 2 of the Appendix to Holte and Sichelman (2019).[108] These regressions supported Holte and Sichelman's conclusion that *KSR* shifted its outcomes substantially (if not as much as the district courts).[109]

Using the same technique as Holte and Sichelman (2019), we applied a probit model using Holte and Sichelman (2019)'s original dataset,[110] the dataset from Rantanen (2013), and the decisions that were only present in both datasets. In all three cases, our results were consistent with those of Holte and Sichelman (2019): there was a significant increase in outcomes of "obvious" after the Supreme Court granted certiorari in *KSR*, regardless of which dataset we used. The complete results of these regressions are in Appendix C.

H. REASONS FOR THE DIFFERENCES BETWEEN DATASETS

By far the biggest issue when comparing these studies was the disagreement on which cases to include in each study. This difference highlights the importance of carefully documenting and describing the method for record selection in empirical legal studies and providing information on which data points scholars include in the study.[111] Including a mechanism to allow for comparison with other studies can also be beneficial for future cross-study comparisons. For example, for studies involving the Federal Circuit, researchers may want to consider using the uniqueID identifier provided in the *Compendium of Federal Circuit Decisions*.[112]

While both studies had documented case inclusion criteria, one possibility is that those criteria may have been unclear or not detailed enough. That most of the inclusion errors were false negatives indicates the coders did not misunderstand the instructions completely but interpreted the inclusion criteria instructions conservatively. For the Rantanen (2013) study, coders were instructed that an "obviousness determination was defined as whether or not the Federal Circuit addressed the obviousness of a utility patent in the context of a validity determination."[113] Coders for Holte and Sichelman (2019) were more simply instructed to determine "if the CAFC actually holds the patent obvious" (or nonobvious).[114] Given the simplicity of both instructions, it is reasonable to believe coders were erring on the side of caution when deciding and would have been less likely to indicate a decision involved nonobviousness unless it was explicitly clear.

---

108. *See* Holte & Sichelman, *supra* note 8, at 169.

109. *Id.* at 154–55.

110. We dropped the two cases that were from the Court of Federal Claims rather than the Federal Circuit.

111. Rantanen, *Response to Holte & Sichelman*, *supra* note 60.

112. Jason Rantanen, *The Landscape of Modern Patent Appeals*, 67 AM. U. L. REV. 985, 1001 (2018) [hereinafter Rantanen, *Modern Patent Appeals*].

113. *Id.*

114. Holte & Sichelman Codebook, *supra* note 51, at 11.

For example, when determining whether to include decisions where the Federal Circuit was reviewing a summary judgment or JMOL, coders may have strictly adhered to the inclusion criteria. Because judges are analyzing summary judgments and JMOLs in the context of whether the accused infringer met the evidentiary burden, coders may not have considered the opinions to be addressing obviousness in the validity or holding the patent to be obvious or nonobvious. *Warsaw Orthopedic, Inc. v. NuVasive, Inc.*[115] illustrates this point. That opinion involved an appeal from a denial of a JMOL on obviousness among other issues. The opinion is succinct in addressing obviousness, with a single paragraph concluding that the district court did not err without an in-depth analysis of the issue.[116]

Indeterminacy is another explanation. It may be that for the question of whether to include a given Federal Circuit decision, there are some decisions that are simply difficult or impossible to categorize into the buckets of "obviousness determination" or "no obviousness determination." Consider, for example, *Flex-Rest, LLC v. Steelcase, Inc.*,[117] a decision that Holte and Sichelman (2019) included but Rantanen (2013) did not. That case involved an appeal of a jury verdict finding the patent in suit invalid for anticipation and obviousness.[118] The patent owner's challenge on appeal was to the district court's ruling to exclude testimony by its expert on the issues of anticipation and obviousness, a decision that the Federal Circuit affirmed.[119] While this decision relates to obviousness, it does not directly decide the issue. Instead, it is collateral. Opinions like these are difficult to classify and even extremely clear coding instructions may not be adequate.

Yet, it is also important to recognize that despite the disagreements in which cases were included and which were not, the reported results were directionally consistent and the coding for most aspects of the court's decisions agreed between the two studies. There were only two notable exceptions: classification of a decision as arising from a jury verdict or JMOL and coding of the TSM_Use judicial reasoning field.

The differences in Jury vs. JMOLs in the two datasets raise two important questions for appellate decision analysis studies. One is the question of when an appellate decision involving a jury should be included at all; the second is the difficulty for coders of differentiating between an appeal involving a "jury verdict" and an appeal involving "JMOL"—particularly because, procedurally,

---

115. Warsaw Orthopedic, Inc. v. NuVasive, Inc., 778 F.3d 1365 (Fed. Cir. 2015).

116. *Id.* at 1370.

117. Flex-Rest, LLC v. Steelcase, Inc., 455 F.3d 1351 (Fed. Cir. 2006).

118. *Id.* at 1356. The patent owner did not move for JMOL or a new trial.

119. *Id.* at 1360.

in order to appeal a "jury verdict," it's almost always necessary to seek a JMOL.[120]

On the first question, an appeal from a jury verdict involves a more complex procedural framework than a review of a bench trial: it incorporates both the jury's verdict and (almost always) a judge's decision on a renewed motion for JMOL.[121] Thus, if the goal is to focus on *judicial* decision-making, then analyzing only appeals from bench trials makes sense. On the other hand, juries play a very important part in the adjudication of disputes involving obviousness, so excluding them entirely means excluding a substantial portion of decisions.

A related concern involves how the appellate dataset is assembled. If the dataset of appellate decisions is being assembled based primarily on district court decisions, and denials of JMOLs are routinely not published, then the identification of appellate decisions will be systematically biased against including appeals from denials of JMOLs. This is a plausible explanation here. It also raises the broader concern that researchers must recognize that some types of orders may not be as available as other orders simply because courts do not publish them.

The second issue—how to code an appeal involving a "jury verdict" versus an appeal involving "JMOL"—is more directly addressed. This is an area where it would be beneficial to have clearly developed objective criteria.[122] When a decision is an appeal from the *grant* of a JMOL, we should code it one way (Rantanen (2013) coded this as "JMOL").[123] When a decision is an appeal from *denial* of a JMOL—in other words, the jury's decision stands—we should code it a different way (Rantanen (2013) coded this as "Jury"). Being able to differentiate in this way may be especially important when analyzing how an appellate court treats district court decisions that grant versus deny JMOLs, an issue that can be important when assessing appellate deference to juries.

A third issue arises specifically in studies of § 103. Nonobviousness determinations are questions of law based on underlying findings of fact.[124] However, sometimes in jury trials the judge will effectively give the entire

---

120. CHILDRESS & DAVIS, *supra* note 71, § 5.07.

121. A renewed motion for judgment as a matter of law will be granted where the jury verdict lacks substantial evidence. ARTHUR R. MILLER, FED. PRAC. & PROC. CIV. § 2537 (3d ed., 2020). On appeal, questions of law, including summary judgments and JMOLs are reviewed de novo on appeal, so no deference is given to the district court's decision. DAVID G. KNIBB, FED. CT. APP. MANUAL § 31.3 (6th ed., 2020). Jury verdicts are reviewed under the rational basis standard, meaning that the court of appeals will set aside the jury verdict if there is no rational basis for the jury's conclusion. *Id.* § 31.5. In other words, the court of appeals may reverse the jury verdict if no reasonable and impartial fact finder could have reached the same conclusion as the verdict. *Id.* When there are mixed findings of law and fact, the applicable standard of review is determined by assessing the questions at issue and whether the questions are primarily factual or legal. *Id.* § 31.6. However, courts also tend to divide rulings to handle each distinct question separately under the appropriate standard of review, but this strategy only works to the extent rulings can be divided into respective parts. *Id.*

122. RANTANEN CODEBOOK, *supra* note 51.

123. *See id.* (providing this standard under "Procedural Posture").

124. *See, e.g.*, Graham v. John Deere, 383 U.S. 1, 17 (1966).

question to the jury (subject to a JMOL),[125] while in others the judge will have the jury render an "advisory" verdict. [126] Yet in others, the jury will make findings of fact while the judge will make the ultimate determination of nonobviousness.[127] These distinct possibilities complicate differentiations between whether a jury decided the procedural posture of the underlying nonobviousness determination or was instead a grant of JMOL. While typically we think of a JMOL as the district judge rejecting the jury's verdict as a matter of law, for nonobviousness disputes a jury might, for example, render an advisory verdict of nonobviousness that the judge agrees with. Is that a jury decision or JMOL? This may lead to indeterminacy as to procedural posture.

The final notable area of disagreement was the identification of the judicial reasoning used in the opinions. Given that the coders most often disagreed on whether the courts used some recognizable form of the TSM test but were relatively consistent when identifying when an opinion was using reasoning falling at each end of the spectrum, it is likely that identification of TSM is more indeterminate than identifying its specific linguistic manifestations. However, when judges stray from formal, established language, even students with significant patent law experience and law professors may find it difficult to agree on the theory behind this more fluid judicial reasoning. So, when there is no formal "test" being clearly articulated in an opinion, it is reasonable to expect variability in coding outcomes. This level of disagreement between the two studies on this aspect of the courts' opinions suggests it is especially important to develop explicit instructions and for studies to report a measure of inter-rater agreement metric for these variables. At the very least, it is important to identify which coded variables are more subjective than others and attempt to account for that subjectivity.

II. The Next Seven Years

Besides comparing the data used in Rantanen (2013) and Holte and Sichelman (2019), we extended the existing Rantanen (2013) dataset to the present. This updated dataset contains all Federal Circuit decisions in appeals arising from the district courts that involved a § 103 validity decision for a utility patent.[128]

---

125. *See* Duro-Last, Inc. v. Custom Seal, Inc., 321 F.3d 1098, 1104–05 (Fed. Cir. 2003).

126. *See* Kinetic Concepts, Inc. v. Smith & Nephew, Inc., 688 F.3d 1342, 1353 (Fed. Cir. 2012).

127. Bard Peripheral Vascular, Inc. v. W.L. Gore & Assocs., Inc., 682 F.3d 1003, 1008 (Fed. Cir. 2012) (noting that where mixed questions of law and fact are concerned, it is permissible to allow the jury to make any underlying factual determinations while leaving the ultimate legal conclusion to the judge).

128. As with Rantanen (2013) and Holte & Sichelman (2019), design patents were not included in this analysis. In addition, interferences and other appeals from the USPTO were not included.

A. METHODOLOGY

To assemble the complete dataset, we first updated the § 103 data from 1996 to April 30, 2012 to reflect everything that we learned from the analysis in Part I. This revised dataset includes all decisions that were in Rantanen (2013) and those from Holte and Sichelman (2019) that were not in Rantanen (2013) and which, after review, we determined met the study criteria. We also reviewed records that were in Rantanen (2013) and coded as "Not sure whether involved a § 103 issue" or "Definitely did not involve a § 103 issue" in Part I to determine whether we should retain them.[129]

In addition, in order to simplify and remain consistent with the analysis in Part I, we changed the unit of analysis to the "decision" level rather than the "analysis" level. As for the analysis conducted in Part I, cases with multiple consistent outcomes from Rantanen (2013) were coded under that outcome. For example, if all analyses were coded as "obvious," we coded the outcome for the case as "obvious." Cases with different outcomes (for example "obvious" and "nonobvious") were coded as "mixed." There were relatively few of these decisions.[130] The final dataset for the period from June 26, 1996, to April 30, 2012, contains 377 decisions.[131]

To update the dataset with decisions from May 1, 2012, to the end of 2019, we followed a process similar to that in Rantanen (2013), but with the benefits of hindsight and the *Compendium of Federal Circuit Decisions*. Because the *Compendium* contains text-searchable copies of all documents published on the Federal Circuit's website,[132] rather than drawing on a word search conducted in Westlaw, we were able to directly search the text of all Federal Circuit written opinions between 2012 and 2019 for opinions containing the string "obvious*" and "nonobvious*" in appeals arising from the district courts.[133] As in Rantanen (2013), one of the study's authors reviewed these opinions to determine whether they might "plausibly involve an obviousness issue," a coarse filter. Three research assistants with substantial experience coding Federal Circuit decisions then reviewed the resulting 273 opinions to determine whether each involved a § 103 issue based on the study selection criteria.[134] All three coders agreed on the yes/no determination for 218 out of the 273 (80% agreement). The kappa coefficient is 0.76 (standard error 0.03), indicating substantial agreement among

---

129. Each of the records from Holte & Sichelman (2019) that were added were independently coded by two research assistants—in other words, while we used Holte & Sichelman (2019) to ensure the completeness of the dataset, we did not copy the substantive coding from Holte & Sichelman (2019).

130. For example, 12 decisions from 1996 to 2012 had a "mixed" outcome at the Federal Circuit on the issue of whether the court concluded the claims on appeal were obvious.

131. As in Rantanen (2013) and Holte & Sichelman (2019), we did not include appeals involving interferences or preliminary injunctions in the analysis. However, these are included with the dataset.

132. Rantanen, *Modern Patent Appeals*, *supra* note 112, at 987–88.

133. For a description of the *Compendium of Federal Circuit Decisions*, see *id.* at 988.

134. The coders followed the same procedures used in the 2013 decisions but with some additions for improved clarity. *See* Rantanen CODEBOOK, *supra* note 51. A copy of the revised coding instructions is available in the project data archive. Rantanen, et al., *Replication Data, supra* note 22.

all three coders.[135] The remaining opinions were reviewed by both Rantanen and Kriz, in consultation, to make a final determination of whether they met the study criteria.

For Rule 36 summary affirmances, two research assistants reviewed the appellant's brief (and if necessary, the appellee's brief) for all Rule 36 summary affirmances arising from the district courts between 2012 and 2019 to determine whether a decision on obviousness was necessary to the court's affirmance. The coders agreed on 591 out of 598 yes/no determinations. Of the seven disagreements, six involved decisions for which one coder could not locate the briefs and thus did not make a determination. One of the study authors reviewed the seven decisions on which the coders did not agree to make the final determination of whether to include the decision in the study.

The complete dataset for 2012–2019 consisted of 154 written opinions and 76 summary affirmances. Two research assistants coded information about the procedural posture, outcome at the district court, outcome at the Federal Circuit, and Federal Circuit disposition. Initial agreement for these fields ranged from 84%–93%. We instructed the research assistants to review their coding for the decisions that contained at least one disagreement; after review, agreement ranged from 90%–94% depending on the field.[136] One of the study's authors reviewed the remaining disagreements to make the final determination.

Inter-rater reliability statistics for all measures show substantial agreement between the two coders. For the initial determination of the procedural posture variable, the kappa coefficient for inter-rater agreement is 0.75 (standard error 0.03). When coding the outcome of the district court, the kappa statistic is 0.85 (standard error 0.03). Agreement was also high for the outcome of the Federal Circuit; the kappa coefficient is 0.79 (standard error 0.03). Finally, there was also substantial agreement in coding the Federal Circuit disposition variable. The kappa statistic is 0.84 (standard error 0.04). Using probabilistic benchmark scale,[137] the two coders had substantial agreement on all variables.

Agreement between coders increased for three of the measures after coders reviewed decisions containing at least one disagreement. Following this review, the kappa coefficient for procedural posture is 0.86 (standard error 0.03). For Federal Circuit disposition variable, the kappa statistic is 0.91 (standard error 0.02). For the district court outcome variable, the kappa statistic is 0.89 (standard error 0.03). The kappa coefficient for Federal Circuit result is 0.84 (standard

---

135. Since there are three raters, we cannot use Cohen's kappa. To calculate this kappa statistic, we use the Fleiss kappa. The Fleiss kappa extends the Cohen's kappa to three or more raters. *See* Joseph L. Fleiss, *Measuring Nominal Scale Agreement Among Many Raters*, 76 PSYCH. BULL. 378, 378–79 (1971).

136. Coders were told that a decision contained a disagreement, but not which variable(s) there was disagreement on.

137. *See supra* note 89 and accompanying text (describing Landis and Koch's benchmark scale).

error 0.04). Appendix B contains cross-tabulation for both rounds of coding and kappa calculations.

We then merged this 2012–2019 dataset with the data from 1997–2012, with duplicate decisions from the overlapping period dropped, to create the final dataset. For the analysis below, we excluded appeals arising from the U.S. Patent and Trademark Office ("USPTO)" (such as in an appeal from a denial of a patent application), interferences, and appeals from the grant or denial of a preliminary injunction because the primary focus of the analysis was on appeals in patent infringement cases.

Finally, while our focus was not on appeals from the USPTO, peer commentators suggested that we need to include data on appeals from the USPTO as a reference point. To provide this comparison, we drew on data from the *Compendium of Federal Circuit Decisions* to obtain numbers of decisions coded in the *Compendium* as involving a § 103 issue. In order to independently assess the reliability of this data, we compared the Issue_103 coding from the *Compendium* to the opinions identified in the 1997–2019 dataset.[138] Table 14 summarizes the 96% agreement between the two ways of capturing whether an opinion involved a § 103 issue.

TABLE 11: COMPARISON DATA ON § 103 CODING

Fed. Cir. Appeals involving § 103 issue	Compendium of Federal Circuit Decisions		
	No	Yes	Total
No	1722	19	1741
Yes	61	262	323
Total	1783	281	2064

Based on this comparison, we concluded that the Issue 103 coding from the *Compendium* provided a reasonable measure of the number of Federal Circuit decisions involving a § 103 issue for opinions. This provided comparative data for appeals arising from the USPTO.

A reminder about selection effects: In reading the data below, it is important to keep in mind selection effects and population biases.[139] The data described below comprises Federal Circuit decisions involving appeals of § 103 issues that arose from the district courts. This context means that it is subject to multiple selective pressures: First, there must have been an infringement case filed, there must be a final decision involving a § 103 issue at the district court,

---

138. This comparison involving matching the decisions in the 1997–2019 dataset with the records in the *Compendium*. Because the *Compendium* only contains Federal Circuit decisions starting in 2004, in practice this comparison only reflects decisions from the period 2004–2019.

139. Petherbridge & Wagner, *supra* note 26, at 2070–71; Lunney, *E-Obviousness*, *supra* note 26, at 374; Schwartz & Kesan, *supra* note 11, at 434–35.

the losing party must have appealed, the parties must not have settled the dispute (which sometimes happens even after appeal), and the losing party must have addressed the § 103 issue in its briefs.[140] In addition, the number of decisions in a year is not large, so caution is warranted in reading-in conclusions based on a single year or two.[141] This is true even though the goal of this project was to collect population-level data: that is, all Federal Circuit decisions involving § 103 that arose from the district courts.

B. SECTION 103 FEDERAL CIRCUIT DECISIONS OVER TIME

For appeals arising from the district courts, the average number of decisions on an annual basis involving § 103 was the same for the post-2019 period as compared to the five-year period immediately after *KSR*. We found 140 opinions and 66 Rule 36 affirmances involving a § 103 issue between May 1, 2012 and December 31, 2019 (an average of 27 decisions/year), as compared with 106 opinions and 27 Rule 36 affirmances in the five years described in the 2013 study (an average of 27 decisions/year). However, closer inspection reveals some striking trends.

First, the number of decisions involving § 103 has not remained constant. Figure 1 shows the number of Federal Circuit decisions involving § 103 in appeals arising from the district courts between 1997 and 2019, broken up by the number of opinions and Rule 36 summary affirmances.

---

140. For a variety of reasons, parties may elect not to appeal issues on which they lose in a lower tribunal. For example, a patent owner may have had two patents invalidated at the district court but for strategic reasons chosen to appeal on only one of those patents. Or an accused infringer may have lost on both infringement and invalidity but decided to only challenge the infringement issue due to its perception of the strength of that issue.

141. For previous discussions of population biases and selection effects, see Rantanen, *Judicial Opinions*, *supra* note 9.

FIGURE 1: NUMBER OF FEDERAL CIRCUIT DECISIONS INVOLVING § 103 IN APPEALS
ARISING FROM DISTRICT COURTS, 1997–2019
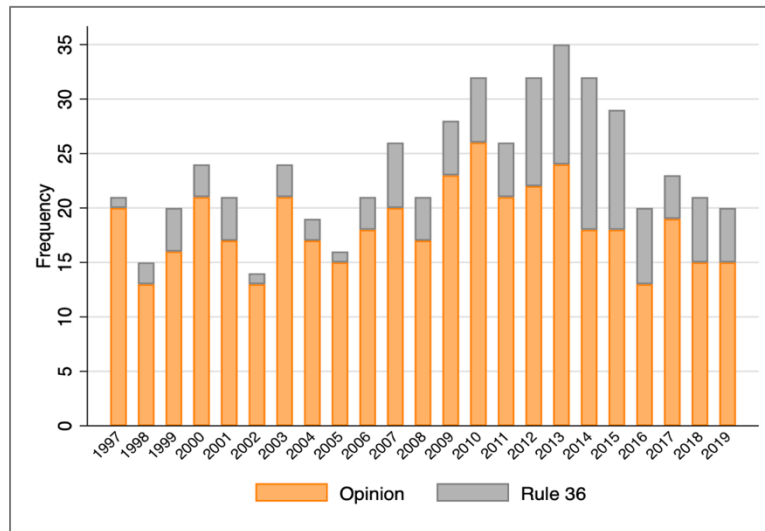
While the average number of decisions between 2007 and 2012 was the same as the average annual number of decisions between 2012 and 2019, a general upward slope characterizes the period from 2007 to 2013 and a general downward slope from 2013 to 2019. The number of Federal Circuit decisions involving § 103 issues appealed from district court decisions peaked around 2013, with over thirty decisions a year between 2012 and 2014, followed by noticeably fewer decisions in recent years.[142] In addition, heightened use of Rule 36 summary affirmances relative to written opinions characterize the period from 2012 to 2016. The court was using the summary affirmance mechanism for § 103 determinations arising from the district court more frequently during this time period. During this period, the Federal Circuit's use of Rule 36 summary affirmances in appeals arising from the district courts has remained relatively constant.[143]

In contrast, the number of decisions involving § 103 that arose from the USPTO increased dramatically over this time period. Table 15 shows the relative

---

142. This pattern is consistent even when the total number of decisions in appeals arising from the district courts on any issue is taken into consideration. See Lunney, *E-Obviousness*, *supra* note 26 for an explanation of why this can matter. To make this comparison, we compared the data from Figure 1 with the total numbers of opinions and Rule 36 affirmances each year from the *Compendium*. As with the absolute numbers, the relative frequency of Rule 36 decisions involving nonobviousness to all Rule 36 decisions was much higher for the years 2009–2013 than for 2014–2019.

143. *See* Jason Rantanen, *Federal Circuit Statistics – 2020 Edition*, PATENTLYO (Jan. 4, 2021), https://patentlyo.com/patent/2021/01/federal-circuit-statistics.html (drawing on data from the *Compendium* to show the relative frequency of Rule 36 summary affirmances and written opinions).

number of opinions (these do not include Rule 36 summary affirmances) involving a § 103 issue as coded in the *Compendium of Federal Circuit Decisions* for appeals arising from the district courts and USPTO.

TABLE 12: FEDERAL CIRCUIT DECISIONS IN APPEALS ARISING FROM THE DISTRICT COURTS AND USPTO THAT INVOLVED § 103.

	Origin of Case		
Year	District Court	USPTO	Total
2004	3	2	5
2005	16	5	21
2006	21	3	24
2007	18	4	22
2008	23	6	29
2009	29	13	42
2010	24	10	34
2011	21	15	36
2012	27	17	44
2013	24	18	42
2014	20	18	38
2015	24	30	54
2016	15	56	71
2017	24	69	93
2018	16	69	85
2019	18	68	86
Total	323	403	726

C. OBVIOUSNESS OUTCOMES

Another perspective on nonobviousness decisions at the Federal Circuit is to examine them in terms of final determinations.[144] Between 2012 and 2019, final outcomes on § 103 at the Federal Circuit in appeals from district court determinations were almost a fifty-fifty split: ninety-seven decisions in which the court reached a final determination of "nonobvious" and ninety-three in

---

144. Final determinations of nonobviousness reflect only decisions in which the court made a final determination; for example, they do not include outcomes of "no final determination." In addition, for purposes of this analysis we have not included the small number of "mixed" decisions. *See, e.g.*, Rantanen*, Obviousness Jurisprudence*, *supra* note 8; Holte & Sichelman, *supra* note 8.

which it reached a final determination of "obvious."[145] However, as Figure 2 shows, most of the "obvious" decisions were from the period 2012–2015, while the last few years have seen relatively more outcomes of "nonobvious." Even as the number of Federal Circuit decisions in appeals arising from the district courts has fallen, so too has the frequency of outcomes of "obvious."

FIGURE 1: OBVIOUS VS. NONOBVIOUS OUTCOMES AT THE FEDERAL CIRCUIT, 1997–2019
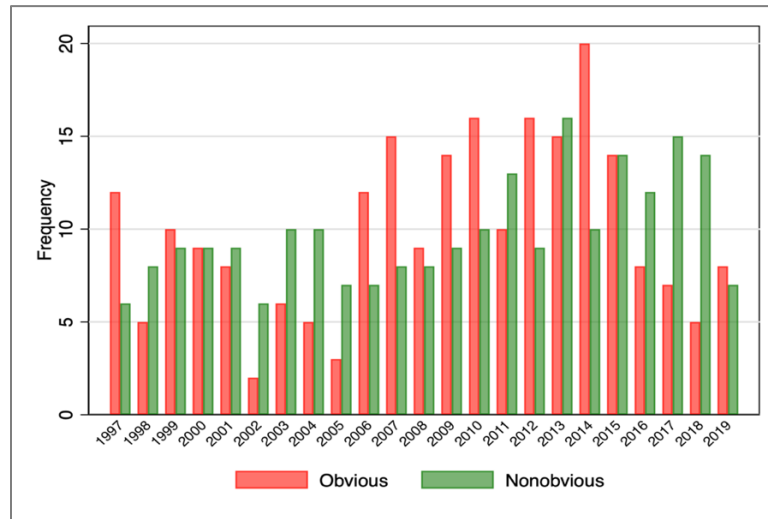
Figure 3 presents a twenty-five-record rolling average graph. This is a moving average of the percentage of times in which the Federal Circuit has reached a conclusion of "obvious" in a final determination in a moving set of twenty-five decisions, beginning with decisions 1–25, then the average for 2–26, etc.[146]

---

145. This analysis is limited to only appeals in which the Federal Circuit reached a final determination on the issue of obviousness. It does not include 16 decisions for which the outcome was coded as "No Final Determination." It also does not include the small number (6) of outcomes coded as "Mixed."

146. Keep in mind that even though years are provided on the X-axis for reference, a moving average shows an average of the record units and thus is not consistent over time. For example, there were more § 103 decisions in 2014 than in 2019. Thus, this chart does not reflect the lower number of § 103 decisions in appeals arising from the district courts in recent years. In addition, the last 25 averages contain successively fewer observations, so there is a potential for data artifacts at the end of the graph.

FIGURE 2: 25-RECORD TRAILING AVERAGE OF FEDERAL CIRCUIT
OBVIOUSNESS DECISIONS, 1996–2019
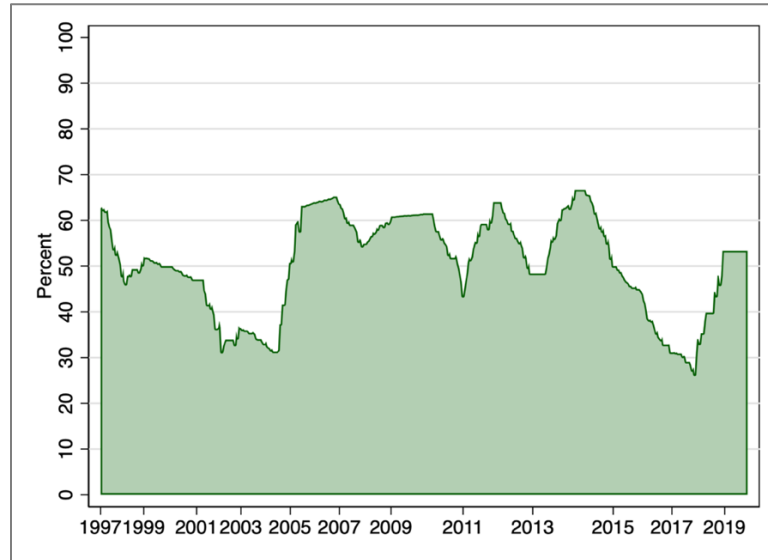
D. PROCEDURAL POSTURE AND OUTCOMES

Until recently, appeals from summary judgment made up a substantial portion of Federal Circuit decisions involving § 103. However, over the past few years the number of decisions involving appeals from summary judgment has declined sharply. Most disputes involving § 103 that have resulted in a Federal Circuit decision have been appeals from full trials on the merits, not summary judgment. These are patents for which there is a real question of the merits of the § 103 issue.

Figure 4 shows the relative proportions of procedural postures for each year. When reading Figure 4, keep in mind that the number of decisions varies for each year.

FIGURE 3: PROCEDURAL POSTURE OF § 103 DECISIONS IN APPEALS
ARISING FROM DISTRICT COURTS, 1997–2019
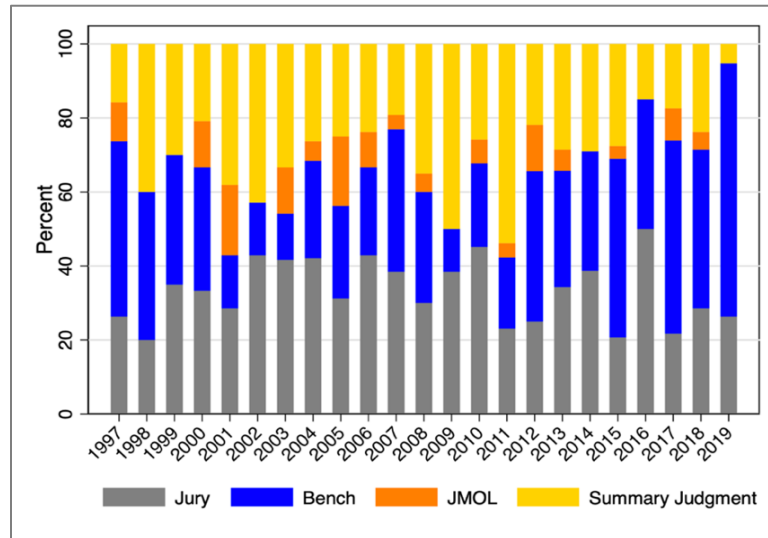
As shown below, more often than not, in recent years these factfinders have concluded that the patents in suit are not obvious.

FIGURE 4: DISTRICT COURT OUTCOMES REVIEWED AT THE FEDERAL CIRCUIT
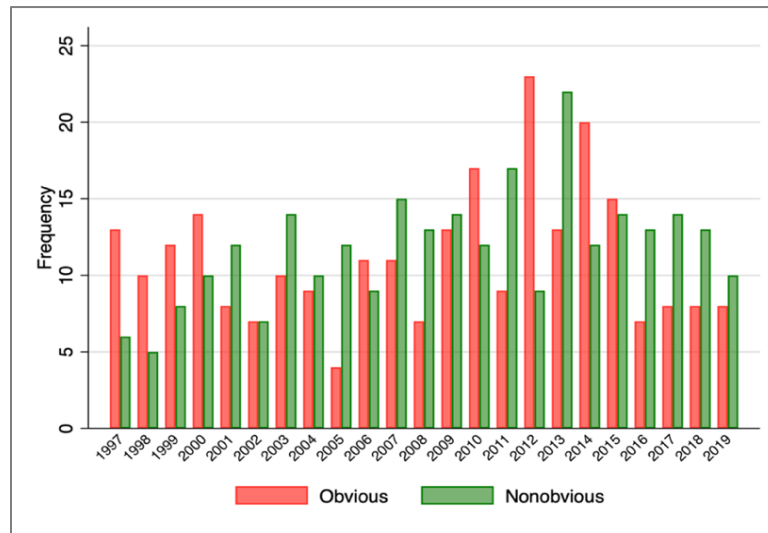
Figure 5 shows that while many of the district court decisions the Federal Circuit reviewed during the years following *KSR* were findings of "obvious," in recent years the Federal Circuit has reviewed fewer district court outcomes of

"obvious" even as the number of outcomes of "nonobvious" has remained about the same.[147]

E. APPELLATE DISPOSITIONS

We also observe that the affirmance rate for district court decisions involving § 103 continues to be high—higher even than pre-*KSR*.[148] Figure 6 reports data on complete affirmances as a function of total Federal Circuit decisions—that is, the denominator includes affirmances, reversals, vacates, and the rare "mixed" disposition in which the district court is affirmed on at least one § 103 decision and reversed on another. Only "affirmances" are counted in the numerator.

FIGURE 5: PERCENTAGE OF DISTRICT COURT DECISIONS INVOLVING § 103 THE FEDERAL CIRCUIT AFFIRMED
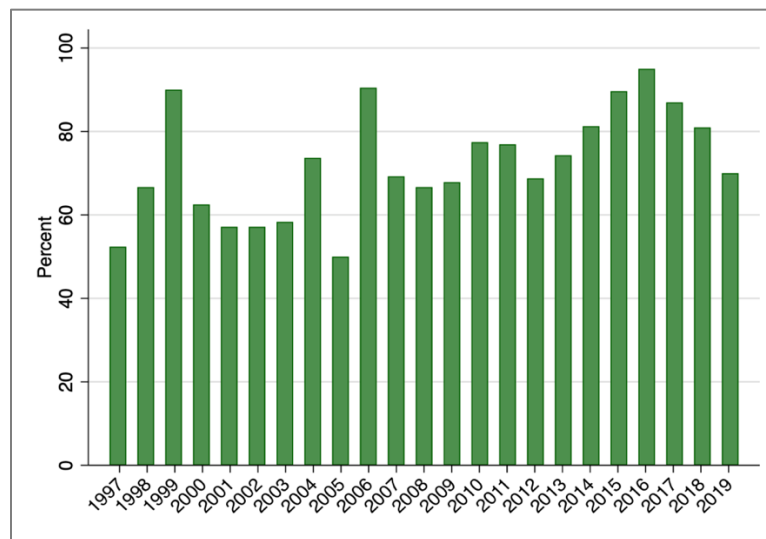
Not only have affirmances been high in recent years, but they involve a shift from the period immediately following *KSR* in *what* is being affirmed. Table 16 compares three time periods: Federal Circuit dispositions between

---

147. *See* Appendix A.

148. Nonobviousness rates are not unique in having a high affirmance rate. *See* Mark A. Lemley & Samantha Zyontz, *Does Alice Target Patent Trolls?*, 18 J. EMPIRICAL LEGAL STUD. 47, 76 (2021) (finding an affirmance rate of around 90% in appeals involving patent eligible subject matter).

January 1, 1997 and June 26, 2006;[149] between April 30, 2007 and April 30, 2012;[150] and since May 1, 2012.

TABLE 13: AFFIRMANCE RATES FOR DISTRICT COURT DETERMINATIONS OF OBVIOUS AND NONOBVIOUS

	1/1/1997– 6/26/2006	4/30/2007– 4/30/2012	5/1/2012– 12/31/2019
Obvious	56%	82%	78%
Nonobvious	71%	65%	83%

Surprisingly, while the period prior to April 30, 2012, is consistent with Table 4 in Rantanen (2013),[151] reflecting an increase in Federal Circuit affirmances of district court determinations of "obvious" following *KSR*, since 2012 there has been a substantial rise in affirmances of district court determinations of "nonobvious." Indeed, since 2015, the affirmance rate for these determinations has been even higher: 89% of all district court determinations of "nonobvious" since January 1, 2015, have been affirmed.[152] Even as the court has been reviewing a higher proportion of district court determinations of "nonobvious," it has been affirming those determinations at a higher rate. This presents a contrast with the immediate aftermath of *KSR* when only district court outcomes of *obvious* were being affirmed at a much higher rate.

F. ANALYSIS OF FEDERAL CIRCUIT DECISIONS INVOLVING § 103

What might be some explanations for these shifts—particularly the decline in the frequency of appellate decisions involving § 103, the increased affirmance rate, and the high frequency of disputed patents being held nonobvious?

One possibility is that litigating parties adapted to the new standard of *KSR*.[153] While the disputes for the first few years after *KSR* may have reflected lawsuits (and appeals) filed before *KSR*, more recent disputes may reflect decisions made in the new normal—the world in which the higher threshold applies. Those cases would thus be closer to the indeterminate boundary between obvious and nonobvious (as opposed to being on the "obvious" side of the post-*KSR* standard).[154] As parties settle the more determinate cases, the consequence

149. The date on which certiorari was granted in KSR Int'l. Co. v. Teleflex, Inc., 548 U.S. 902 (2006).

150. For example, from the date the Court issued its opinion in *KSR* until the end of Rantanen (2013), five years after the Court's opinion.

151. *See* Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 744 tbl.4. Note that several of the decisions from Holte & Sichelman (2019) that were added to the Rantanen (2013) dataset were Rule 36 summary affirmances, thus increasing the overall affirmance rate slightly.

152. This constitutes 57 out of 64 decisions. For reference, the affirmance rate for district court determinations of obvious has been 83% (38/46) since 2015.

153. *See, e.g.*, Lunney & Johnson, *supra* note 8, at 76–79.

154. *See* Rantanen, *Obviousness Jurisprudence*, *supra* note 8, at 764.

708 HASTINGS LAW JOURNAL Vol. 73:3

would be a return to what Glynn Lunney refers to as "the underlying equilibrium," which Lunney calculates at around a 30% patent owner win rate.[155]

There are a few reasons to be skeptical of this explanation, however. The settlement theory on which we base this explanation is a theory of *disputes*, not *issues*.[156] Unless a § 103 question is essentially the sole issue involved in a litigated dispute, there are other considerations that will affect which disputes parties settle and which they litigate. The consequence is that equilibrium-based strategic settlement selection theories such as Priest-Klein do not provide a theory for selection for outcomes on individual issues.[157] Another reason to be skeptical that the changes in § 103 decisions are due to an equilibrium-based selection theory is that the decline in the number of decisions involving § 103 and increase in the nonobvious rate began around eight years after the Court's decision in *KSR*. It was also sudden and dramatic. Even accounting for the fact that disputes can take time to percolate up to a Federal Circuit decision, that is a long time for the shift to manifest. Selection theories such as Priest-Klein assume that most disputes are settled, thus resulting in substantial selection for those disputes that are litigated. However, the data in Holte and Sichelman (2019) indicates that most substantive obviousness determinations at the district courts are appealed to the Federal Circuit, limiting the effect that selection at the appellate level could have.[158]

Another explanation is that these trends reflect changes in the substance and process of patent law itself. Three major events have dominated patent law over the last ten years: Congress's passage of the America Invents Act in 2011, which changed substantive elements of patent law; the creation and rise of inter partes review proceedings ("IPRs") at the USPTO; and the emergence of patent eligible subject as an explicit limit on patentability.[159] Of these, we think the latter two had the greatest potential to affect appellate decisions involving § 103.

There are several reasons IPRs may lead to fewer § 103 appeals from the district courts. IPRs are significantly faster than district court litigation. Once

---

155. Glynn S. Lunney, Jr., *Patent Law, the Federal Circuit, and the Supreme Court: A Quiet Revolution*, 11 SUP. CT. ECON. REV. 1, 10, 14 (2003) [hereinafter Lunney, *A Quiet Revolution*].

156. *See* Lunney & Johnson, *supra* note 8, at 76–78; Jason Rantanen, *Why Priest-Klein Cannot Apply to Individual Issues in Patent Cases* 2 (Univ. of Iowa Coll. of L. Legal Stud. Rsch. Paper Series, Paper No. 12-15, 2013).

157. *See* Rantanen, *supra* note 156, at 3–7 (discussing the application of the Priest-Klein selection effect theory to multi-issue appeals).

158. *See* Holte & Sichelman, *supra* note 8, at 136 (reporting 319 district court opinions and 192 Federal Circuit opinions that had actual obviousness determinations between January 1, 2003 and December 31, 2013). There is certainly still selection at the district court level. *See* John R. Allison, Mark A. Lemley & David L. Schwartz, *Our Divided Patent System*, 82 UNIV. CHI. L. REV. 1073, 1080 (2015). However, here the hypothesized selection is from the district court to the appellate court rather than selection from all possible disputes.

159. *See* DUFFY & MERGES, *supra* note 23.

the USPTO grants a petition for IPR, the Patent Trial and Appeal Board must issue a final determination within, at most, eighteen months.[160] A resolution in this time period is very expeditious as compared to the many years patent litigation disputes last in district courts.[161] It can also mean that the dispute may end with a decision that a patent is invalid well before it reaches a final judgment at the district court.[162] In addition, any district court action filed on or after the petition date challenging the validity of a claim of the patent is automatically stayed for the pendency of the IPR.[163] District courts also have inherent power to control their dockets and therefore can, and often do, stay district court proceeding in favor of subsequently filed IPRs.[164]

Further, a petition for IPR will not be granted unless the petitioner has shown a reasonable likelihood of success on at least one of the challenged claims, which may indicate the possibility of a favorable outcome.[165] Once the petition is granted, the petitioner is only required to prove invalidity by a preponderance of the evidence, as opposed to the clear and convincing evidence standard in district courts.[166] A substantial number of § 103 disputes that *could* happen in the district courts may simply be resolved in IPRs,[167] and with fewer district court rulings on nonobviousness, there are fewer appeals to the Federal Circuit involving the same.

Another explanation for the changes we observe in nonobviousness decisions at the Federal Circuit is the tightening of subject eligibility following four major Supreme Court decisions between 2010 and 2014.[168] For many years, courts loosely interpreted patent eligible subject matter under § 101[169] to encompass nearly anything that was a "process, machine, manufacture, or

---

160. 35 U.S.C. § 316(a)(11) (2018); Christian Helmers & Brian Love, *The Effect of New Information on Patent Litigation: Evidence from U.S. Inter Partes Review* 7 (June 29, 2021), https://papers.ssrn.com/sol3/ papers.cfm?abstract_id=3720709; Brian J. Love, Shawn P. Miller & Shawn Ambwani, *Determinants of Patent Quality: Evidence from Inter Partes Review Proceedings*, 90 UNIV. COLO. L. REV. 67, 103–04 (2019).

161. Helmers & Love, *supra* note 160, at 8 (showing that, on average, it takes three to four years after a complaint is filed for an appellate decision to issue).

162. Love et al., *supra* note 160, at 101 n.151 ("According to LexMachina.com, the median time to summary judgment in patent cases filed since 2000 is about 660 days.").

163. 35 U.S.C. § 315 (2018); Helmers & Love, *supra* note 160, at 3.

164. *See* Jonathan Stroud, Linda Thayer & Jeffrey C. Totten, *Stay Awhile: The Evolving Law of District Court Stays in Light of* Inter Partes *Review, Post-Grant Review, and Covered Business Method Post-Grant Review*, 11 BUFF. INTELL. PROP. L.J. 226, 237–38 (2015).

165. 35 U.S.C. § 314(a); Helmers & Love, *supra* note 160, at 6; Love et al., *supra* note 160, at 98–99.

166. 35 U.S.C. § 316(e); Love et al., *supra* note 160, at 103.

167. Love et al., *supra* note 160, at 96 (finding that almost 6,500 petitions for IPR have been filed since September 2012, which exceeds the total number of patent cases filed in all district courts except the Eastern District of Texas during the same time).

168. Alice Corp. v. CLS Bank Int'l, 573 U.S. 208, 217–18 (2014); Ass'n for Molecular Pathology v. Myriad Genetics, Inc., 569 U.S. 576, 589–90 (2013); Bilski v. Kappos, 561 U.S. 593, 609 (2010); Mayo Collaborative Servs. v. Prometheus Lab'ys, Inc., 566 U.S. 66, 71–72 (2012).

169. 35 U.S.C. § 101.

composition of matter."[170] Critics at the time also believed patents being issued during this time were not only directed to ineligible subject matter but also invalid as obvious.[171] But under the strict application of the TSM test prior to *KSR*, invalidating patents as obvious was quite difficult.[172] So for many years, patents that should have otherwise been invalid were easily satisfying the relaxed § 101 standard, and litigants often failed to meet the burden for invalidation under § 103 with the strict application of the TSM test.

The Supreme Court's 2007 decision in *KSR* gave the courts—particularly the district courts—increased ability to rely on the obviousness requirement to resolve disputes.[173] The decision also brought attention to the nonobviousness requirement, invigorating accused infringers to raise this ground of invalidity.[174] This led, initially at least, to a surge in decisions involving nonobviousness.[175] That decision, however, was followed by *Alice Corp. v. CLS Bank International*, *Mayo Collaborative Services. v. Prometheus Laboratories, Inc*, and *Bilski v. Kappos*[176] which offered accused infringers an alternative—and sometimes even more potent—way to challenge patents that may have had some of the same underlying issues that could give rise to obviousness problems.[177] Indeed, since the shift to using the *Alice/Mayo* framework to assess patentability, courts have found patents were directed to ineligible subjects in a majority of cases involving § 101.[178] The strict application of *Alice/Mayo* allows courts to quickly dispose of patents that are clearly invalid, even if doing so reaches into the territory of questions that are better addressed under § 103.[179] But courts are inclined to decide cases under § 101 because patent eligibility is purely a question of law that can be addressed on a motion to dismiss or judgment on the pleadings—as compared to § 103, which at the earliest can be dismissed at the summary judgment stage.[180] Therefore, what we are observing may be, as some scholars

---

170. *See* Paul R. Gugliuzza, *Quick Decisions in Patent Cases*, 106 GEO. L.J. 619, 649 (2018); Dennis Crouch & Robert P. Merges, *Operating Efficiently Post-*Bilski *by Ordering Patent Doctrine Decision-Making*, 25 BERKELEY TECH. L.J. 1673, 1677 (2010).

171. *See* Gugliuzza, *supra* note 170, at 650.

172. *See id.*

173. *See id.* at 651; Crouch & Merges, *supra* note 170, at 1685–86 (suggesting that obviousness should be considered prior to alternate grounds for invalidity).

174. Rantanen, *supra* note 8, at 739 (2013); Holte & Sichelman, *supra* note 8, at 139–40.

175. Holte & Sichelman, *supra* note 8, at 141.

176. Alice Corp. v. CLS Bank Int'l, 573 U.S. 208 (2014); Mayo Collaborative Servs. v. Prometheus Lab'ys, Inc., 566 U.S. 66 (2012); Bilski v. Kappos, 561 U.S. 593 (2010).

177. *See* Gugliuzza, *supra* note 170, at 652.

178. *See* Paul R. Gugliuzza & Mark A. Lemley, *Can a Court Change the Law by Saying Nothing?*, 71 VAND. L. REV. 765, 774 (2018).

179. *See id.* at 777; *see also* Gugliuzza, *supra* note 170, at 655.

180. *See* Gugliuzza, *supra* note 170, at 651; *see also* Gugliuzza & Lemley, *supra* note 178, at 777–78; Paul Gugliuzza, *The Procedure of Patent Eligibility*, 97 TEX. L. REV. 571, 614–616 (2019).

have suggested, patent eligibility serving as a rough filter for other questions of patentability—particularly § 103.[181]

A final explanation for what we observe draws on the first *en banc* decision by the Federal Circuit since 1994[182] to address § 103: the court's *Apple v. Samsung* opinion in 2016.[183] At the district court, the jury found the '721 and '172 patents were nonobvious and accordingly denied Samsung's requests for a JMOL on the issue.[184] Samsung appealed both JMOL denials, which the Federal Circuit affirmed in an 8-3 decision.[185] Because obviousness is a question of law based on underlying findings of fact, the Federal Circuit's analysis focused on whether the jury had sufficient evidence to support their findings of nonobviousness.[186] The jury's findings, and the Federal Circuit's analysis, regarding the '721 patent is interesting in this respect.

At trial, Samsung presented evidence of two prior art references that together taught every element of the disputed claim, yet the jury still found the claim was nonobvious.[187] In reviewing the evidence, the Federal Circuit focused not only on the *Graham*[188] factors but also put a significant emphasis on the secondary considerations of nonobviousness.[189] In particular, Apple presented large amounts of evidence on the industry praise, copying, commercial success, and long-felt unresolved need.[190] The outcome here could suggest that the Federal Circuit is reluctant to overcome jury findings, but also could be evidence of where the judges on the Federal Circuit stand on obviousness.

Another possibility is that *Apple* served as a signal for the judges' normative views on § 103. The decision itself was 8-3 in favor of nonobviousness (Judge Moore authoring the majority opinion), with Chief Judge Prost and Judges Reyna and Dyk all dissenting.[191] Looking at the period from 2013 to 2019, those three judges had the second, third, and fourth-highest ratio of obvious to nonobvious outcomes when authoring the majority

---

181. *See* Gugliuzza, *supra* note 170, at 655.

182. That opinion was In Re Donaldson Co. Inc., 16 F.3d 1189, 1194 (Fed. Cir. 1994) (en banc); *see also* In re Dillon., 919 F.2d 688, 198 (Fed. Cir. 1990) (en banc). Thanks to David Taylor and the dataset maintained by the Federal Circuit Blog for data on Federal Circuit en banc decisions. *En Banc Cases*, FEDCIRCUITBLOG, https://fedcircuitblog.com/en-banc/cases (last visited Mar. 18, 2022).

183. Apple Inc. v. Samsung Elecs. Co., 839 F.3d 1034 (Fed. Cir. 2016) (en banc).

184. *Id.* at 1038.

185. *Id.* at 1038–39.

186. *See id.* at 1047.

187. *Id.* at 1050–52.

188. Graham v. John Deere, 383 U.S. 1, 17–18 (1966). ("Under § 103, the scope and content of the prior art are to be determined; differences between the prior art and the claims are to be ascertained; and the level of ordinary skill in the pertinent art resolved. Against this background, the obviousness or nonobviousness of the subject matter is determined. Such secondary considerations as commercial success, long felt but unsolved needs, failure of others, etc., might be utilized to give light to the circumstances surrounding the origin of the subject matter sought to be patented.").

189. *Apple*, 839 F.3d at 1052–53.

190. *See id.* at 1053–57

191. *See id.* at 1034.

opinion.[192] Only Judge Lourie has a higher rate (47% of the opinions that Judge Lourie authored had an outcome that the patents are obvious). All the other judges in the majority had ratios between 0–25% for majority-authored opinions. *Apple* may be less notable for legal pronouncements about obviousness law than what it says about the Federal Circuit judges' normative schema on obviousness.[193]

To be clear, we are not suggesting that the Federal Circuit judges are close-minded on the issue of obviousness, nor that they have pre-decided the issue in a case. Every active judge has authored at least one opinion in which the outcome was that the patent was obvious and one in which the outcome was that the patent was not obvious (with one exception who has only authored three opinions involving the issue).[194] Overall ratios of obviousness to nonobviousness decisions may still be instructive as a proxy for a judge's general attitudes regarding § 103.

CONCLUSION

The data suggest we are in a new "cycle of obviousness," to use Holte and Sichelman's term, one in which the outcome of most appeals from district court decisions is the patents-in-suit are nonobvious. We are also in a period of greater affirmance of district court decisions on obviousness. However, this shift may also result from the use of IPRs and invalidity challenges based on § 101 that effectively weed out patents that are also unlikely to survive a nonobviousness analysis in a district court proceeding.

The return to a new normal in which the Federal Circuit rejects § 103 challenges in district court appeals is consistent with long-term historical studies of the Federal Circuit that have found it holds in favor of the validity of patents. As Glynn Lunney, Jr. and others have observed, the Federal Circuit era has been one in which the court has been reluctant to invalidate patents, especially on § 103 grounds.[195] While *KSR* had an effect at the district courts, and to some extent on the Federal Circuit, that effect has since faded, and now other factors overshadow it.

This does not mean that patent owners *win* at the Federal Circuit: that is a different issue from whether patents are invalid based on § 103. As Lunney and

---

192. Chief Judge Prost: 8 out 18 (44%); Judge Reyna: 3 out of 7 (43%); Judge Dyk: 3 out of 8 (38%). For additional details, see Appendix D.

193. Melissa F. Wasserman & Jonathan D. Slack, *Is Too Much Specialization a Bad Thing? Specialization in Specialized Courts*, 115 NW. U. L. REV. 1405, 1419 ("[O]pinion specialization on specialized courts increases the chances that doctrine may reflect idiosyncratic preferences of a few judges.").

194. *See* Appendix D.

195. *See* Lunney, *A Quiet Revolution*, *supra* note 155, at 15, 26; Lunney, *E-Obviousness*, *supra* note 26, at 374–75; Matthew D. Henry & John L. Turner, *The Court of Appeals for the Federal Circuit's Impact on Patent Litigation*, 35 J. LEGAL STUD. 85, 112 (2006).

others have observed, often the Federal Circuit concludes a patent is valid but not infringed.[196]

If the new normal is primarily being driven by IPR proceedings and the rise of § 101-based challenges, it may not be as meaningful of a return to pre-*KSR* trends as it might appear. Since a sizeable portion of IPR proceedings involves challenges based on § 103, a decline in appeals from district courts that involve § 103 does not say all that much about the relative importance of the doctrine.[197] A detailed examination of appeals from IPRs, however, is for another article.

More broadly, as the analysis in Part I demonstrates, there can be a high level of consistency between independent coders of the same judicial opinions. This finding helps to validate macro-level studies of judicial opinions. However, our analysis reveals the importance of decisions about whether to include individual judicial decisions within the study population. This suggests that scholars should give greater attention to this stage of studies of judicial opinions. Well-developed and documented instructions to guide decisionmakers in the inclusion determination should be a primary focus for all researchers.

---

196. *See* Lunney, *A Quiet Revolution*, *supra* note 155, at 14.

197. Love et al., *supra* note 160, at 96 (finding that almost 6,500 petitions for IPR have been filed since September 2012, which exceeds the total number of patent cases filed in all district courts except the Eastern District of Texas during the same time period).

APPENDIX A

TABLE A1: HOLTE AND SICHELMAN AND RANTANEN
DATASET INTER-RATER RELIABILITY

Variable	Kappa Coefficient	Standard Error	Probabilistic Benchmark Interval
Procedural Posture	0.72	0.05	0.60 – 0.80
Lower Court Result	0.79	0.04	0.60 – 0.80
Federal Circuit Result	0.82	0.05	0.60 – 0.80
Federal Circuit Disposition	0.87	0.05	0.60 – 0.80
Some Use of TSM	0.38	0.09	0.20 – 0.40
TSM Formal	0.67	0.09	0.40 – 0.60
Common Sense	0.80	0.10	0.60 – 0.80

APPENDIX B

TABLE B1: PRE-REVIEW INTER-RATER RELIABILITY

Variable	Kappa Coefficient	Standard Error	Probabilistic Benchmark Interval
Procedural Posture	0.75	0.03	0.60 – 0.80
Lower Court Result	0.85	0.03	0.60 – 0.80
Federal Circuit Result	0.79	0.03	0.60 – 0.80
Federal Circuit Disposition	0.84	0.04	0.60 – 0.80

TABLE B2: PRE-REVIEW PROCEDURAL POSTURE

Coder 1 Procedural Posture	Coder 2 Procedural Posture						
	Jury	Bench	JMOL	SJ	PI	PTO	Total
Jury	47	8	7	2	0	0	64
Bench	1	85	0	0	0	1	87
SJ	3	2	7	48	0	1	61
PI	0	0	0	0	7	0	7
PTO	0	0	0	0	0	3	3
Other	1	3	0	2	3	0	9
Total	52	98	14	52	10	5	231

TABLE B3: PRE-REVIEW LOWER COURT RESULT

Coder 1 Lower Court Result	Coder 2 Lower Court Result				
	Obvious	Nonobvious	Both	No Final Determination	Total
Obvious	104	3	0	2	109
Nonobvious	3	106	0	3	112
Both	2	0	1	3	6
No Final Determination	2	1	0	1	4
Total	111	110	1	9	231

TABLE B4: PRE-REVIEW FEDERAL CIRCUIT RESULT

Coder 1 Federal Circuit Result	Coder 2 Federal Circuit Result				
	Obvious	Nonobvious	Both	No Final Determination	Total
Obvious	96	4	0	2	102
Nonobvious	4	85	0	8	97
Both	3	1	1	4	9
No Final Determination	0	4	0	19	23
Total	103	94	1	33	231

TABLE B5: PRE-REVIEW FEDERAL CIRCUIT DISPOSITION

Coder 1 Federal Circuit Disposition	Coder 2 Federal Circuit Disposition				
	Affirmed	Reversed	Vacated	Mixed	Total
Affirmed	176	0	0	7	183
Reversed	1	25	1	1	28
Vacated	0	0	11	2	13
Mixed	0	1	1	5	7
Total	177	26	13	15	231

TABLE B6: POST-REVIEW INTER-RATER RELIABILITY

Variable	Kappa Coefficient	Standard Error	Probabilistic Benchmark Interval
Procedural Posture	0.86	0.03	0.80 – 1.00
Lower Court Result	0.89	0.03	0.80 – 1.00
Federal Circuit Result	0.91	0.02	0.80 – 1.00
Federal Circuit Disposition	0.84	0.04	0.60 – 0.80

TABLE B7: POST-REVIEW PROCEDURAL POSTURE

Coder 1 Procedural Posture	Coder 2 Procedural Posture							
	Jury	Bench	JMOL	SJ	PI	PTO	Other	Total
Jury	56	2	6	0	0	0	0	64
Bench	0	87	0	0	0	1	0	88
JMOL	3	1	7	0	0	0	0	11
SJ	0	1	0	49	0	1	0	51
PI	0	0	0	0	6	0	0	6
PTO	0	0	0	0	0	3	0	3
Other	1	3	0	0	4	0	1	9
Total	60	94	13	49	10	5	1	232

TABLE B8: POST-REVIEW LOWER COURT RESULT

Coder 1 Lower Court Result	Coder 2 Lower Court Result				
	Obvious	Nonobvious	Both	No Final Determination	Total
Obvious	109	2	0	0	111
Nonobvious	1	105	0	0	106
Both	2	0	3	0	5
No Final Determination	4	4	1	1	10
Total	116	111	4	1	232

TABLE B9: POST-REVIEW FEDERAL CIRCUIT RESULT

Coder 1 Federal Circuit Result	Coder 2 Federal Circuit Result				
	Obvious	Nonobvious	Both	No Final Determination	Total
Obvious	99	1	0	1	101
Nonobvious	1	94	0	0	95
Both	2	1	5	0	8
No Final Determination	2	4	1	21	28
Total	104	100	6	22	232

TABLE B10: POST-REVIEW FEDERAL CIRCUIT DISPOSITION

Coder 1 Federal Circuit Disposition	Coder 2 Federal Circuit Disposition				
	Affirmed	Reversed	Vacated	Mixed	Total
Affirmed	179	0	0	1	180
Reversed	2	21	3	1	27
Vacated	0	0	11	2	13
Mixed	3	2	0	7	12
Total	184	23	14	11	232

APPENDIX C

TABLE C1: PROBIT ESTIMATES OF THE LIKELIHOOD THE FEDERAL CIRCUIT
FINDS OBVIOUSNESS POST-*KSR* GRANT OF CERT

	Original Data		Comparison Data	
	Holte & Sichelman (2019)	Rantane (2013)	Holte & Sichelman (2019)	Rantanen (2013)
Post-KSR Cert	0.22***	0.13**	0.29***	0.31***
	(0.08)	(0.06)	(0.10)	(0.10)
No. of Observations	190	290	125	120
Log-likelihood	-125.15	-198.01	-77.27	-74.13

Marginal effects reported with discrete change of indicator variables from 0 to 1.

* p<0.10, ** p<0.05, *** p<0.01

TABLE C2: PROBIT ESTIMATES OF THE LIKELIHOOD THE FEDERAL CIRCUIT
FINDS OBVIOUSNESS POST-*KSR* GRANT OF CERT

	Original Data		Comparison Data	
	Holte & Sichelman (2019)	Rantanen (2013)	Holte & Sichelman (2019)	Rantanen (2013)
Post-KSR Cert	0.19**	0.12*	0.25**	0.31***
	(0.08)	(0.06)	(0.10)	(0.11)
Bench	0.10	-0.19	-0.04	-0.38*
	(0.19)	(0.13)	(0.13)	(0.20)
Jury	-0.03	-0.21	-0.12	-0.39**
	(0.19)	(0.13)	(0.14)	(0.19)
JMOL	0.06	0.10		
	(0.18)	(0.16)		
SJ	0.38***	0.22*	0.33***	0.04
	(0.14)	(0.12)	(0.10)	(0.21)
No. of Observations	190	288	125	120
Log-likelihood	-114.4	-178.83	-68.59	-65.16

Marginal effects reported with discrete change of indicator variables from 0 to 1.

* p<0.10, ** p<0.05, *** p<0.01

TABLE C3: PROBIT ESTIMATES OF THE LIKELIHOOD THE FEDERAL CIRCUIT
FINDS OBVIOUSNESS POST-*KSR* GRANT OF CERT

	Original Data		Comparison Data	
	Holte & Sichelman (2019)	Rantanen (2013)	Holte & Sichelman (2019)	Rantanen (2013)
Post-KSR Cert	0.22***	0.13**	0.29***	0.31***
	(0.08)	(0.06)	(0.10)	(0.10)
No. of Observations	190	290	125	120
Log-likelihood	-125.15	-198.01	-77.27	-74.13

Marginal effects reported with discrete change of indicator variables from 0 to 1.

* p<0.10, ** p<0.05, *** p<0.01

TABLE C4.

	Original Data		Comparison Data	
	Holte & Sichelman (2019)	Rantanen (2013)	Holte & Sichelman (2019)	Rantanen (2013)
Post-KSR Cert	0.19**	0.12*	0.25**	0.31***
	(0.08)	(0.06)	(0.10)	(0.11)
Bench	0.10	-0.19	-0.04	-0.38*
	(0.19)	(0.13)	(0.13)	(0.20)
Jury	-0.03	-0.21	-0.12	-0.39**
	(0.19)	(0.13)	(0.14)	(0.19)
JMOL	0.06	0.10		
	(0.18)	(0.16)		
SJ	0.38***	0.22*	0.33***	0.04
	(0.14)	(0.12)	(0.10)	(0.21)
No. of Observations	190	288	125	120
Log-likelihood	-114.4	-178.83	-68.59	-65.16

Marginal effects reported with discrete change of indicator variables from 0 to 1.

* p<0.10, ** p<0.05, *** p<0.01

APPENDIX D

TABLE D1: OBVIOUSNESS DECISIONS BY AUTHORING JUDGE, 2013–2019

Judicial Opinion Author	CAFC Result				
	Obvious	Nonobvious	Mixed	No Final Determination	Total
Chen	2	4	1	2	9
Dyk	3	5	0	0	8
Hughes	1	3	0	0	4
Lourie	8	7	2	0	17
Moore	2	6	0	1	9
Newman	1	3	0	0	4
O'Malley	0	2	0	1	3
Prost	8	7	1	2	18
Reyna	3	3	0	1	7
Stoll	2	5	0	1	8
Taranto	3	6	0	0	9
Wallach	1	2	0	1	4
Total	34	53	4	9	100

***