Scientometrics (2021) 126:2903–2929
https://doi.org/10.1007/s11192-020-03849-z

Out of sight: patents that have never been cited

Neil Gandal[1] · **Michal Shur-Ofry**[2] · **Michael Crystal**[1] · **Royee Shilony**[2]

Received: 12 May 2020 / Published online: 28 February 2021

Abstract

Patent citations have become an acceptable proxy for inventions' quality. Our study offers the first systematic exploration of uncited patents. Analyzing data on all US patents issued between 1976 and 2008, we examine the ratio of uncited patents out of all patents granted each year. We find a robust pattern, consistent across technological fields, whereby the percentage of uncited patents declined between 1976 and the mid-1990s, but has been significantly increasing since then. We discuss policy implications of these findings and suggest that the ratio of uncited patents can serve as a complementary measure for evaluating the patent system.

Keywords Uncited patents · Patents · Patent citations · Patent quality · Innovation · Networks · Big data · Negative knowledge

Introduction

Patent registration generates huge data repositories that constitute an important source of knowledge about innovation. A growing body of literature in economics, business management, network science, and the legal field, suggests that analyses of patent data can provide ample information about innovation processes, about the traits of specific inventions, and about technological domains. This growing awareness, together with developments in big data analyses have led to an upsurge in the exploration of information stored in patent databases in the recent decades.[1]

A significant part of the research in this area is devoted to the study of patent citations, and specifically to the study of highly cited patents. The perception underlying this line of scholarship is that patent citations reflect technological relations between the citing and the

[1] For pioneering studies see, e.g., Schmookler, 1966; Griliches, 1984; Trajtenberg, 1990. For more recent studies see, e.g., Lanjouw and Schankerman, 2001; Allison et al. 2004; Érdi et al. 2013; Shur-Ofry, 2017; Ashtor, 2019; Pedraza-Farina and Whelan, 2020.

✉ Neil Gandal
gandal@tauex.tau.ac.il

Michal Shur-Ofry
michalshur@huji.ac.il

[1] Berglas School of Economics, Tel Aviv University, Dr George Wise St. 40, Tel Aviv-Yafo, Israel

[2] Hebrew University of Jerusalem Law Faculty, Jerusalem, Israel

cited patents (e.g. Trajtenberg 1990; Harhoff et al. 1999; Jaffe and Trajtenberg 2003; Hall et al. 2005). Therefore, a large number of follow-on citations received by a patent is perceived as an indication of the impact, technological quality, value, and even breakthrough nature, of the patented technology (e.g., Fleming, 2001; Dahlin and Behrens, 2005; Érdi et al. 2013; Arts and Veugelers 2015).[2]

Yet, there is another side to patents citations that can provide important information about innovation and about the patent system, but has been essentially ignored in the literature to date: patents that do not receive any citations. In this Article we term them "uncited patents". While some scholarship incidentally observed that patents without citations could be a proxy for low quality inventions,[3] no study has systematically focused on uncited patents. How many uncited patents are there? Does the percentage of uncited patents change over time, or across technological fields? Can we identify traits that characterize uncited patents? These and additional questions have been left unanswered in the literature to date. The lack of focus on uncited patents may be understandable, in light of the general human tendency to concentrate on positive information (in our case: successful patents that receive many citations), and ignore negative information.[4] Yet, if patent citations provide information about the quality and value of patented technologies, then studying patents without citations is necessary, in order to get a fuller picture about the patent system, and the way it fulfills its ultimate task: promoting valuable innovation.[5]

In this Article we aim to fill-in this gap, by providing the first systematic exploration of uncited patents. We analyze all U.S. patents granted between 1976 and 2008 (3,079,587 observations overall). First, we focus on the ratio of uncited patents out of all patents granted each year, and track the changes in this percentage during a period of more than three decades. Second, we inquire whether there are differences in uncited patents across technological fields, comparing, primarily, the fields of pharmaceuticals and software-related patents. Third, we inquire whether there is a link between an "uncited status" and specific patent traits. As part of the latter inquiry we examine a series of factors —including the number of backward citations a patent has, the number of inventors listed on the patent, the number of subclasses to which the patent belongs, the number of claims in the patent, and the degree of similarity between the uncited patents and the prior art which they cite—in order to explore whether there are differences between uncited and cited patents. We also explore the relations between an "uncited status" and the payment of patent renewal fees (that prevents the expiration of the patent before the end of the term).

Our findings reveal a robust pattern in which the ratio of uncited patents out of all granted patents decreased between 1976 and the mid-1990s, but has been significantly increasing since then. Graphically, then, the ratio of uncited patents across the period of

---

[2] In the analysis below we use the term "quality" or "patent quality" in a broad and expansive manner, to refer not only to a patent's technological quality but also to aspects such as its impact on subsequent technologies, or its value. We discuss the evidence on the connection quality and a large number of forward citations *infra*.

[3] *See, e.g.,* Arts and Veugelers, 2015:1218 ("[P]atents receiving a disproportionately large number of citations can be considered as breakthroughs while patent receiving no citations as failures").

[4] For a discussion of the general disregard of negative information in the innovation system, see Shur-Ofry, 2016.

[5] The role of the patent system as a major vehicle for promoting innovation traces back to the U.S. Constitution—See U.S. Const. Art. I, Sec. 8, Cl. 8. (empowering Congress to "promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Rights to their respective Writings and Discoveries").

our study forms a rough “U” shape. These findings are robust after controlling for patents' age and additional factors, and are consistent across technological fields. Since patent citations are perceived as an indication of technological quality, impact, and value, these findings are concerning.

Interestingly, we observe differences among technological fields in the ratio of uncited patents. Prominently, the percentage of uncited patents is significantly higher for pharmaceutical drug patents, in comparison to patents in the software and tech-related fields.[6]

We also find that uncited patents are *negatively* associated with the number of backward citations, the number of subclasses, and the number of claims in the patent.[7] In other words, large numbers of backward citations, subclasses, and claims increase a patent's chances of being cited. Because numbers of claims, subclasses and backward citations were recognized in the literature as positively related to high-quality patents, these latter findings indicate that on average, the quality of uncited patents may indeed be lower, in comparison to cited patents.[8] In addition, we find that uncited patents are *negatively associated* with the payment of patent renewal fees by the patent owners, relative to cited patents, a finding which provides another indication for lower quality.

The general picture emerging from our study is concerning. Out of all patents granted by the USPTO, the percentage of uncited patents across domains has been increasing significantly from 1996 through 2008, which is the last year in our study. We discuss possible explanations for this trend, and, while we cannot offer a definitive interpretation, we locate our findings in the context of the literature concerning “patent explosion” and patent quality.

More broadly, our systematic focus on uncited patents presents a new, additional measure to evaluate the shape of the patent system, and opens up a new research avenue in the analysis of patent data. We conclude by sketching, in broad strokes, additional directions for exploration that would shed further light on the negative information embedded in patent repositories, and complement the research on the positive aspects thereof. Such studies could provide us with a deeper and more nuanced understanding of the patent system, as well as the innovation processes it seeks to promote.

This Article proceeds as follows: We begin with the theoretical background on the scientific uses of patent citations as a proxy for processes, traits, and trends of innovation. We then proceeds to quickly review the scholarly discussion of patent quality that is relevant for the following analysis. The next section describes our dataset, methodology, and findings, which are also graphically presented in a series of figures and detailed in technical appendices. We then discuss the potential significance and potential policy implications of these findings, describe several limitations of our study, and sketche potential directions for further research.

[6] *Infra*. As explained herein, for the purpose of our study we broadly define this field to include patents in information and communications technology (ICT), fin-tech and med-tech.

[7] *Infra*, tables 1 and 2, and accompanying text. As we explain below, the associations between cited/uncited status and additional two factors (number of inventors and “degree of similarity”) are weak.

[8] *Infra*. Notably, even after controlling for these characteristics, the “U” shaped pattern we find continues to hold, which implies that the increase in the ratio of uncited patents cannot be attributed to any of the aforesaid factors.

Theoretical background: patent data and patent citations

Our study of uncited patents draws on several strands of literature. The first is the large body of economic, network science, and legal scholarship that explores patent data, in order to gain valuable insights about the patent system, and the innovation ecosystem. These studies are based on the understanding that patent repositories are not merely registries of legal rights, but also an excellent source for data, which can provide us with new insights about innovation processes, about the traits of specific inventions, and about the shape of the patent system more generally (e.g., Trajtenberg, 1990; Jaffe and Trajtenberg 2003; Érdi et al., 2013). The registration requirement, which is a pre-condition for obtaining patent rights, implies that patent repositories include significant and highly structured information with respect to each and every registered patent. In addition to information about the invention itself, the registration includes a host of additional data, such as data about the inventors, the patent owners, the technological classes to which the invention is assigned by the Patent Office, and additional factors.[9]

A prominent thread within this body of research is devoted to patent citations, and particularly to highly cited patents. Patent citations are citations of prior art pertaining to the invention. These citations are commonly comprised of previous patents, and infrequently also of scientific literature. Because a patentable invention must be "new " and "non-obvious" relative to the prior art,[10] prior art plays a crucial legal role in the decision whether the invention deserves patent protection (e.g., Érdi et al., 2013). Therefore, the citation of relevant prior art is required as part of submitting a patent application.[11] The applicant's citations are reviewed by the Patent Office examiners, who often contribute additional citations (e.g., Alcacer et al., 2009).

Citations, therefore, reflect relations between inventions: broadly speaking, *backward citations*—citations made by a patent—reflect the previous technologies related, or providing building blocks to the patented invention, while *forward citations*—citations received by a patent—reflect the invention's impact on subsequent technologies (e.g., Trajtenberg et al. 2003b; Lanjouw and Schankerman, 2001). The underlying assumption is that if the technology embedded in the patent is valuable for technological progress, future patents relying on that technology would cite the original patent (e.g., Trajtenberg 1991; Fleming 2001; Érdi et al. 2013; Arts and Veugelers 2015).

Indeed, numerous studies found positive correlations between large numbers of forward citations and various external indications of value. Trajtenberg's pioneering study of patent citations in the field of CT scanners (1990) showed a close association between high number of forward citations and the amounts expended on R&D. Alberta et al. (1991)

9 *See* the USPTO Manual of Patent Examining Procedure (MPEP) (9th. edition, last revised January 2018), available at https://mpep.uspto.gov/RDMS/MPEP/current#/current/d0e18.html.

10 U.S. Patent Act, 35 USCS § 101 ("Whoever invents or discovers any *new* and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title"), and § 103 ("A patent may not be obtained..[…]..if the differences between the subject matter sought to be patented and the *prior art* are such that the subject matter as a whole *would have been obvious* at the time the invention was made to a person having ordinary skill in the art to which the claimed invention pertains") (emphases added).

11 *See* "Duty to Disclose Information Material to Patentability" 37 C.F.R. § 1.56 ("Each individual associated with the filing and prosecution of a patent application has a duty of candor and good faith in dealing with the Office, which includes a duty to disclose to the Office all information known to that individual to be material to patentability..[…]..").

found that highly cited patents were valued by experts as technically important; Harhoff et al. (1999) and Moore (2005) found a positive association between patent citations and the payment of patent renewal fees, so that patents renewed to their full term were "significantly more highly cited than patents allowed to expire before their full term" (Harhoff et al. 1999: 515). Allison et al. (2004) found that litigated patents had higher numbers of forward citations, relative to non-litigated patents. Since the involvement of a patent in litigation is a strong indication of its value, this finding provides another indication for the link between forward citations and patent quality (broadly defined). Likewise, Hall et al. (2005) found that firms whose technology was based on highly cited patents had higher market values.[12] Forward citations have therefore become an acceptable, if noisy, indication for "patent quality", in a broad sense which refers to the technical quality, impact or value of inventions (e.g., Fleming 2001: 142; Strandburg et al. 2006; Csardi et al. 2009: 108).

A related line of research investigates various traits of highly cited patents, in an effort to understand what "building blocks" make a successful invention. Studies in this vein found positive association between high numbers of forward citations and the number of technological sub-classes to which the patent belongs, as determined by the patent office (Schoenmakers and Duysters 2010; Yoshikane et al. 2012). Similarly, research found that highly cited patents have, on average, more backward citations (namely, they cite more patents), and also have more claims, relative to other patents (Kelley et al. 2013; Lanjouw and Schankerman, 2004). Likewise, Wuchty et al. (2007) found that patents invented by more than one inventor, are more likely to be highly cited than patents with a single inventor. However, the relations between these factors and *uncited* patents were not systematically examined.

Additional studies, of particular relevance to our research, concentrate on patent citation *patterns*. These studies examine temporal changes that occur in patent citations over the years in order to draw broader insights about the patent system and about innovation processes. Several studies indicate that patents reach the peak of their citations during the early years after patent grant, with certain variations among technological domains (Strandburg et al. 2006; Mehta et al. 2010).

Kuhn et al. (2018) detect a recent change in patent citation patterns whereby a small minority of patent applications are generating a large majority of patent citations, and argue that the technological similarity between citing and cited patents has significantly weakened in recent years. Strandburg et al. observed an increasing gap between the least and most cited patents since the late 1980s, and suggested that this increase may result from issuing patents on more trivial advances, of lesser technical value. A subsequent work showed that this trend has leveled around 2000 (Strandburg et al. 2006; Csardi et al. 2009). Other research found that although newer patents have more backward citations (i.e., cite more patents) in comparison to older patents, the average number of citations received by a patent has been declining over the years, and proposed that this trend may stem from the grant of lower quality patents (Mehta et al. 2010).

Concomitantly, numerous legal scholars expressed concerns about a decline in patent quality, and argued that the grant of low-quality patents contributes to a sharp increase in the numbers of patent application and patents granted over the past decades, a trend often referred to as a "patent explosion" (e.g., Allison et al. 2010; Masur 2011: 472; Chien 2018). This scholarship further submits that this phenomenon buttresses non-practicing

[12] Fleming (2001) and Arts and Veugelers (2015) found a positive association between patent citations and a high level of recombinations in the patent's backward citations and subclasses.

entities (commonly known as patent trolls), and produces other negative externalities that overall hinder, rather than promote, innovation (e.g., Sichelman 2010).

However, the rich literature on patent citations, whether its focus is on overall citation patterns or on the attributes of inventions, largely concentrates on positive information and on highly cited patents. Conversely, patents that receive *no* citations have not been the subject of substantial scholarly attention. While several recent studies focused on uncited *scientific papers* (Larviere et al. 2009; Noorden 2017), uncited patents have not received such focus. Yet, negative information—in our case, information about patents with no citations—is essential for getting a full picture of the patent system, and more broadly of the innovation ecosystem.[13] We therefore seek to fill in this gap in the literature, by providing a first systematic inquiry of uncited patents. The next Section details the methods of our investigation and describes our findings.

Data, methods and findings

What percentage of patents receive no citations at all? Are there differences in these percentage across technological domains? Does the pattern change over time? And are there specific traits which can be associated with uncited patents? In order to explore these questions, we analyzed data extracted from the USPTO database, on all US patents granted between 1976 and 2008. Our data includes a total number of 3,079,587 observations.[14]

For each year in our study period, we identified the absolute numbers and percentage of patents granted that have not been cited. As we explain in the following section, we control for patents' age and include citations for 10 years following patent grant.

In order to check the robustness of our findings, we further performed several robustness analyses. Primarily, (1) we repeated our analysis while excluding self-citations from the citation counts; (2) we repeated our analysis while considering citations to published patent applications in addition to citations to granted patents; and (3) we repeated our analysis with a universe of patents comprised of uncited patents and patents with only a single citation ("once-cited patents"). The results we describe in the following paragraphs are robust with respect to all these robustness analyses, which are reported in detail below and in the Appendix.

Temporal pattern of non-citation

As indicated above, according to previous studies patents usually reach their citation peak within approximately four years after issuance (Strandburg et al. 2006; Mehta et al. 2010; cf. Hall et al. 2003). Nevertheless, an "uncited" status is never final. The exploratory and combinatory nature of innovation, together with the legal requirement to cite any relevant prior art, imply that a patent can receive citations at any point during its lifetime, and also after its expiry (Strandburg et al. 2006; cf. Fleming and Sorenson 2001). Thus, older

[13] Two prominent studies that focused on other negative aspects embedded in patent data are Moore (2005), who examined patents that were *not* renewed by their owners, and Cotropia and Schwartz (2020), who explored abandoned patent applications that did *not* mature into issued patents.

[14] http://www.patentsview.org/download/. A small number of patents (two percent) did not have complete data needed for the analysis; hence, they were not included in our data. This small amount of missing data is typical in such analyses and does not affect our results.

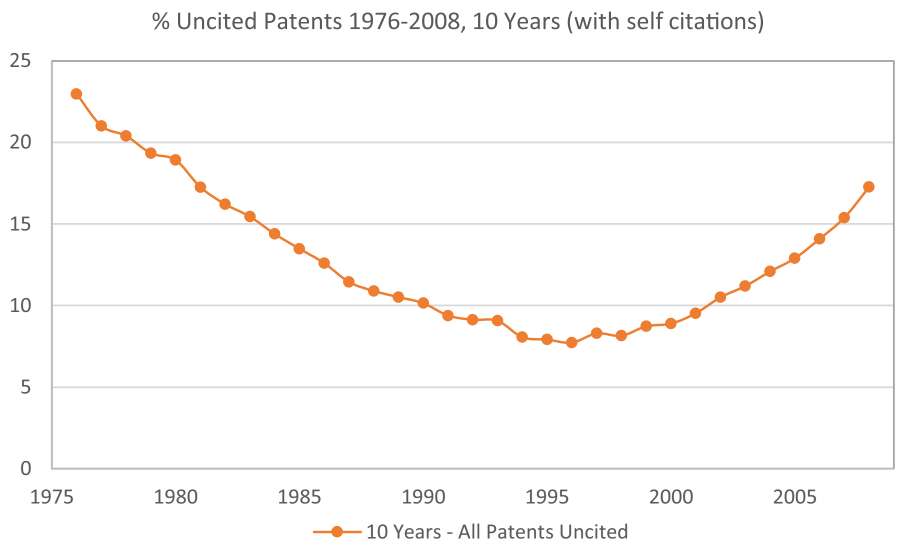

Fig. 1 Uncited Patents 1976–2008 (10 years)

patents have more opportunities to gain citations. As a result, in order to accurately compare the percentages of uncited patents across the period of our study we needed to control for patents' age. We therefore considered citations received in the first 10 years following the grant of the patent. Consequently, our data includes patents issued no later than 2008, so as to allow all patents in our database an equal ten-year period to gain citations.

Our data indicate that, overall, eleven percent of the patents issued during the period of our study did not receive any citations during ten years from issuance. Yet, our data reveals an interesting temporal pattern. Figure 1 shows this temporal pattern, namely the ratio of uncited patents to total patents issued each year during our study period, after controlling for patents age.

Figure 1 demonstrates a U-shaped pattern: From 1976 to (roughly) 1996, the percentage of patents that were uncited *decreased* over time from essentially 22 percent in 1976 to less than 8 percent in 1996. Since 1996, the percentage of uncited patents increased steadily over time, reaching 17 percent in 2008.

Interestingly, this trend is contrary to the trend observed in studies of scientific papers (Larviere et al. 2009). While the percentage of uncited science has been decreasing over the years, the percentage of uncited patents has been continuously increasing since the mid-1990s. In the "Discussion" section we discuss the significance of this pattern and suggest possible explanations.

Non-citation patterns across technological fields

Does this temporal pattern subsist across technological fields? Are there differences in the percentages of uncited patents among the different fields? We performed an initial inquiry of this question, by distinguishing between three categories of patents: (1) pharmaceutical drug patents; (2) software-related patents, which we defined in a broad manner, as including patents in

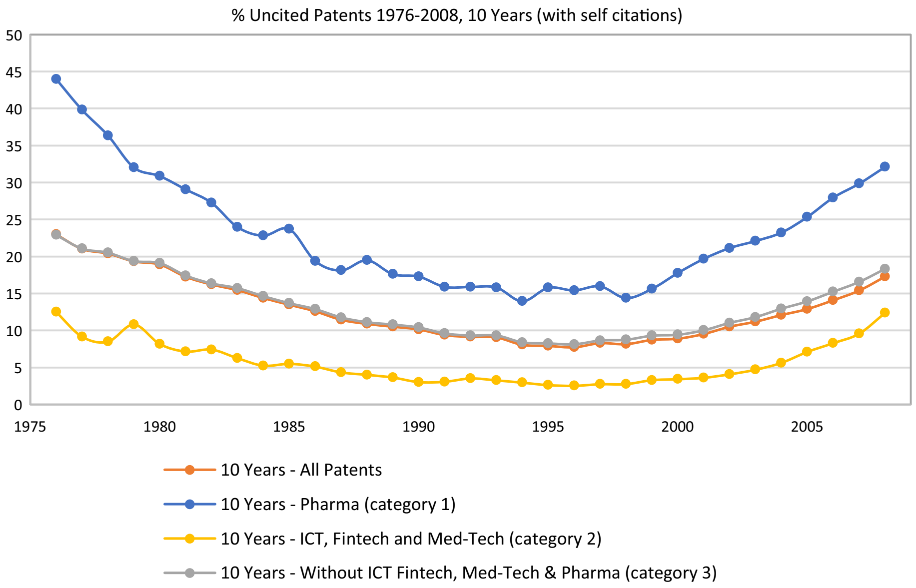

Fig. 2 Uncited patents by categories (10 years)

information and communications technology (ICT), fin-tech and med-tech; and (3) all other patents, namely patents not included in categories (1) and (2). In order to identify the patents which belong to each of these groups, we used the patent classification system, which assigns each patent into subclasses, in accordance with the invention's technological features.[15]

Figure 2 shows the percentage of uncited patents for each of the three categories, together with the general, all-patents data.

Similar to our overall data, the pattern of uncited patents in all the three categories—software-related (yellow curve) pharmaceuticals (blue curve), and others (grey curve)—exhibits a U-shape. The percentage of uncited patents in each of these categories decreased between 1976 and (roughly) 1996, and increased since 1996 and until the end of our period.

However, despite the general similarity in the pattern, there are striking differences among the three groups in the percentages of uncited patents. The percentage of uncited patents in the pharmaceutical drugs field is significantly higher across the entire period, in comparison to software related patents. The ratio of uncited patents in our third, residual, category lies somewhere between the two other categories, and closely tracks the overall non-citation ratio. The fact that a higher percentage of pharmaceutical patents remain uncited, in comparison to software related patents, is somewhat counter intuitive, given

[15] For an overview of the US Patent Classification System (USPC), see http://www.uspto.gov/patents/resources/classification/overview.pdf. For identifying subclasses relevant for pharmaceutical drug patents (our first category) we used the aforesaid classification system and included in this group USPC Classes 424 and 514. For identifying subclasses relevant for med-tech, fin-tech, and information and communications technology patents (our second group) we relied on the list of subclasses previously compiled by Branstetter et al. (2019), and by Gandal and Cohen (2019), which includes ICT/Information Security (USPC), Fin-Tech (IPC) and Med-Tech (IPC). Our third category is a residual group and includes patents that are not included in the other two groups (for example, certain mechanical patents). Notably, because inventions can be classified into more than a single subclass our categories are not completely exclusive.

that patent protection in the pharmaceutical field is often perceived in the literature as more necessary and justified, in comparison to the field of software. We return to this finding in the "Discussion" section.

Lack of citations and patent traits

Is there a possible association between lack of citations and certain patent traits? We examined this question with respect to the following five factors, which we include as explanatory variables in our study:

(1) Number of backward citations–the number of citations made by a patent to preexisting patents;
(2) Number of subclasses—the number of subclasses to which the patent was assigned by the USPTO examiners, in accordance with the invention's technological traits;
(3) Number of claims listed on the patent. The claims determine the scope of the exclusive rights afforded to the patentee. Their initial number is determined by the applicant, and then reviewed by the USPTO examiners;
(4) Number of inventors appearing on the patent; and
(5) Backwards similarity—a variable which reflects the degree of similarity between the patent and its backward citations.

The information about the first four factors appears on the patent. In order to calculate backwards similarity we used the similarity index developed by Lanjouw and Schankerman (2001). Simply put, the index determines similarity on a scale between 0 and 1, in accordance with the fraction of backward citations that belong to the same class as the citing patent, out of the total number of backward citations. To illustrate, a similarity index that equals 1 implies that all cited patents belong to the same subclasses as the citing patent, while a value of 0 implies that all backward citations belong to subclasses that are different from the subclasses of the citing patent.

Intuitively, one would expect that higher values of the first three factors would increase the likelihood of a patent to be cited, since these factors are perceived as indications of the invention's technological breadth, quality, and impact. Indeed, previous studies found positive associations between higher numbers of backward citations (e.g., Yoshikane et al. 2012; Kelley et al. 2013), sub-classes (e.g., Schoenmakers and Duysters 2010; Yoshikane et al. 2012; Kelley et al. 2013), and claims (e.g., Tong and Frame 1994; cf. Lanjouw and Schankerman 2004) on the one hand, and a high number of forward citations on the other. In addition, Allison et al. (2004) found that litigated patents (which are perceived as valuable patents), have higher numbers of claims and backwards citations relative to non-litigated patents. Using data on patents issued in 1991, Moore (2005) further found that patents which expired early due to non-payment of renewal fees by their owners generally had fewer backward citations and fewer claims, relative to patents that were maintained to their full term. Previous research also indicates that patents produced by more than one inventor are more likely to be highly cited than those produced by a single inventor, when self-citations are included (Wuchty et al. 2007). As for the fifth factor, namely the similarity between the patent and its backward citations, existing studies indicate that the relations between this factor and the patent's quality and impact are more nuanced (see, e.g. Phene et al. 2006; Nemet et al. 2012).

Table 1 Descriptive statistics: 1976–2008: cited patents

	# of Obs	Mean	Std. dev.	Min.	Max.
Backward citations	2,754,321	11.75095	20.36809	1	1328
Number of subclasses	2,754,321	4.369612	3.207579	1	260
Number of claims	2,754,321	15.65761	13.24943	1	887
Number of inventors	2,754,321	2.242832	1.611063	1	51
Backward similarity	2,754,321	0.6145774	0.3541722	0	1

Table 2 Descriptive Statistics: 1976–2008: uncited patents

	# of Obs.	Mean	Std. Dev.	Min.	Max.
Backward citations	325,266	7.853425	10.57848	1	773
Number of Subclasses	325,266	3.981225	3.165645	1	164
Number of claims	325,266	12.21976	9.965565	1	706
Number of inventors	325,266	2.209788	1.611371	1	32
Backward similarity	325,266	0.6070431	0.3809577	0	1

The two tables above provide summary data for these variables. Table 1 described the cited patents, while Table 2 describes the uncited patents:

The data in Tables 1 and 2 indicate that uncited patents have, on average, fewer backward citations and less claims than cited patents, and also belong to fewer technological subclasses. Backward similarity values and the number of inventors are roughly the same for cited and uncited patents.

We further conducted a regression analysis of these data. The details of our regression and formal econometric analysis, including a summary of our regression results, appear in Appendix 1.

Our analysis demonstrates that patents with more backward citations, more subclasses, and more claims are much more likely to receive at least one citation, and this effect is statistically significant.[16] Since numbers of claims, subclasses, and backward citations have been previously recognized as positively related to highly cited patents, which are considered high-quality patents, these findings are not entirely surprising. Yet, they provide an initial indication that the quality of uncited patents is indeed lower.

Importantly, the regression results in Tables 3 and 4 in the Appendix also demonstrate that our findings concerning the "U" shape temporal pattern of uncited patents continue to hold, even after we control for all these factors. In other words: after controlling for a series of prevalent factors that might affect citations, the likelihood of not being cited decreases from 1976 to (roughly) 1996, and increases consistently from 1996 through 2008.

[16] See Table 3 in Appendix 1, *infra*. The Table as well as other Tables in the Appendix demonstrate that the additional two factors (numbers of inventors and backwards similarity) do not have a substantial association with "uncited" status.

Lack of citations and patent renewals

In addition to the aforesaid traits, we also examined whether uncited patents are less likely to be renewed by their owners, relative to cited patents.

Patent owners are required to pay periodical maintenance fees (also known as "renewal fees") during the term of the patent.[17] The non-payment of those fees results in the expiration of the patent before the end of the statutory term (although, or course, the patent still appears in the repository and can be cited as prior art). The payment of renewal fees therefore constitutes an external indication for the value of the patent. Previous studies indeed found positive associations between the payment of renewal fees and high number of forward citations (Harhoff 1999; Moore 2005). Against this background, we examined whether an "*uncited*" status is associated with the *non-payment* of patent renewal fees, using USTPO data about patents which expired due to non-renewal.[18]

We find a negative and significant correlation between patents that were cited at least once and the nonpayment of patent maintenance fees, which indicates that uncited patents, on average, are less likely to be renewed, relative to cited patents. Further, the partial correlation between uncited patents and nonpayment of patent renewal fees is higher (in absolute value) than the partial correlations between uncited patents and all but one of the other traits we employed in the analysis.[19]

We further conducted a regression analysis which demonstrates that the "nonpayment of patent maintenance fees" is negatively associated with whether a patent has been cited and that this effect is statistically significant. The analysis further shows that these findings continue to hold after we control for the five patent traits we examined above. These results provide another external indication, in addition to the patent traits described above, that the quality of uncited patents may indeed be lower in comparison to cited patents.

Our findings concerning patent renewals are further detailed in the Appendix, Tables 5 and 6.

Robustness analyses

In order to check the robustness of our findings, we performed the following three analyses.

1) Excluding self-citations

As explained above, in our main analysis we considered self-citations as citations. In order to check the robustness of our findings we performed an additional analysis with self-citations excluded. In other words, under this second analysis citations to a patent by a subsequent patent with the same patent holder or the same inventor are not considered "citations", and a patent whose sole citations are self-citations would be considered "uncited". Figure 3 below shows the percentages of uncited patents by categories, with self-citations excluded.

---

[17] *See* https://www.uspto.gov/patents-maintaining-patent/maintain-your-patent. The fees are due for patents issued after 1980, and are payable in three intervals during the life of the patent, occurring after 3–4 years, 7–8 years and 11–12 years from issuance.

[18] https://uspto.data.commerce.gov/dataset/Patent-Maintenance-Fee-Events-1981-Present-/95ij-9exb. Note that this data is available only with respect to patents issued after 1982.

[19] A partial correlation between two variables is the correlation between them after removing the effects of all other variables. See Table 5.

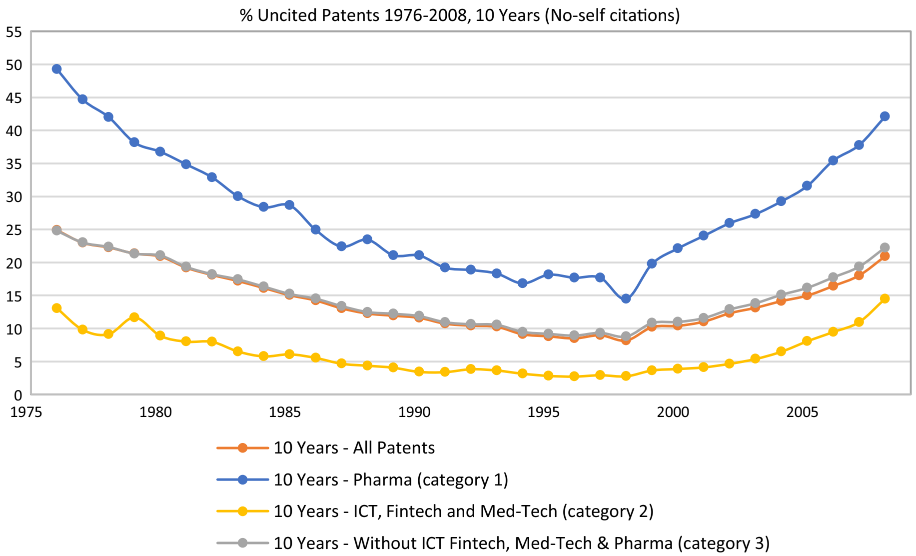

Fig. 3 Uncited patents by categories (10 years), no self-citations

As is apparent from Fig. 3, while excluding self-citations somewhat raises the percentages of uncited patents, the temporal U-shape pattern for all categories remains unchanged. Additional details about this analysis appear in the Appendix.

2) Citations to patent applications

Following a statutory amendment in U.S. patent law, applications filed on or after November 29, 2000 are generally published after the lapse of eighteen months from submission, whether or not a patent has been granted.[20] Therefore, when patent applications were published before grant, citations may refer to the patent application, and not only to the granted patent (Kuhn et al. 2018; and Cotropia and Schwartz, 2020). In order to verify that the increase in the rate of uncited patents did not result from citations to patent applications (which replaced citations to granted patents) we repeated our analysis while taking into account citations made to patent applications: For patents published since 2002, we linked and united the citations to the published applications with the citations to the patents themselves. In order to maintain our age limit, we counted all citations (either to the patent application or the patent itself) during the ten year period following the publication of the patent application. The details of our analysis appear in the Appendix ("Robustness analysis: citations to published patent applications"), Tables 7 and 8. As these tables demonstrate, the U-shape we identified continues to hold even when considering citations to patent applications, and importantly, the percentage of uncited patents still increases consistently from (roughly) 1996 through 2008. Thus, our results cannot be attributed to the aforesaid change in patent policy.

[20] *See* 35 U.S.C. 122; https://www.uspto.gov/web/offices/pac/mpep/s1120.html. As detailed therein, there are certain nuances in the calculation of the 18 months period, yet these are immaterial for the purpose of this study.

3) Uncited versus once-cited patents

Finally, we repeated our analysis while considering uncited patents versus patents that received a single citation during a ten-year period ("once-cited patents"). The details of this analysis appear in the Appendix ("Robustness analysis: uncited versus once-cited patents", Tables 9 and 10)

As these tables show, the results of our analysis are qualitatively similar. Most importantly, the U-shape pattern of non-citations continues to hold even when our "universe" of patents is comprised only of patents that were uncited and patents that received just a single citation: between 1976 and (roughly) 1996, the percentage of uncited patents decreased, whereas between 1996 and 2008, this percentage has been increasing (essentially monotonically).[21]

In addition, examining the patent traits described above, when the"universe of patents" is comprised only of uncited patents and once-cited patents, shows that backwards citations, claims, and the number of technological subclasses are very significantly associated with whether the patent was cited.[22] As expected, the magnitudes of the coefficients are smaller in comparison to our "base case" where we compare uncited patents with the entire universe of all cited patents. These results support the robustness of our findings and provide yet another indication that the quality of uncited patents may indeed be lower.

A long tail of patent citations?

Finally, we take a look at our raw data, without controlling for patents' age. While our 10-year limit is necessary in order to compare and evaluate citation patterns, looking at the raw data without controlling for age can provide us with a different insight regarding the actual existence of citations beyond the 10-year period. Figure 4 shows our results for patents that did not receive any citations, by technological categories, without controlling for patent's age.

The oldest patents in our database had more than forty years to gain citations, in comparison to the youngest patents, that had only ten years. Despite this distorting factor, this figure still demonstrates a decrease in the percentage of uncited patents between 1976 and 1996, and a change of this trend from 1996 onwards. Interestingly, however, the percentage of 'old' uncited patents drops significantly in all categories when we lift the 10-year limitation. For example, 22 percent of all patents issued in 1976 were uncited after 10 years (Fig. 1), but only 5 percent remained uncited in 2018, 42 years from issuance. Similarly, for pharmaceutical patents issued in 1976 the percentage of uncited patents drops from 44 percent after 10 years to roughly 20 percent after 42 years.[23]

---

[21] Additionally, our results are robust when comparing "rarely cited patents", namely patents that obtained either zero citations or a single citation, to patents that receive more than one citation.

[22] The associations between uncited patents and the two other factors-backwards similarity and number of inventors-are relatively small or negative.

[23] Obviously, with respect to younger patents issued after 1996 the gap between uncited after "10 years" and uncited after "all years" narrows.

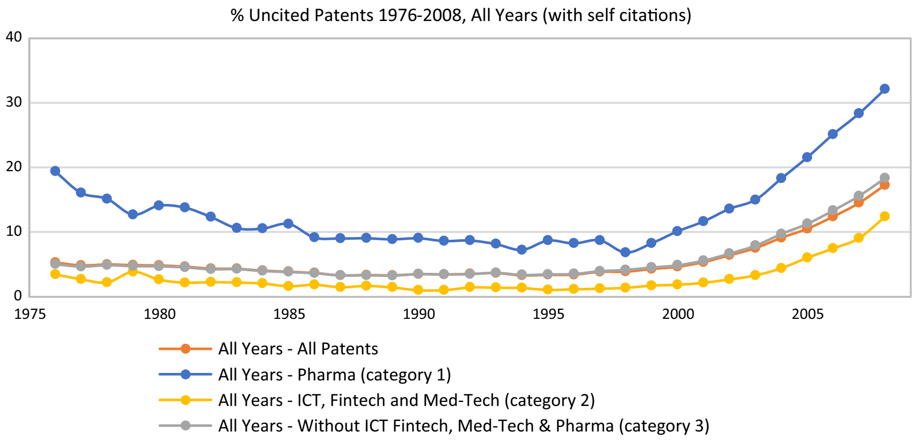

Fig. 4 Uncited patents by categories (all years, self-citations included)

These results imply that even if patents remain uncited for ten years after grant, many of them can still receive at least one citation after that period. In other words, the patent citation tail is a 'long tail'. These findings are consistent with a previous study by Standburg et al., which observed that "patent citability dies out surprisingly slowly with age" (Strandburg et al. 2006: 1340–41). They also strengthen and reinforce a growing stream of research maintaining that innovation is not a strictly cumulative and linear process, but rather a combinatory, exploratory and networked process, whereby old technologies may gain new significance at a later stage (e.g., Fleming and Sorenson 2001; Fleming 2001; Shur-Ofry 2017). This line of inquiry deserves further research.

Discussion

What are the possible interpretations of our findings, and what is their significance for innovation policy?

Several factors may explain the U-shape temporal pattern. A plausible explanation for the left side of the U, namely the decline in the ratio of uncited patents between 1976 and 1996, may be the substantial improvement in patent search tools. During the first years of our period patents were published on paper, and had to be searched manually (Grigg 2003). This state of affairs gradually changed with the introduction of a system that allowed a computerized CD-ROM-based search in 1988, and an internet-based search in 1996. The website launched by the USPTO in 1996 was first limited to bibliographic patent information, but since 2000 enabled to search the full-text of all patents issued from 1976 onwards (and some more limited search of earlier patents) (Grigg 2003). This timeline is largely consistent with our findings that the percentage of uncited patents reached its lowest value around 1996.

The increase in the ratio of uncited patents since 1996 is possibly connected to the sharp increase in the overall number of patents issued by the USPTO during that period. According to USPTO data, in 1976, the first year of our study, the cumulative number of patents in the USPTO registry, which started in 1790, was approximately four million patents.

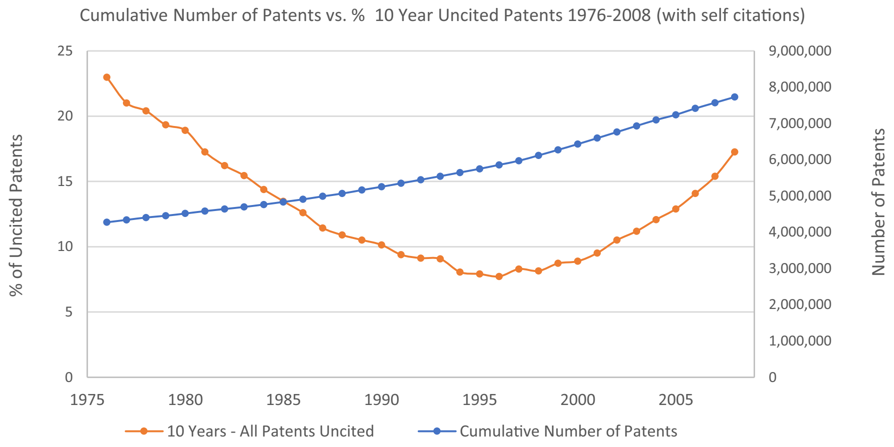

Fig. 5 Cumulative number of US patents vs. Percentage of Uncited Patents 1976–2008

In 2008 the number of patents exceeded seven million. Figure 5 shows the percentage of uncited patents against the *cumulative* number of US patents during the period of our study.

A closer look at the number of patents issued *per year* during our study period reveals a sharp rise in issued patents during the second half of our period, from the mid-1990s. Figure 6 shows the percentage of uncited patents against the yearly numbers of US patents issued during the entire period of our study.

This significant increase in patents issued since the mid-1990s is largely consistent with the increase we find in the percentage of uncited patents during the same period. Hence, one possible interpretation of our results is that despite the improvement in search capabilities, the search for prior art becomes more difficult the more patents there are in the registry. Therefore, in an era of "patent explosion" more patents are left uncited.

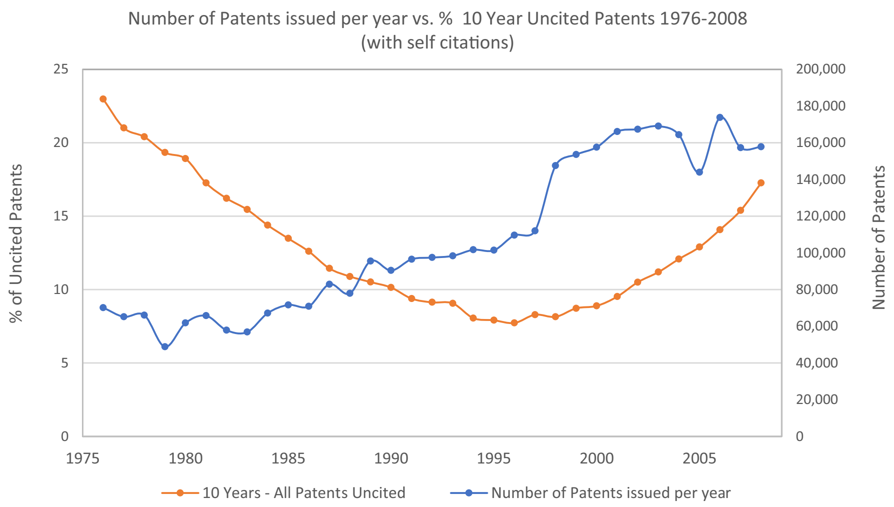

Fig. 6 Yearly Numbers of Patents Issued vs. Percentage of Uncited Patents, 1976–2008

A second interpretation, which is not contradictory to the first one, concerns patent quality. In recent years, numerous legal scholars expressed concerns about a decline in patent quality. This literature generally maintains that the issuance of low quality patents is a major cause for the recent patent explosion. It argues that low quality patents often find their way to the hands of non-practicing entities (commonly known as patent trolls), create "patent thickets" that hinder technology commercialization, and produce other negative externalities that overall impede, rather than promote, innovation (e.g., Merges 1999; Lemley and Sampat 2008; Hemphill and Sampat 2012; Lemley 2013; Chien 2018; Frakes and Wasserman 2019). The growing consensus that the quality of patents granted deserves the attention of policy makers has led the U.S. Patent Office to declare patent quality as a priority, and to consider various ways to improve it.[24]

While we cannot offer a definitive conclusion, to a certain extent the picture emerging from our study reinforces patent-quality concerns. Because citations are perceived as an indication of quality (in the broad sense used in this paper, that includes, beyond technological quality, properties such as impact, or value), the increase in the percentage of patents with zero citations could imply an increase in low-quality patents that do not serve as building blocks for subsequent innovation.

Our analysis further reveals negative relationships between uncited patents and several patent traits that were previously associated with successful, highly cited, patents, namely numbers of backward citations, subclasses, and claims. We found that uncited patents largely possess "less" of those traits. We also found that uncited patent are less likely to be maintained by their owners through the payment of periodical renewal fees. These findings may not be surprising, yet they reinforces the concerns that uncited patents are indeed of lower quality, and that the increase in the rate of uncited-ness signals a decrease in patent quality. The fact that, unlike patents, the percentage of uncited scientific papers is continuously decreasing (Larviere et al. 2009; Noorden 2017) may provide further support for this explanation.

Cautionary notes are warranted. First, the factors we examined are not exhaustive, and further studies could expand this line of exploration and examine the links between uncited patents and a series of additional traits and external factors, which were previously linked with highly cited patents.[25] These factors include, for example, expert evaluations (Albert et al. 1991), R&D investments in the underlying technologies (Trajtenberg 1990) and the owners' market value (Hall 2005). We outline below additional directions for such explorations that would deepen our understanding of the association between uncited-ness on the one hand, and lack of social or private value, on the other. Secondly, patent citations are a rough proxy, and cannot provide completely precise information of the value of specific patents: put

[24] *See* United State Patent and Trademark Office, Patent Quality, https://www.uspto.gov/patent/patent-quality (stating that "[t]o ensure we continue to United State Patent and Trademark Office, Patent Quality issue high-quality patents that will fuel innovation well into the future, the Office of the Deputy Commissioner for Patent Quality, along with our partners across the Patents organization, promotes and supports the continuous improvement of patent products, processes and services through collaboration with internal and external stakeholders of the intellectual property community.") United State Patent and Trademark Office, Office of the Deputy Commissioner for Patent Quality https://www.uspto.gov/about-us/organizational-offices/office-commissioner-patents/office-deputy-commissioner-patent-19 ("The Deputy Commissioner for Patent Quality is responsible for optimizing the quality of patent products, processes and services to build a culture of process improvement and enhanced patent quality ").

[25] For example, Fleming (2001) and Arts and Veugelers (2015) found a positive association between patent citations and a high level of recombinations in the patent's backward citations and subclasses.

differently, "zero citations" does not necessarily imply "zero quality". And while our examination paints a general picture, specific cases may of course deviate from that picture. Nevertheless, the "big picture" emerging from our data raises concerns about patent quality.

The differences we find in the percentage of uncited patents between categories, particularly between drug patents and software related patents are counter-intuitive. The pharmaceutical field is considered the "poster child" of the patent system, and the area in which strong patent protection is most justified (e.g., Ouellette 2010: 300). Conversely, there is a longstanding policy debate that casts doubt on the justifications for and necessity of software patents (e.g., Landes and Posner 2003: 326; Bessen and Meurer 2008; Bessen and Maskin 2009). Yet, our analysis demonstrates that the percentages of uncited patents are consistently higher for drug patents in comparison to software-related patents.

One possible explanation may relate to the tendency of pharmaceutical companies to protect drugs by a series of secondary patents that do not involve new active ingredients, in order to prolong their life-cycle. Studies indicate that this phenomenon, also known as "ever-greening", has been growing over time. For example, Ouellette (2010) observed that the average number of patents per drug increased from 2.5 in the late 1980s to nearly 3.5 in 2005. Kapczynski et al. (2012) found that around 50% of drugs are protected by secondary patents, while Feldman (2018) noted that 78% of the drugs associated with new patents in the FDA's records between the years 2005 and 2015 were not new drugs coming on the market, but existing drugs that are "recycled" by their owners. These secondary patents are often considered low-quality patents (e.g. Hemphill and Sampat, 2012: 345), which may explain the relatively high percentage of uncited drug patents. Another explanation may be that the differences we find are related to the much larger number of software-related patents relative to drug patents in the groups we examined.[26]

Yet another explanation may concern references to non-patent literature. Studies suggest that patents in the life-sciences tend to cite more scientific, non-patent literature, relative to patents in other fields (e.g., Ahmadpoor and Jones 2017; Verbandt and Vadot 2018). Possibly, such references to non-patent literature come, de facto, "at the expense" of citation of drug-patents.[27] Relatedly, Glänzel and Meyer (2003) found that scientific articles in the field of "Drugs and Medical" were more likely to cite patents, relative to articles in the field of "Computer and Communications", which may imply that some of the value of some pharmaceutical patents is actually manifested in contribution to science.[28] Finally, our findings possibly reflect different dynamics of innovation, for example that stakeholders in the pharmaceutical area "rush to patent" at an early stage, and register patents in numerous directions, many of which do not mature into valuable innovation, or that innovation in this field has less cumulative traits than in the software-related domains.[29] All these hypotheses, however, require further exploration.

[26] For the patents included in each category, see *supra*, note 21 and accompanying text. As explained therein, our definition of software-related patents includes patents in the fields of information and communications technology, fin-tech and med-tech.

[27] One should note, however, that the legal requirement is to consider patent applications against all relevant prior art, namely patents *and* NPL. Therefore, this hypothesis certainly warrants further validation.

[28] Note, however, that according to Glänzel and Meyer (2003: 422), 98.5% of US patents have not been cited in scientific literature.

[29] *Cf.* Noorden (2017) who discusses uncited science and suggests that certain domains may be less "cumulative" than others.

The finding that the percentage of uncited patents in software-related fields is *lower* than the average for all technological fields is also surprising, in light of prominent criticisms that software patents are often trivial, and have limited use as sources for subsequent developments (e.g., Bessen and Meurer 2008). One should note that our definition of software-related patents is broad and includes patents in the fields of information and communications technology, fin-tech and med-tech. Subject to this broad definition, our analysis seems to reflect the cumulative nature of innovation in the tech-related fields, and suggests that software-related patents provide building blocks to subsequent technologies, no less (and even more) than other categories of patents. It should be noted that because the youngest patents in our database were issued in 2008, our results do not reveal the potential impact of the recent U.S. Supreme Court cases that raised the threshold of patentability in the software field.[30] The question whether and how these decisions, which, presumably, could raise the quality of software related patents, will impact the rate of uncited patents in this area certainly warrants future examination.[31]

Overall, although our study indicates that a substantial majority of patents receive at least one citation within 10 years from grant, the trend we identify is disturbing. If innovation is a networked process, the fact that more and more patents remain outside the network is a cause for concern. We therefore believe that a systematic exploration of the rate of uncited patents and the tracing of temporal changes in this rate can provide a complementary, evidence-based measure, for evaluating the shape of the patent system and the overall quality of the patents granted by the USPTO.

Nevertheless, the exploration we present in this Article is merely a first step in this direction and one should be cautious in the interpretation of our findings. One limitation relates to the period of our study, which ends in 2008. As we explain earlier, this limitation stems from the need to allow patents a sufficient period of time to gain forward citations. This implies that our findings do not reflect changes in non-citation rates that may have occurred after the passage of the America Invents Act of 2011.[32] The legislation introduced a series of amendments to the U.S. Patent Act designed to improve patent quality, including a legal mechanism to challenge patents post grant (see, e.g., Carrier 2012). The future tracing of the rates of uncited patents before and after the reform may provide insight as to its effect, and certainly warrants future research.

Similarly, our study examined citations of patents by subsequent patents. Yet patents can also be cited in scientific literature (Glänzel and Meyer 2003), which implies that the cited patent served as a source of knowledge, and therefore has some social value (Ouellette 2012: 555; Fromer 2016: 1717). Future research can examine whether and to what extent any of the uncited patents received citations in scientific literature. Nevertheless, since previous research indicates that the vast majority (98.5%) of US patents have not been cited in scientific literature (Glänzel and Meyer 2003: 422), this inquiry is unlikely to significantly alter our main findings.

[30] Two prominent decisions are Bilski v. Kappos, 561 U.S. 593 (2010) (holding that patents can be rejected on subject-matter eligibility grounds and denying a patent over a method of risk hedging); Alice Corp. Pty. v. CLS Bank Int’l, 134 S. Ct. 2347 (2014) (holding that implementing a merely abstract idea on a computer is unpatentable).

[31] *Cf.* Chien (2018: 90–92) (reviewing the recent case law that elevated the patentability standards that apply to software, and suggesting that their long-term impact on patent quality is yet to be evaluated).

[32] Leahy-Smith America Invents Act, 125 Stat. 284.

Likewise, our inquiry is limited to patents issued in the U.S. However, U.S. patents can also be cited by foreign patents, or by applications that did not mature into an issued patent.[33] Thus, another line of research would be to inquire whether, and to which extent, patents that receive no citations from subsequent U.S. patents still receive citations from other sources, including scientific literature, foreign patents, or abandoned applications.

Finally, inventions could potentially have commercial value despite lack of citations to their underlying patents. Therefore, another direction for future exploration would be to cross the data on uncited patents with data concerning patent transactions,[34] with data about patents involved in litigation (in itself a proxy for commercial value—cf. Allison et al. 2004), and with data about patents held by non-practicing entities.[35]

These directions for future research are of course non-exhaustive. Rather, they illustrate the broad potential of the approach we choose in this study. Overall, the focus on uncited patents opens up myriad research questions, the study of which can illuminate a shadowed side of the patent system, and provide important evidence that would allow us to better evaluate it.

Conclusion

Our systematic study of patents that receive no citations yields three principal insights. First, we find a robust U-shape pattern, whereby the percentage of uncited patents decreased between 1976 and 1996, but has been constantly increasing since then. Second, we find counter-intuitive differences in the rates of uncited patents between different technological fields, primarily drug patents and software-related patents. Third, our analysis reveals that uncited patents are negatively associated with several indications for patent quality. From the perspective of innovation policy, these findings are troubling. They raise, and reinforce, concerns regarding patent quality and 'patent explosion'. On a more general note, this study's systematic focus on patents that are "out of sight" opens up new avenues for future research, and demonstrates how exploration of negative information embedded in patent data can provide us with important knowledge and a deeper, more nuanced, understanding of our ecosystem of innovation.

33 *Cf.* Cotropia & Schwartz (2020)(discussing patent applications that were abandoned for various reasons, and demonstrating that "abandoned applications" still receive significant citations from patent examiners).

34 *E.g.*, the data concerning patent assignments recorded with the USPTO: Patent Assignment Dataset, https://www.uspto.gov/learning-and-resources/electronic-data-products/patent-assignment-dataset.

35 *See, e.g.*, the Stanford NPE Litigation Database, which tracks patent litigation initiated by non-practicing entities and patent assertion entities (often referred to as "patent trolls")—https://law.stanford.edu/projects/stanford-npe-litigation-database/#slsnav-brief-dataset-methodology.

Appendix: Formal econometric (regression) analysis

Variables

The dependent variable is "Cited", where Cited equals one if the patent receives one or more citations in the first ten years following its issuance. If the patent receives no citations during the first ten years following its issuance, Cited equals zero.[36]

Since the dependent variable is a binary variable, we run a Logistic regression. The same qualitative results are obtained using a Probit regression.

The independent variables in the regressions are number of backward citations, number of subclasses, number of claims, number of inventors and backwards similarity.

Because all independent variable except for backwards similarity are highly skewed, we enter these variables in logarithms. The independent variables included in the regression are.

l_back_cites–the natural logarithm of the number of citations the patent made to preexisting patents.

l_subclasses–the natural logarithm of the number of subclasses listed on the patent.

l_claims–the natural logarithm of the number of claims.

l_inventors–the natural logarithm of the number of inventors on the patent.

b_similarity–the backward similarity as defined above.

Finally, we include dummy variables for the grant year. These are the primary variables of interest. We include data from 1976 to 2008; therefore we have dummy variables for each year from 1977 to 2008.[37]

Regression analysis

The estimation equation is as follows, where the subscript "j" refers to each patent. For compactness we do not list the dummy variables for year.

$$\begin{aligned}\text{Cited}_j = &\beta_0 + \beta_1 * \text{l\_back\_cites}_j + \beta_2 * \text{l\_inventors}_j \\ &* \text{l\_subclasses}_j + \beta_3 * \text{l\_claims}_j + \beta_4 * \text{b\_similarity}_j\end{aligned}$$

The results of the logistic regression are shown in Tables 3 and 4.

Table 3 Regression Results

Dependent variable: Cited	Coefficient	Standard error	Z-Statistic
Independent variables			
l_back_cites	0.378	0.002	172.31***
l_inventors	0.029	0.003	9.18***
l_subclasses	0.200	0.003	67.16***
1_claims	0.321	0.002	140.59***
Backward similarity	0.153	0.005	29.92***
3,079,587 observations	Pseudo R-squared	0.04	

***means significant at the 99% level of confidence.

[36] The variable CITED includes self-citations unless noted otherwise.

[37] Since we have a constant in the regression, we cannot include a dummy variable for 1976. Otherwise, there would be perfect multicollinearity.

Table 4 Coefficients on the yearly dummy (binary) variables

Year	Coefficient
1977	0.081
1978	0.005
1979	0.033
1980	0.023
1981	0.064
1982	0.147
1983	0.165
1984	0.241
1985	0.276
1986	0.348
1987	0.446
1988	0.470
1989	0.483
1990	0.527
1991	0.591
1992	0.602
1993	0.585
1994	0.679
1995	0.675
1996	0.693
1997	0.606
1998	0.612
1999	0.526
2000	0.482
2001	0.394
2002	0.250
2003	0.143
2004	0.037
2005	– 0.071
2006	– 0.166
2007	– 0.256
2008	– 0.378

All of the estimated parameters are highly significant (*** means significant at the 99% level of confidence.) The estimated coefficients on the yearly dummy (binary) variables from the regression in equation I are shown in Table 4 below.

Thus after controlling for patent characteristics Table 4 shows that the pattern is exactly as in the raw data. From 1976 to 1996, a higher percentage of the patents are being cited over time. This is because the estimated coefficients on the yearly dummy variables increase (essentially monotonically) over that period. From 1996 through 2008, a lower percentage of the patents are cited over time and this decline is also essentially monotonic.

Lack of citations and patent renewals

The partial correlations between whether a patent was cited at least once and the explanatory variables are shown in Table 5. The first explanatory variable refers to the expiry of the patent due to non-payment of maintenance fees.

We further conducted formal regression analysis in which the variable "nonpayment of maintenance fees" was added to the regressions as an explanatory variable, to denote patents that expired due to non-payment of maintenance fees. In this part of the analysis, we included patents from 1982 through 2008, since the information on patent renewals is only available beginning in 1982. Our regressions results in Table 6 show that the estimated coefficient on the variable "nonpayment of maintenance fees" is negative and statistically significant (at 99 percent level of confidence). This means that, other things being equal, uncited patents were less likely to be renewed. We further found that after controlling for patent traits including the non-payment of renewal fees the "U-shape" pattern continues to hold. The latter results are available from authors upon request.

All these result hold regardless of whether we include or exclude self-citations.

Table 5 Partial Correlations between the variable "Cited" and the explanatory variables (grant year 1982 or later)

Variable	Partial correlations with the variable Cited(with self-citations)	Partial correlations with the variable Cited (no self-citations)
Nonpayment of maintenance fees	−0.08	−0.08
l_back_cites	0.09	0.09
l_inventors	−0.01	−0.02
l_subclasses	0.04	0.04
1_claims	0.06	0.06
Backward similarity	0.02	0.02

Table 6 Regression Results –with the variable "nonpayment of patent maintenance fees included.)

Dependent variable:	Cited (with self-citations)	Cited (no self-citations)
	Coefficient (Std. Error)	Coefficient (Std. Error)
Independent Variables		
Nonpayment of maintenance fees	−0.570 (0.004)***	−0.518 (0.004)***
l_back_cites	0.378 (0.002)***	0.361 (0.002)***
l_inventors	−0.004 (0.003)	−0.074 (0.003)***
l_subclasses	0.183 (0.003)***	0.165 (0.003)***
1_claims	0.299 (0.002)***	0.286 (0.002)***
Backward similarity	0.143 (0.005)***	0.115 (0.005)***
Pseudo R-squared	0.05	0.05
2,935,152 observations		

***means significant at the 99% level of confidence

Robustness analysis: citations to published patent applications

See Tables 7 and 8

Table 7 Regression Results taking into account citations to published patent applications since 2002

Dependent variable: Cited	Coefficient	Standard error	Z-Statistic
Independent Variables			
l_back_cites	0.363	0.002	171.08***
l_inventors	0.025	0.003	8.02***
l_subclasses	0.190	0.003	65.44***
1_claims	0.317	0.002	141.73***
Backward similarity	0.160	0.005	32.04***
3,079,587 observations	Pseudo R-squared	0.05	

*** means significant at the 99% level of confidence.

Table 8 Coefficients on the yearly dummy (binary) variables

Year	Coefficient
1977	0.079
1978	0.005
1979	0.032
1980	0.023
1981	0.065
1982	0.148
1983	0.166
1984	0.242
1985	0.277
1986	0.349
1987	0.447
1988	0.473
1989	0.485
1990	0.530
1991	0.593
1992	0.606
1993	0.589
1994	0.683
1995	0.681
1996	0.698
1997	0.612
1998	0.618
1999	0.533
2000	0.489
2001	0.402
2002	0.254
2003	0.076
2004	−0.095
2005	−0.255
2006	−0.389
2007	−0.494
2008	−0.619

Robustness analysis: uncited versus once-cited patents

See Tables 9 and 10

Table 9 Regression Results for Uncited Patents versus Once-Cited Patents ("***" means significant at the 99% level of confidence.)

Dependent variable: Cited	Coefficient	Standard error	Z-Statistic
Independent Variables			
l_back_cites	0.136	0.0030	45.81***
l_inventors	−0.057	0.0042	−13.66***
l_subclasses	0.044	0.0039	11.39***
1_claims	0.095	0.0031	30.62***
Backward similarity	0.030	0.0065	4.63***
671,974 observations	Pseudo R-squared	0.01	

Table 10 Coefficients on the yearly dummy (binary) variables

Year	Coefficient
1977	−0.020
1978	−0.061
1979	−0.026
1980	−0.017
1981	−0.007
1982	0.014
1983	0.010
1984	0.050
1985	0.042
1986	0.056
1987	0.099
1988	0.084
1989	0.106
1990	0.104
1991	0.119
1992	0.124
1993	0.069
1994	0.130
1995	0.092
1996	0.101
1997	0.039
1998	0.073
1999	0.034
2000	0.013
2001	−0.045
2002	−0.093
2003	−0.141
2004	−0.170
2005	−0.212
2006	−0.245
2007	−0.299
2008	−0.339

Additional robustness analyses

We also ran the following robustness regressions:

- We excluded self-citations for all regressions, that is, citations to patents with the same patent holder or the same inventor.
- We included dummy variables for the eight IPC classes.
- We included dummy variables for software and pharmaceutical patents
- We used a linear rather than log linear functional form for the explanatory variables

The results are qualitatively unchanged, except that in the case where we exclude self-citations, the coefficient associated with the number of inventors is negative and significant, rather than positive and significant. The temporal graph of uncited patents is still U-shaped. These results are available from the authors upon request.

Acknowledgements We are grateful to the two anonymous referees for extremely helpful suggestions and comments. For valuable comments and input we also thank Ronen Avraham, Eran Bareket, Miriam Marcowitz-Bitton, Reuven Cohen, Rebecca Eisenberg, Roger Ford, Jeanne Fromer, Richard Gruner, Sharon Hausdoff, Dmitry Karshedt, Paul Kaye, Mark Lemley, Michael Livermore, David Schwartz, Ofer Tur-Sinai, Saurabh Vishnubhkat, and the participants of the European Law and Economics Association Conference (2019), the Data Science and Law Conference at Bar-Ilan University (2019), the WIPO Conference on Emerging Technologies (2019), and the Intellectual Property Scholars Conference at Stanford University (2020, Online). We are also grateful Andrew Toole for assistance in obtaining relevant data. Neil Gandal is very grateful for a grant from the Foerder Institute for Economic Research.

References

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: patented inventions and prior scientific advance. *Science, 357*(6351), 583–587.

Albert, M. B., Averyb, D., Narina, F., & McAllister, P. (1991). Direct validation of citation counts as indicators of industrially important patents. *Research policy, 20*(3), 251–259.

Allison, J. R., Lemley, M. A., Moore, K. A., & Trunkey, R. D. (2004). Valuable Patents. *GEO L J., 92,* 435.

Allison, J. R., Lemley, M. A., & Walker, J. (2010). Patent quality and settlement among repeat patent litigants. *Geo L J., 99,* 677.

Allison, J. R. (2018). Patent value. In P. Menell & D. Schwartz (Eds.), *Research handbook on the economics of intellectual property law*. Cheltenham: Edward Elgar Publishing.

Arts, S., & Veugelers, R. (2015). Technology familiarity recombinant novelty and breakthrough inventions. *Industrial and Corporate Change, 24,* 1215.

Ashtor, J. (2019). Investigating cohort similarity as an ex ante alternative to patent forward citations. *Journal of Empirical Legal Studies, 16,* 848.

Beck, & Roger, L. (1981). Competition for patent monopolies. *The RAND Journal of Economics, 3,* 91.

Bessen, J., & Maskin, E. (2009). Sequential innovation, patents, and imitation. *The RAND Journal of Economics, 40*(4), 611.

Bessen, J., & Meurer, M. J. (2008). *Patent failure: How judges bureaucrats and lawyers put innovators at risk*. New Jersey: Princeton University Press.

Branstetter, N. Gandal, & N. Kunievsky*, Network-mediated knowledge spillovers: A cross-country comparative analysis of information security innovations,* CEPR discussion paper #12268, available at https://cepr.org/active/publications/discussion_papers/dp.php?dpno=12268.

Carrier, M. (2011). *UCDL Rev, 45,* 103.

Cotropia, CA & Schwartz, L, *The hidden value of abandoned applications to the patent system*, Boston College L Rev. Vol. 61 (forthcoming, 2020) available at https://ssrn.com/abstract=3465737.

Chien, C. V. (2018). *Comparative Patent Quality. Arizona St LJ, 50,* 71.

Csardi, G., Tobochnik, J., Erdi, P., Zalányi, L., & Strandburg, K. (2009). Patent citation networks revisited: Signs of a twenty-first century change? *North Carolina L Rev, 87,* 1657.

Dahlin, K. B., & Behrens, D. M. (2005). When is an invention really radical?: Defining and measuring technological radicalness. *Research Plolicies, 34*(5), 717.
Érdi, P., Kinga, M., Zoltán, S., Katherine, S., Jan, T., Péter, V., et al. (2013). Prediction of emerging technologies based on analysis of the U.S patent citation network. *Scientometrics*, *95*(1), 225–242.
Feldman, R. (2018). May your drug price be evergreen. *Journal of Law and Biosciences, 5,* 590.
Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Sciences, 47,* 117.
Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Rearch Policy, 30,* 1019.
Frakes, M., & Wasserman, M. F. (2019). Irrational ignorance at the patent office. *Vanderbilt L Rev, 72,* 975.
Fromer, J. C. (2016). *Dynamic Patent Disclosure. Vand L Rev, 69,* 1715.
Gambardella, A., P. Giuri, and M. Mariani, *The value of European patents. Evidence from a survey of European inventors*, final report of the PatVal-EU project. Pisa: Laboratory of Economics and Management (LEM), Sant' Anna School of Advanced Studies. Retrieved from www.alfonsogambardella. it/patvalfinalreport.pdf.
Gandal, N., & Cohen, S. *Networks and spillovers in software in Israeli Hi-Tech,* CEPR discussion paper #13467. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3319788.
Glänzel, W., & Meyer, M. (2003). Patents cited in the scientific literature: An exploratory study of Ereverseí citation relations. *Scientometrics, 58,* 415.
Golden, J. M. (2010). Innovation dynamics, patents, and dynamic-elasticity tests for the promotion of progress. *Harvard Journal of Law & Technology, 24,* 47.
González, A. G. (2006). The software patent debate. *Journal of Intellectual Property Law and Practice, 1*(3), 196.
Grigg, K., *The evolution of the patent and trademark depository library and the role of the patent and trademark librarian in the digital age* (2003), Retrieved from https://peer.asee.org/12462.
Grriliches, Z. (Ed.). (1984). *R&D patents and productivity*. Chicago: University of Chicago Press.
Hall, B. (2005). Exploring the patent explosion. *The Journal of Technology Transfer, 30,* 35–48.
Hall, B., Griliches, Z., & Hausman, J. (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica, 52,* 909.
Hall, B., Jaffe, A. Jaffe, and Trajtenberg, M., *Market Value and Patent Citations: A First Look*, RAND Journal of Economics 36: (2005).
Hall, B., Jaffe, A., & Trajtenberg, M. (2003). The NBER patent citations data file: Lessons insights and methodological tools. In A. Jaffe, B. Adam, & M. Trajtenberg (Eds.), *Patents, citations, and innovations: A window on the knowledge economy*. Cambridge: MIT Press.
Harhoff, D. F., Narin, F., Scherer, F. M., & Vopel, K. (1999). Citation frequency and the value of patented inventions. *Review of Economics and Statistics, 81*(3), 511–515.
Hemphill, C., Scott, S., & Bhaven, N. (2012). Evergreening, patent challenges, and effective market life in pharmaceuticals. *Journal of Health Economics., 31,* 327.
Jaffe, A. B., & Trajtenberg, M. (2003). *Patents, citations & innovations: A Window on the knowledge economy*. Cambridge: MIT press.
Kapczynski, A., Park, C., & Sampat, B. (2012). Polymorphs and prodrugs and salts (Oh My!): An empirical analysis of secondary pharmaceutical patents. *PLoS ONE, 7,* 49470.
Kelley, D. J., Ali, A., & Zahra, S. A. (2013). Where do breakthroughs come from? Characteristics of high-potential inventions. *Journal of Product Innovation Management, 30*(6), 1212.
Kuhn, Jeffrey M., Younge, Kenneth A., and Marco, Alan, *Patent Citations Reexamined: New Data and Methods*, available at http://dx.doi.org/https://doi.org/10.2139/ssrn.2714954 (2018)
Kuhn, J., & Younge, K. (2016). Patent citations: An examination of the data generating process. *Academy of Management Proceedings, 1,* 18173.
Landes, W. M., & Ponser, R. A. (2003). *The economic structure of intellectual property law*. Cambridge: Harvard University Press.
Lanjouw, J. O., & Schankerman, M. (2001). Characteristics of patent litigation. *The RAND Journal of Economics, 32,* 129.
Lanjouw, J. O., & Schankerman, M. (2004). Patent quality and research productivity: Measuring innovation with multiple indicators. *The Economic Journal, 114,* 495.
Larviere, V., Gingras, Y., & Archmbault, E. (2009). The decline in the concentration of citations, 1900–2007. *Journal of the American Society of Information Science and Technology, 60*(4), 858.
Lemley, M. A. (2001). Rational Ignorance at the Patent Office. *SSRN Electronic Journal, 95,* 1495.
Lemley, M. A. (2013). Fixing the patent office. *Innovation Policy and the Economy, 13,* 83.
Lemley, M. A., & Sampat, B. (2008). Is the patent office a rubber stamp? *Emory LJ, 58,* 181.
Masur, J. (2011). Patent inflation. The. *Yale Law Journal, 121,* 470.

Merges, R. P. (1999). As many as six impossible patents before breakfast: Property rights for business concepts and patent system reform. *Berkeley Technology Law Journal, 14,* 577.

Mehta, A., Rysman, M., & Simcoe, T. (2010). Identifying the age profile of patent citations: New estimates of knowledge diffusion. *Journal of Applied Econometrics, 25*(7), 1179–1204.

Motohashi, K. (2008). Licensing or not licensing? An empirical analysis of the strategic use of patents by Japanese firms. *Research Policy, 37,* 1548–1550.

Moore, K. A. (2005). Worthless patents. *Berkeley Technology Law Journal, 20,* 1521.

Nemet, G. F., & Johnson, E. (2012). Do important inventions benefit from knowledge originating in other technological domains. *Research Policy, 41*(1), 190.

Ouellette, L. L. (2010). How many patents does it take to make a drug? Follow-on pharmaceutical patents and university licensing. *Michigan Telecommunication and Technology Law Review, 17,* 299.

Ouellette, L. L. (2012). Do patents disclose useful information? *Harvard Journal of Law & Technology, 25,* 545.

Pedraza-Farina, L., & Whelan, R. (2020). A network theory of patentability. *University of Chicago Law Review, 87,* 63.

Phene, A., Fladmoe-Lindquist, K., & Marsh, L. (2006). Break through innovations in the U.S. biotechnology industry: The effects of technological space and geographic origin. *Strategic Management Journal, 27*(4), 369.

Ponchek, T. (2015). Does the patent system promote scientific innovation—empirical analysis of patent forward citations. *Albany Law Journal of Science and Technology, 25,* 289.

Saunders, K. M. (2002). Patent nonuse and the role of public interest as a deterrant to technology supression. *Harvard Journal of Law and Technology., 15,* 389–390.

Schmookler, J. (1966). *Invention and economic growths*. Cambridge: Harvard University Press.

Schoenmakers, W., & Duysters, G. (2010). The technological origins of radical inventions. *Research Policy, 39,* 1051.

Shur-Ofry, M. (2016). Access to error. *Cardozo Arts & entertainment Law Review, 34,* 357.

Shur-Ofry, M. (2017). Connect the dots: Patents and interdisciplinarity. *Michigan Journal of Law Reform, 51,* 55.

Sichelman, T. (2010). Commercializing patents. *Stanford Law Review, 62,* 341.

Strandburg, K. J., Csárdi, G., Tobochnik, J., Érdi, P., & Zalányi, L. (2006). Law and the science of networks: An overview and an application to the patent explosion. *Berkeley Tech LJ, 21,* 1293.

Tong, X., & Davidson, J. (1994). Measuring national technological performance with patent claims data. *Research Policy Journal, 23,* 133.

Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovation. *RAND Journal of Economics, 21,* 172.

Trajtenberg, M., Manuel, H., Manuel, Jaffe, & Adam, B. (2013). University versus corporate patents: A window on the basicness of invention. In Jafe, B. Adam, & Trajtenberg (Eds.), *Patents citations innovations: A window on the knowledge economy*. Cambridge: MIT Press.

Noorden, V. (2017). Richard, *the science that's never been cited*. *Nature, 552,* 163.

Verbandt, Y., & Vadot, E. (2018). Non-patent literature search at the European patent office. *World Patent Information, 52,* 72.

Yoshikane, F., Suzuki, Y., & Tsuji, K. (2012). Analysis of the relationship between citation frequency of patents and diversity of their backward citations for Japanese patents. *Scientometrics, 92*(3), 721.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316,* 1036.