An attempt to identify technologically relevant papers based on their references

Yasuhiro Yamashita[1]

Abstract

In this study, two indicators derived from references in papers were proposed to characterize the papers regarding technological relevance: (1) the number of reference papers that obtained citations from patents by the time of observation, i.e., the publication years of papers to be assessed (NR-PCP), and (2) the number of reference papers authored by the firms' researchers (NR-FP). Next, the two indicators were applied to papers published in 2001 to assess their performance. The results obtained by the two indicators were evaluated by citations from patents until 2016 in various conditions: scientific field, institutional sector, and period of measurement. Results showed a robustness of both indicators in many conditions. NR-PCP showed better results in most cases than NR-FP, although its recall was inferior to NR-FP for papers in which all references were newer than 1996. Based on the result that NR-PCP was preferred as an indicator, the rationale of using reference papers cited in the patent by the period of observation (R-PCP) as an indicator was considered based on the papers' potential distances from the border between science and technology, which was obtained from an extended version of the citation network originally proposed by Ahmadpoor and Jones (Science 357:583–587, 2017. https://doi.org/10.1126/science.aam9527). Finally, issues to be addressed were discussed.

Keywords Patent-paper citations · References · Indicator · Technological relevance

Introduction

Today, information on the latest research output is essential in the research planning and formulation of science, technology, and innovation strategies. For example, the European Union's Future Emerging Technology program tried to utilize bibliographic coupling clustering of the newest papers in its horizon scanning activity (Warnke et al. 2016). Actually, they could not utilize the result of the activity because of the huge computation time; they utilized instead the Research Fronts database provided by Clarivate Analytics, which contained various indicators of co-citation clusters of the recent

✉ Yasuhiro Yamashita
yasuhiro.yamashita@jst.go.jp

[1] Japan Science and Technology Agency, Science Plaza, 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-8666, Japan

6 years' highly cited papers (Clarivate Analytics refers to them as "research fronts"). The author's organization (Japan Science and Technology Agency) also introduced the Research Fronts and Hot Papers provided by Clarivate Analytics, consisting of recently published papers that obtained the top 0.1 percent of citations during the last two months, to find emerging research topics.

On the other hand, scientific research has been expected to be a driving forth of innovation. For example, the 5th Japanese Science and Technology Basic Plan for FY 2016–2020 listed four pillars: (1) acting to create new value for the development of future industry and social transformation; (2) addressing economic and social challenges; (3) reinforcing the "fundamentals" for STI (science, technology, and innovation); and (4) building a systemic virtuous cycle of human resources, knowledge, and funding for innovation (Council for Science, Technology and Innovation 2015). To formulate research plans in accordance with such innovation-oriented goals, the current situation concerning the relationship between science and technology should be monitored.

Citation linkages between patents and papers have been deemed as a quantitative indicator of relevance of patented inventions and scientific knowledge since Narin and his colleagues utilized science linkage (e.g., Narin and Noma 1985; Narin 1991). As illustrated in van Raan's review article (van Raan 2017a), many studies were published on various aspects concerning relationships between science and technology. However, most studies were done retrospectively based on existing citation linkages from patents to papers (hereafter, PPCs). As van Raan found, observation of PPCs requires 3–20 years, depending on the fields' characteristics, and only a small fraction of papers were cited in patents. Such nature of PPCs hampered their use in research planning.

Tijssen (2010) defined three domains of journals' orientation of application based on rate of authors representing specific institutional sectors (i.e., clinical, industrial, and civic). By extension, he classified journals into six categories by their so-called "journal application domain" based on a combination of two of those domains—clinical and industrial—which together. These categories can be used as a proxy for technological relevance. He also identified 25 application-oriented fields of science (AoFs), which Hu and Rousseau (2018) later used to analyze the transfer of knowledge from discovery-oriented science to technology via AoFs. More recently, Bikard and Marx (2019) developed two measures of papers' applicability: appliedness of the paper and the Journal Commercial Impact Factor (JCIF). The former is calculated based on pairs of keywords derived from the citing patents and the cited papers. It needs 5-year citation window; therefore, it is not be suitable for application to the newest papers. On the other hand, the latter, JCIF, is derived by the same formula as the Journal Impact Factor, using PPCs instead of paper-paper citations. They showed JCIF's positive correlation with PPCs. Considering PPCs' relatively strong bias to specific journals as van Raan (2017a) stated, Tijssen's classification system and JCIF can be expected to be used for characterizing relatively young papers. However, because papers in multidisciplinary journals should tend to obtain citations from patents in ways that differ according to their specific fields, methods of characterization in individual papers should be explored as well. Methods of characterization in individual papers should be explored as well.

Many researchers utilized regression analysis to shed light on factors that positively correlate to PPCs (such as Fukuzawa and Ida 2016; Veugelers and Wang 2016; Yamashita 2018). However, coefficients they obtained could not be directly extrapolated to other objects or periods. Moreover, each independent variable used in their studies did not represent "technological relevance" itself.

Therefore, alternative feature values of papers' technological relevance, which can be obtained at time of publication or a short time afterwards, should be identified. As Gingras (2014) argued, an indicator should "correspond to the object (or concept) being evaluated." The author attach great importance to new indicators being simple and intuitive. In developing indicators, I assumed that "technologically relevant" papers frequently (but not always) obtained citations from patents, while "technologically non-relevant" papers rarely obtained them.

In this study, I attempted to utilize reference information (backward citations). References contain rich information regarding background knowledge on which papers are based. Although recent studies shed light on many aspects of PPCs as mentioned above, the relationship between "technologically relevant" references of papers as assessed and their future citations from patents were still unknown. By exploring citation-based indicators, the author looked to shed light on the nature of "technologically relevant" references as well.

There was another approach to investigating papers' similarity to patents. As to a content-level relationship, Kajikawa and his colleagues (Shibata et al. 2010; Ogawa and Kajikawa 2015) utilized clustering by direct citations within papers and patents, respectively, then matched keywords of paper clusters and those of patents to each other to calculate similarities between them. By utilizing text similarities between papers and patents, the time lag caused by PPCs was avoided. Their method should be reasonable to make a strategic plan for a specific research area. However, if one wishes to execute a calculation of direct citation linkage for the whole collection of citation index databases, such as the Web of Science (WoS) produced by Clarivate Analytics, large computational resources are required. Although the author utilized custom data compiled in-house in this study, there are some existing databases providing PPC data, such as Marx and Fuegi (2020). Therefore, researchers do not have to compile PPC data themselves. Thus, a method based on PPCs might also be useful for overviewing the research trend in macro level.

Here, I explored papers' feature values, which should correlate to their future possibility of being cited in future patents, proposed two indicators based on the papers' references, and compared them to show which is preferred as an indicator of technological relevance. This paper is an extended version of Yamashita (2019), originally presented at the 17th International Conference on Scientometrics and Informetrics in Rome, Italy.

Methodology and data

Definitions of indicators

In this study, a simple count of reference papers regarded as technologically relevant was adopted for the calculation of indicator values of the papers to be assessed (hereafter, "focal papers"). This was based on the idea that the closer a paper's content was to technology, the more technologically relevant knowledge was needed for the paper's background. I focused on the following two features of papers as indicators of technological relevance. The indicator values were measured at the end of the focal papers' publication year (in this research, it was set to 2001), because two indicators were aimed to be measured just after their publication.
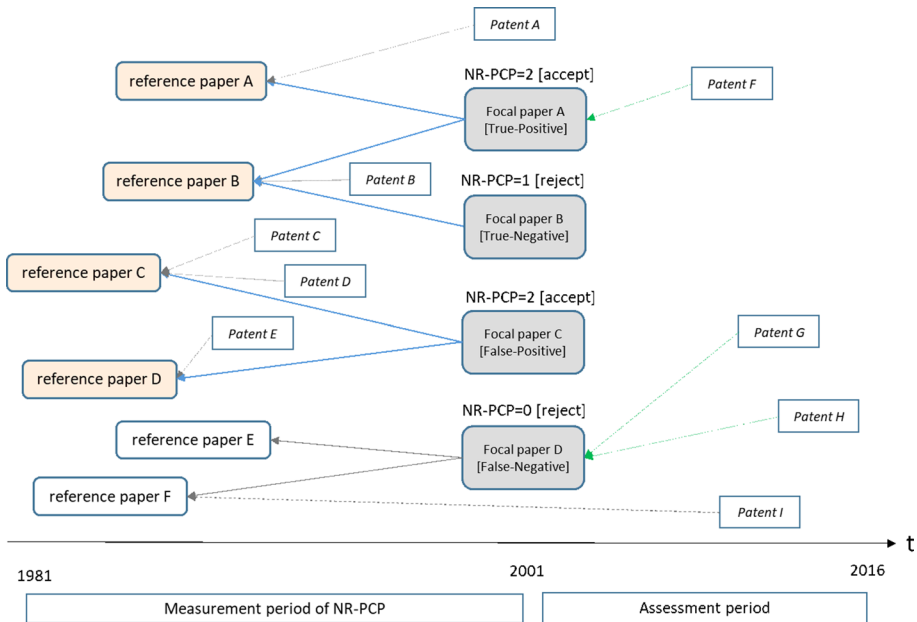
Fig. 1 Scheme for measuring NR-PCP (case of threshold = 2)

1. Reference papers that were cited in the patents by the period of observation (R-PCPs).
    If it is assumed that papers cited in patents (PCPs) have singular characteristics of
    the border between science and technology, as Ahmadpoor and Jones (2017) defined,
    then candidates for PCPs that have the same nature as PCPs might frequently need them
    as background knowledge for their research. As van Raan (van Raan 2017a, b) utilized
    citing papers of specific PCPs (he called it SNPR) to study diffusion of research topics
    centered on PCPs, papers citing PCPs (i.e., papers referencing PCPs) partially share
    research topics with the PCPs to which they refer. The number of R-PCPs in focal papers
    (NR-PCP), an indicator based on R-PCP, was obtained, as shown in Fig. 1 (threshold
    value was set to 2). In Fig. 1, R-PCPs are colored in orange, citation linkages contribut-
    ing to calculation of NR-PCP values are thick blue arrows, and green arrows are citation
    linkages used to decide whether focal papers were future PCPs or not.

    (a) Focal paper A referred to two papers (reference papers A and B) that were cited
        by patents published during the period of measuring NR-PCP. Thus, its NR-PCP
        was 2, which was equal to the threshold value. Focal paper A was deemed to be
        a predicted "technologically relevant" paper (hereafter "candidate paper"), and
        it was cited in a patent by 2016 (within the period of assessment); therefore, the
        result was "true-positive."
    (b) Focal paper B referred to one paper (reference paper B) that was cited by patent
        B; hence, its NR-PCP = 1 was below the threshold value 2. Therefore, it was not
        deemed to be a candidate paper. Focal paper B was not cited by any patent during
        the period of assessment; therefore, the result was "true-negative."
    (c) Focal paper C referred to two papers cited by three patents (patents C, D, and E).
        How many citations each reference paper obtained was ignored (only cited or not

was taken into account). Therefore, its NR-PCP was 2 (accepted). But it was not cited by any patent afterwards. The result was "false-positive."

(d) Focal paper D referred to two reference papers that obtained no citations from patents published during the measurement period. Actually, reference paper F was cited by patent I, which was published outside the measurement period. This citation was not counted because the NR-PCP value was measured at the end of 2001. Therefore, focal paper D was not deemed a candidate paper. However, it was cited by patents G and H during the assessment period. The result was "false-negative."

2. Reference papers written by firms' authors (R-FPs).

Tijssen (2010) considered the rate of authors from the industrial sector in the journals as a proxy for industrial relevance. In the author's previous study of Japanese papers, firms showed relatively strong positive correlations to PPCs (Yamashita 2018). Papers written by researchers in the firm sector can be assumed to be close to technology in nature as firms are the core sector of technological development. Therefore, it can be assumed that PCP candidates frequently need firms' research as part of their background knowledge. The number of R-FPs (NR-FP), an indicator based on R-FP, was obtained, as shown in Fig. 2, where R-FPs are colored orange, citation linkages contributing to the calculation of NR-FP are presented as thick blue arrows, and citation linkages used to assess the result are presented as green arrows, as in Fig. 1.

(a) Focal paper A referred to three reference papers (co)authored by firms; thus, its NR-FP was 3. This value was greater than the threshold of 2; therefore, it was deemed a candidate paper (accepted). Focal paper A was cited in a patent by 2016; therefore, the result was "true-positive."
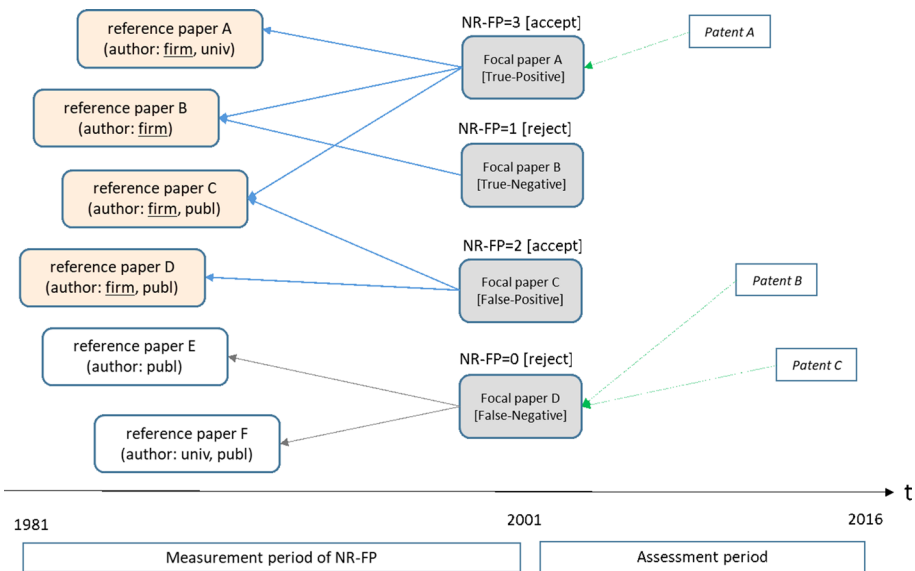

Fig. 2 Scheme for measuring NR-FP (case of threshold = 2)

(b)   Focal paper B referred to one reference paper (C) which was (co)authored by firms; thus, its NR-FP was 1. This value was below the threshold value of 2. Therefore, focal paper B was not deemed a candidate paper. This paper was not cited in patents during the assessment period, and the result was "true-negative."

(c)   Focal paper C referred to two reference papers (co)authored by firms, so its NR-FP was 2 (accepted). However, it obtained no citation from patent during the assessment period. Therefore, the result was "false-positive."

(d)   Focal paper D referred to no papers authored by firms, only by universities and public institutes. Therefore, its NR-FP was 0; however, it was cited by two patents during the assessment period. The result was "false-negative."

Results were assessed from both perspectives: (1) the number of focal papers judged as "true-positive" divided by the number of papers judged as either "true-positive" or "false-positive" (precision), and (2) the number of focal papers judged as "true-positive" divided by the number of focal papers judged as either "true-positive" or "false-negative" (recall). Usually, they show opposing relationships to each other; therefore, their harmonic mean (i.e., F-measure) was used to show the "balanced" total score of the two measures for convenience.

Data

Papers published from 1981 to 2001 available from Web of Science (WoS) were used as paper data. Among them, papers published in 2001 were used as focal papers, and those published between 1981 and 2001 were used as sources for additional references. For both focal and reference papers, the document type searched for was limited to "Article." All papers were classified into 22 categories within WoS's Essential Science Indicators (ESI) by following a six-step procedure. (1) Papers recorded in ESI were classified according to their categories in ESI; (2) if journals in which the papers were published were listed in ESI, then the papers were classified by their specific journal's categories in ESI; (3) the most frequent their citing papers' categories were decided as their category; (4) the most frequent their references' categories were deemed as their categories; (5) the most frequent Steps 3 and 4 were deemed as their categories; otherwise, (6) papers were classified into the category "Multidisciplinary science." Papers without references or author affiliation(s) were excluded from the focal papers. In total, 716,584 papers were submitted for analysis.

For citing patent data, the 2016 spring edition of Patstat was used. All patent authorities and citation kinds in the Patstat were included in the analysis, since "technological relevance" was taken in broad meaning in this study. The type of intellectual property rights of patents was limited to the "patent of invention" (PI) to unify the statistical nature of the PPC. All non-patent literature appearing in the Patstat was matched to each paper in the WoS to obtain the PPC linkages. Therefore, the data contained paper citations from patents published until early 2016.

The identification of three sectors (firm, university, and public institute) was indispensable for deriving NR-FP indicators and analyzing their effects. This was executed based on a data table that the author and colleagues elaborated to classify the world's organizations, and organizations not covered in the table were classified based on the keywords shown in Table 1. According to the classification process, of the 12,166,213 papers published from 1981 to 2001, 890,451, 9,041,035, and 1,654,871 papers covering both focal and reference

Table 1 Keywords for identifying the three sectors in the papers

Sector	Keyword or description
Firm (suffix)	INC, LTD, CORP, CO, GMBH, AG, GRP, SA, AB, SPA, PLC, AS, MBH, BV, KG, KK, PC, LLC, NV, OY, LLP, SRL, R&D, KGAA, SRO, LTDA, CONSULTANTS, CONSULTING, CONSULTANCY, CONSULTANT, CONSULT, GROUP, COMPANY, LP, LIMITED, SARL, SAPA, VOS
University	UNIVERSITY, UNIV, UNIVERSITE, UNIVERSITAT, ECOL, ECOLE, ECOLES, MED SCH, COLL, INST TECHNOL, CHU, CHRU, TH, ETH, FAC, GRAD SCH, POLYTECH, POLYTECN
Public Institute	NATL, ACAD, FED, NACL, NATL, MINIST, EUROPEAN, GOVT, BUNDES, CITY, MUNICIPAL, PUBL, PREFECTURAL
Other	Organizations not classified in the three sectors above, such as hospitals, non-profits and unknown.

papers were attributed to firm, university, and public institute sectors, respectively (including duplication caused by co-authorship across different sectors).

Comparison of NR-PCP and NR-FP

Threshold values and obtained results

Figure 3 shows three assessment measures (precision, recall and F-measure) of the results obtained by applying NR-PCP and NR-FP (hereinafter, the set of candidate papers obtained
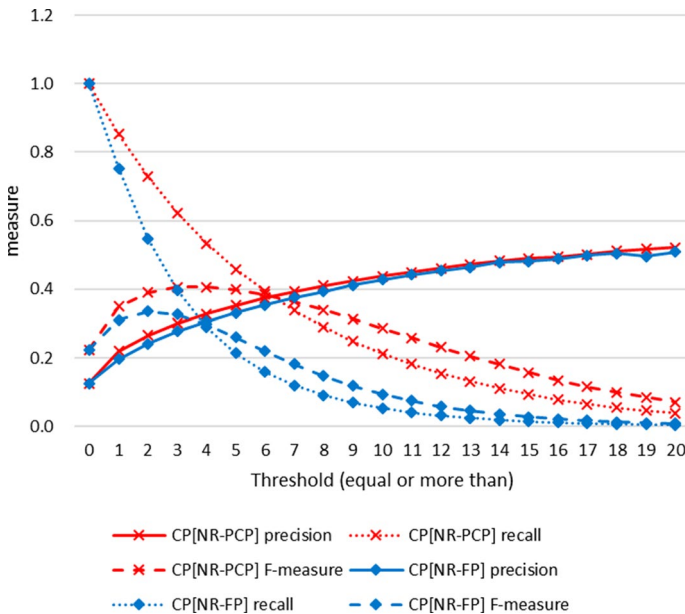

Fig. 3 Precision, recall and F-measure of results by threshold value (CP[NR-PCP] and CP[NR-FP])

by applying NR-PCP and NR-FP was called CP[NR-PCP] and CP[NR-FP], respectively) to the focal papers published in 2001. The resulting dataset was identified by extracting papers of which indicator values were equal to or above a threshold value (X-axis). Y-axis designated the values of the three assessment measures. Therefore, this figure illustrates when a specific threshold value (the lowest limit of NR-PCP and NR-FP) was selected and, to what extent the accurately predicted dataset of future PPC would be obtained.

The precision of results increased along with the increase in threshold values of both NR-PCP and NR-FP, as shown in Fig. 3. Although the two indicators did not necessarily indicate identical entities, their precision curves were similar (CP[NR-PCP] was slightly superior to CP[NR-FP] in all ranges of threshold values below 20). If the papers containing at least one R-PCP (or R-FP) were deemed technologically relevant (i.e., threshold = 1), then CP[NR-PCP] and CP[NR-FP] marked precision of 0.220 and 0.196, respectively, which were much higher than the rate of PCPs in all papers (Hereafter, I refer to the rate of PCPs in the sample as "sample PCP rate". Precision at threshold = 0 is always identical to sample PCP rate), which was 0.126. The precision of results from NR-PCP and NR-FP indicators reached 0.521 and 0.509, respectively, when threshold values were set to 20.

As for recall, CP[NR-PCP] showed values that were much higher than CP[NR-FP] in all ranges of threshold values, as shown in Fig. 3. Recall of CP[NR-FP] rapidly decreased with an increase in the threshold of NR-FP. Relatively low values of recall might be partially caused by the coverage of firms' papers.

Generally, NR-PCP showed better results than NR-FP, since F-measures of CP[NR-PCP] were much higher than those of CP[NR-FP]. The maximum value of the F-measure for CP[NR-PCP] was 0.406 (threshold = 4; precision = 0.328, recall = 0.533), while that of CP[NR-FP] was 0.335 (threshold = 2; precision = 0.242, recall = 0.546). F-measure of the whole sample was 0.223 (see F-measure value of both indicators at threshold = 0); therefore, NR-PCP and NR-FP could improve F-measure by 0.183 and 0.112 from that of the sample, respectively.

How did the results obtained by both indicators overlap? The Spearman's rank correlation coefficient between NR-PCP and NR-FP was 0.525. The inclusion index between the candidate papers was 0.70, and that of the correct answers between them was 0.91 at threshold = 1. These inclusion indexes decreased as threshold values increased; nevertheless, inclusion indexes measured at threshold = 20 remained at 0.48 and 0.59, respectively. Therefore, both indicators modestly correlated with each other.

Although the two indicators showed positive correlation to precision in Fig. 3, we should know whether NR-PCP and NR-FP were essential factors to predict the events of papers becoming PCP in the future. Therefore, logistic regression analyses were executed to show to what extent NR-PCP and NR-FP could explain future citation from patents. For comparison, the case of all references included as independent variable was also shown.

As control variables, the 2001 edition of Journal Impact Factor compiled by Clarivate Analytics (proxy of scientific quality), reprint authors' region, authors' institutional sectors, and scientific fields were included in the models. Considering citation behavior of patents, reprint authors' regions were decided by the trilateral patent office, the United States, European Patent Office member states, Japan, and others, according to Veugelers and Wang (2016). As EPO member states, 38 countries belonging to it as of January 2020 were used.

Results are shown in Table 2. NR-PCP showed a relatively high coefficient (0.7454), which was higher than that of impact factor (Model 1). Model 1 marked the highest pseudo R^2 value of the three models shown in Table 2. In Model 2, pseudo R^2 was below that of

Table 2 Logistic regressions of probability of being PCP by 2016

Variable	Model 1 (Cited by 2016)	Model 2 (Cited by 2016)	Model 3 (Cited by 2016)
NR-PCP (normalized)	0.7454*** (0.0056)	–	–
NR-FP (normalized)	–	0.4081*** (0.0048)	–
Number of references (normalized)	–	–	0.1398*** (0.0061)
(Intercept)	− 2.2681*** (0.0255)	− 2.0461*** (0.0247)	− 2.0208*** (0.0243)
Impact factor (normalized)	0.3981*** (0.0058)	0.5238*** (0.0055)	0.5542*** (0.0057)
Sector			
University	− 0.2356*** (0.0190)	− 0.1823*** (0.0186)	− 0.2550*** (0.0183)
Firm	0.7157*** (0.0163)	0.5398*** (0.0163)	0.8717*** (0.0157)
Publ. Inst.	− 0.0823*** (0.0156)	− 0.0911*** (0.0152)	− 0.1057*** (0.0151)
Reprint authors' region			
The United States	0.3435*** (0.0145)	0.3132*** (0.0142)	0.3313*** (0.0141)
Europe	0.1207*** (0.0144)	0.0611*** (0.0141)	0.0542*** (0.0140)
Japan	0.1409*** (0.0180)	0.1479*** (0.0176)	0.2080*** (0.0174)
Field			
Agricultural Sciences	0.5296*** (0.0345)	0.3559*** (0.0337)	0.3421*** (0.0334)
Biology and Biochemistry	0.1051*** (0.0192)	0.3566*** (0.0186)	0.4118*** (0.0184)
Chemistry	0.5660*** (0.0165)	0.3520*** (0.0160)	0.4250*** (0.0158)
Computer Science	1.2348*** (0.0333)	1.0429*** (0.0325)	1.1805*** (0.0325)
Engineering	0.8457*** (0.0232)	0.4975*** (0.0225)	0.5895*** (0.0227)
Immunology	0.1242*** (0.0308)	0.2955*** (0.0302)	0.5163*** (0.0294)
Microbiology	0.1046** (0.0327)	0.5064*** (0.0316)	0.5227*** (0.0311)
Molecular Biology and Genetics	− 0.3586*** (0.0256)	0.0241 (0.0250)	0.0320 (0.0248)
Materials Science	0.7238*** (0.0256)	0.4322*** (0.0250)	0.5989*** (0.0249)
Neuroscience and Behavior	− 0.3501*** (0.0267)	− 0.3754*** (0.0261)	− 0.3925*** (0.0259)
Pharmacology and Toxicology	0.5049*** (0.0270)	0.4206*** (0.0266)	0.6213*** (0.0261)
Physics	− 0.0196 (0.0212)	− 0.5727*** (0.0204)	− 0.5160*** (0.0203)

Table 2 (continued)

Variable	Model 1 (Cited by 2016)	Model 2 (Cited by 2016)	Model 3 (Cited by 2016)
Plant and Animal Science	− 0.2383*** (0.0298)	− 0.2981*** (0.0289)	− 0.4142*** (0.0286)
Observations	370,747	370,747	370,747
Pseudo R^2	0.155	0.118	0.098

Countries of reprint authors neither the United States, Europe nor Japan were classified into "Other" region, and used as reference

For the field, clinical medicine was used as reference

NR-PCP, NR-FP and Impact Factor were neglog-transformed, then normalized into z-scores

Signif. codes: 0 ***' 0.001 ** 0.01 * 0.05 '.' 1

Cases were restricted to those that matched the following conditions:

Impact factor Papers in which journals had impact factors in 2001. Impact factors were obtained from the Journal Citation Report compiled by Clarivate Analytics

Sector Papers in which all authors were affiliated with either a university, public institute, or firm (i.e., papers (co)authored by "other" sector were excluded)

Field Field in which number of papers cited by patents by 2016 (i.e., correct answer) was greater than one thousand. Multidisciplinary field was excluded

Model 1, and coefficient of NR-FP (0.4081) was also below that of NR-PCP. However, pseudo R^2 and coefficient of NR-FP of Model 2 were above pseudo R^2 and "number of references" of Model 3, respectively. Therefore, NR-FP had more predictive power for future PPCs than whole references, but it was relatively weaker than that of NR-PCP. In all three models, firms' papers tended to be more cited than those of the other two sectors, this tendency also gave rationale for NR-FP.

Tendencies by scientific field

Whether the two indicators, NR-PCP and NR-FP, work reliably in any scientific field is essential for their actual use. Figure 4 showed maximum F-measure obtained in each scientific field by the application of the two indicators. Gray bars in the right side designate F-measure of the whole sample. The line graph in Fig. 4 designated the sample PCP rate in each field.

All cases of both CP[NR-PCP] and CP[NR-FP] showed improvement from the F-measure of the whole samples, and CP[NR-PCP] outperformed CP[NR-FP]. While for scientific field those rates of PCPs were relatively high (i.e., sample PCP rate exceeded 0.2), showing limited improvement of F-measures, some fields of the medium rate of PCP (i.e., sample PCP rates between 0.05 and 0.2) showed relatively large improvement. Of such fields of the medium sample PCP rates, CP[NR-PCP] showed a relatively high improvement in Plant and Animal Sciences (0.255), Clinical Medicine (0.171), and Physics (0.164). Furthermore, CP[NR-FP] showed a relatively large improvement in Plant and Animal Sciences (0.136), Clinical Medicine (0.121), and Agricultural Sciences (0.091). The Multidisciplinary field, which consisted of papers that could not be classified into any of the 21 fields, also showed a high improvement of F-measures for both CP[NR-PCP] (0.211) and CP[NR-FP] (0.159).
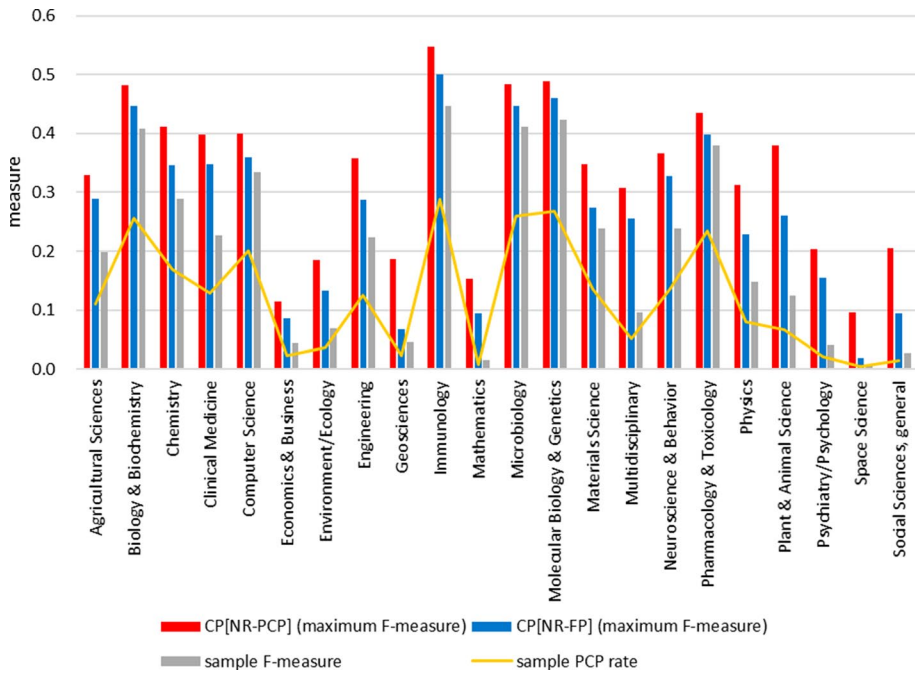
Fig. 4 Maximum F-measures of CP[NR-PCP] and CP[NR-FP], sample F-measure and sample PCP rate by scientific field

Tendencies by authors' sector of focal papers

To clarify whether NR-PCP and NR-FP could be applied to any sector, I showed trends in three major sectors: firms, universities, and public institutes (Fig. 5). To eliminate the influence of sectors other than focal sectors, papers written by only focal sectors were counted.

All cases except the precision of CP[NR-FP] in public institutes showed similar tendencies. Precision of public institutes showed a decrease when the threshold value exceeded
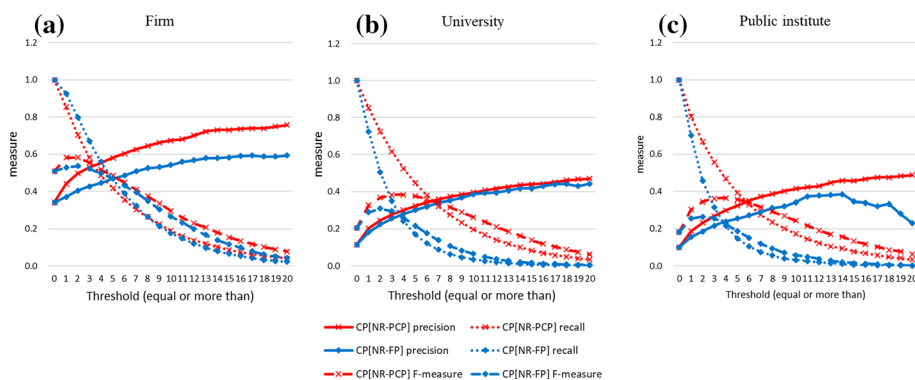
Fig. 5 Precision, recall and F-measure of results in the three sectors (CP[NR-PCP] and CP[NR-FP])

14. For public institutes, the number of focal papers in which NR-FP was more than 14 was only 27. Therefore, the decrease in precision of CP[NR-FP] was caused by the small number of identified papers. Generally, the indicators showed robust trends in small values of thresholds for all three sectors.

NR-PCP and NR-FP yielded much better results in the firm sector than in the other two sectors; their precision rates reached 0.758 and 0.593 at threshold = 20, and their maximum F-measures were 0.582 and 0.537 at threshold = 2, respectively. Therefore, both indicators worked effectively for firms' research.

Since NR-FP deemed firms' papers as technologically relevant, its result should cover most of the firms' PCPs; otherwise, the firms' papers themselves should be directly used as indicators. The firms' high recall value (92.6%) for CP[NR-FP] at threshold = 1 proved the efficiency of this indirect application of firms' papers as indicators. Even when including co-authored papers with other sectors, the recall value of firms' CP[NR-FP] was as high as 90.0% at threshold = 1 (not presented).

Influence of recency of references on the results

NR-PCP depends on reference papers' forward citations from patents. It suggests that papers that consist of relatively young references do not tend to contain R-PCP, since all references have a minimal chance of obtaining citations from patents by the publication year of focal papers (i.e., 2001).

Figure 6 shows how precision, recall and F-measures obtained at threshold = 1 by applying both indicators changed by the oldest publication years of reference papers in focal papers. It clearly shows that CP[NR-PCP] sharply decreased its F-measure in the range that the oldest publication year of references was equal to or later than 1999, while CP[NR-FP]
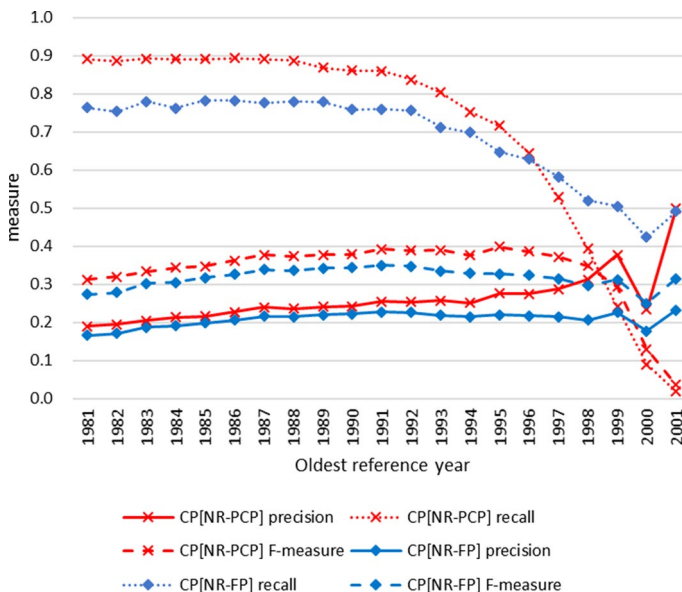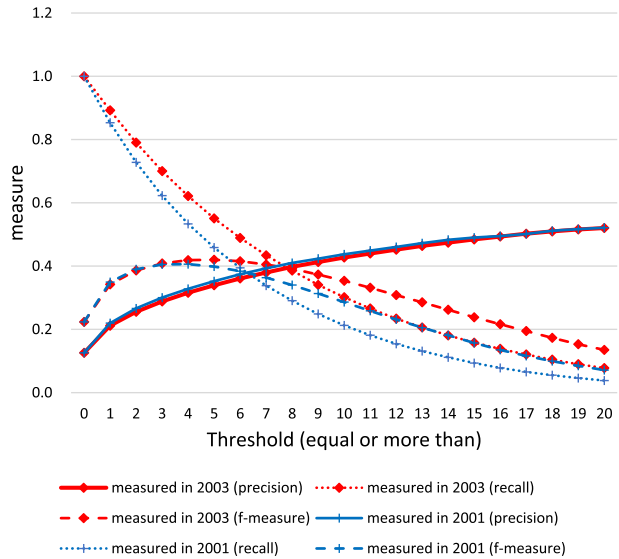

Fig. 6 Precision, recall and F-measure of results by the oldest reference year of focal papers (CP[NR-PCP] and CP[NR-FP], threshold = 1)

Fig. 7 Comparison of results measured in 2003 and 2001 (CP[NR-PCP])

showed relatively stable tendency regardless of the distribution of publication years of references. Therefore, NR-PCP requires enough references published more than 2 years before the focal papers' publication year.

For practical use of indicators, enough recall values should be secured. Recall of CP[NR-PCP] fell below that of CP[NR-FP] in the range in which the oldest publication year of references was equal to or later than 1997. The rate of focal papers that only cited papers published from 1997 to 2001 was relatively small (4.6%), and their sample PCP rate was 9.1%, which was much lower than that of all the focal papers (12.6%). Therefore, the influence of the recency of references seemed to be relatively limited.

It is noteworthy that CP[NR-PCP] showed an increase in F-measure before its trend turned negative in 1996. It suggests that NR-PCP functioned well for focal papers with relatively (but not quite) new references. For NR-PCP, securing a few years to observe the PPC of reference papers should improve results. To grasp how the three measures (precision, recall, and F-measure) of the results improved by adding a short observation period, the results measured in the publication year of the focal paper (2001) and 2 years later (2003) were compared to each other (Fig. 7).

Precision was almost identical, while recall improved visibly. The difference in recall increased as the threshold increased, and reaching a peak value of 9.6 points at threshold = 7. Maximum F-measure improved slightly, from 0.406 (threshold = 4) to 0.420 (threshold = 5). The results suggested that an observation period after the publication year of focal papers ensured higher recall.

Scrutiny of the results of the comparison

Though NR-PCP and NR-FP generally showed similar tendencies, NR-PCP seemed to be preferred as an indicator in terms of stability and operability. NR-FP needs an appropriate definition and identification of firms. The author checked its robustness under

different settings (without an organization classification table or the inclusion of non-profit organizations); however, these results did not surpass those of NR-PCP (not presented). It should be noted that the combinatorial use of both NR-PCP and NR-FP (not presented) did not yield any result of which F-measure exceeded the maximum F-measure obtained by the sole use of NR-PCP (i.e., 0.406).

On the other hand, NR-PCP needed only accurate identification of citation linkages, which required no knowledge outside the databases. The reference-age-dependent nature was a shortcoming of NR-PCP. Therefore, in the cases where most of the references were published in recent years (e.g. the research area of those cycles was remarkably short, or a remarkable breakthrough occurred recently), NR-PCP might not be able to show good performance. However, its influence may visibly lessen by ensuring a few years of citation measurement after focal papers' publication.

Consideration of papers' natures in relation to the D-metric of their references

.

NR-PCP, which was based on existing PPC linkages, provided some insights into the relationship between science and technology in a citation network of patent-patent, patent-paper, and paper-paper citations. Here, I tried to clarify its conceptual position in the citation network, extending the D-metric defined by Ahmadpoor and Jones (2017). Figure 8 shows the structure of the extended Ahmadpoor and Jones network. The distances "D" from the border between science (papers) and technology (patents) were defined based on hierarchies using citation linkages in the network. Papers attributed to D-metric are illustrated in the upper left part of Fig. 8. They define $D=1$ papers as those cited by patents (i.e., PCPs), $D=2$ papers as those cited by papers of $D=1$ but not by patents; $D=3$ papers were cited by those of $D=2$ but by neither those of $D=1$ nor patents, and so on.

Papers referencing R-PCPs (PR-PCPs) in this study rarely obtained citations from papers or patents because of they were too young; therefore, they did not have any position
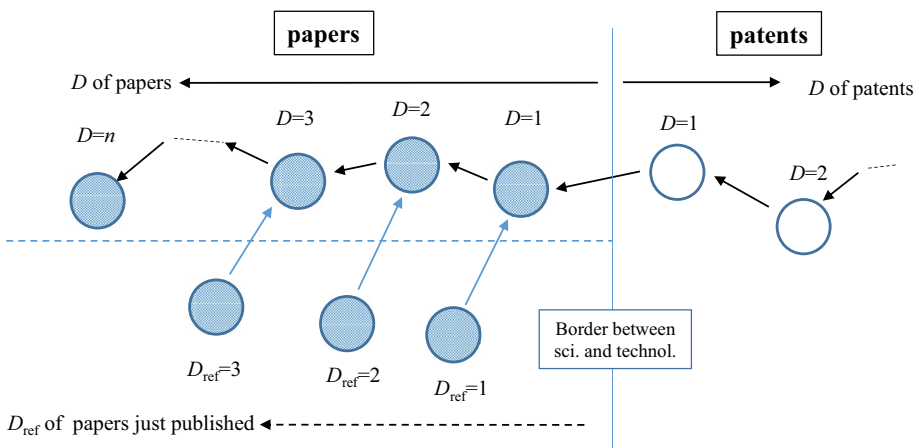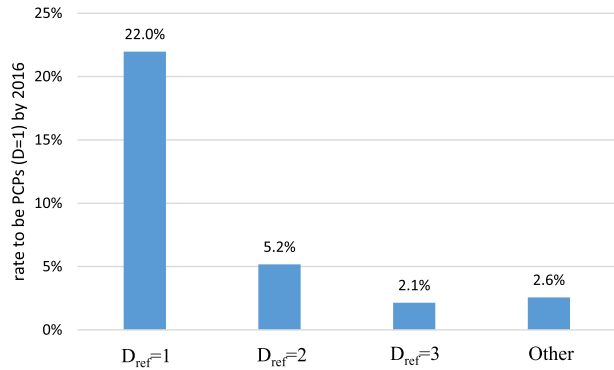

Fig. 8 Extended Ahmadpoor and Jones citation network

Fig. 9 Rate of papers published in 2001 of being PCP by 2016 according to their D_{ref}-metric (Only D_{ref}-metrics 1 to 3 were described). The reference papers' D-metric used for the calculation of D_{ref}-metric were measured in the period between 1981 and 2001

in Ahmadpoor and Jones original network in the period of their publications. However, their probability of being $D=1$ papers in the future was higher than papers published in the same year that did not cite PCPs, so their potential distance from technology was relatively closer to $D=1$. Here, I defined another type of distance from the border, D_{ref}, for stratifying recently published papers in Ahmadpoor and Jones citation network. In the case that recently published papers cited $D=1$ papers, their distance from the border of science and technology was defined as $D_{ref}=1$. Papers of $D_{ref}=n$ $(n>1)$ were defined as those cited papers of $D=n$ that did not cite any paper of D_m $(1 \leq m < n)$.

By attributing the D_{ref}-metric to recently published papers, a part of the role of reference papers in the citation network was revealed. Figure 9 shows how the rates of papers published in 2001 to be PCP $(D=1)$ by 2016 were affected by their D_{ref}-metric. It shows that the deeper the D_{ref}-metrics of papers, the less likely they would be classified as $D=1$ in the future, although the rate of "other" papers $(D_{ref}>3$, and the papers themselves or their successors not cited by patents) of being PCP by 2016 was slightly greater than that of $D_{ref}=3$.

It should be noted that the results contained some truncation noise caused by the limited observation period (21 years) of references. If a paper cited another paper of $D=1$ published before 1981, its distance from the border was not measured as $D_{ref}=1$. Therefore, the results shown in this section are rough estimates. However, even if the observation period was moved to 2010 (i.e., observation period for references was set to 1981–2010, of citations from patents to 2010–2016), tendencies remained similar, although the rate to be PCP by 2016 decreased totally (not presented).

The results suggested that papers could be stratified according to their reference papers' D-metric (i.e., D_{ref}-metric), and that R-PCP was a better predictor of future PCP than papers of $D_{ref}>1$. The results also suggest that both forward and backward citations should be considered to understand the mechanism in which PPCs occur since PPCs typically occurred near existing PCPs $(D=1)$.

Discussion and conclusion

This study attempted to develop two reference-based indicators (NR-PCP and NR-FP) of technological relevance. A comparison of both indicators' behaviors showed NR-PCP's relative advantage in steadily obtaining better results in many cases. To understand the rationale of using R-PCP as a predictor of events that lead focal papers to become PCPs

in the future, the focal papers' rate of being PCPs was analyzed by their potential distance from the border between science and technology (D_{ref}-metric), which was obtained from an extended Ahmadpoor and Jones citation network.

Those reference-based indicators should be helpful to grasp early on whether a research topic or area was sufficiently related to technology. As such, they should also support a convenient method of characterizing research areas for people responsible for science and technology planning, including funders. Even within a specific research area, the phase of the research should vary remarkably between subareas, such that funders can identify subareas of the phase of interest by applying indicators such as basic research area distant from commercialization or research area intended to be applied to specific products. By overlaying those indicators on a science map, users can grasp such information intuitively.

Two indicators provide just rough clue of future contribution to the technological application (or use for patent examination) of scientific knowledge they yielded, but not a precise prediction of future PPCs since I focused on only "input knowledge" for focal research. As many other reference-based indicators, references were deemed as proxy for characteristics of knowledge which the authors of focal papers yielded, but neither their contents nor proxy for their quality in the technological aspect.

These reference-based indicators were intended to be used for characterizing recent research papers for overviewing a research topic or area, but not for performance assessment of specific research papers or projects. Information extracted from papers' references does not directly indicate the impact of specific research and can be manipulated by the authors of papers.

The present research is a first step for the elaboration of monitoring indicators of technological relevance for young papers, and many issues should be addressed for their practical use.

1. In this research, each reference paper was regarded as a unit of knowledge, and only a simply counted number of reference papers was adopted as an indicator. However, the number of PPCs obtained by reference papers was discarded to simplify the scheme. To improve the performance obtained by NR-PCP, more sophisticated indicators that may consider the number of PPCs obtained by reference papers should be developed.

2. In addition, the likelihood of whether a specific type of paper was often missed by the indicators should be investigated. One of the possible biases was the presence or absence of the authors' and inventors' self-citations. As Tijssen et al. (2000) inferred regarding Dutch patents and papers, many self-citation linkages should be in my data. My finding that PPCs tended to occur around existing PCPs might also imply an influence of self-citation linkages. However, NR-PCP and NR-FP are aimed at measuring the introduction of technologically relevant knowledge to focal papers, not assuming the extraction of only self-citation linkages. Can the indicators appropriately predict any type of future PPCs, even if there is no self-citation? The extent to which the authors and inventors shared in focal papers, their reference papers, and patents that cited either focal or reference papers should be investigated to verify the versatility of reference-based indicators such as NR-PCP.

3. For NR-PCP, I compared results measured in the publication year and 2 years afterward. The results showed that a 2-year observation period secured a visible improvement of recall; however, the exact time that this should be measured to obtain reliable results remains unclear. The period of measurement should be determined by considering the balance between the immediacy and validity of the results.

4. The results were assessed only regarding the future PPC that papers would obtain. Nonetheless, it should only be one of the features that represent the technological relevance of papers. The indicators should be assessed from various perspectives, such as the co-occurrence of keywords between focal papers and patents, presence or absence of citations from firms' papers, and assessment using correct answers (e.g., papers regarded to have promoted technological innovation).

5. In this research, all PPCs, regardless of patent authorities and citation kind (by applicants or examiners), were included in the analysis, since a broad range of scientific knowledge was taken into account. For a more detailed understanding of the nature of technologically relevant references, an analysis should be executed by patent authority and by citation kind, although it might cause a decrease in F-measure since sample PPC rates should decrease by limiting the correct answer. In the case where patent authority was limited to USPTO only, maximum F-measure decreased from 0.406 to 0.326 for CP[NR-PCP], and from 0.335 to 0.282 for CP[NR-FP] (not presented); however, filtering by two indicators showed improvement from the decreased sample's rates.

References

Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science, 357,* 583–587. https://doi.org/10.1126/science.aam9527.

Bikard, M., & Marx, M. (2019). Bridging academia and industry: How geographic hubs connect university science and corporate technology. *Management Science*. https://doi.org/10.1287/mnsc.2019.3385.

Council for Science, Technology and Innovation. (2015). *Report on the 5th science and technology basic plan*. Retrieved December 10, 2019, from https://www8.cao.go.jp/cstp/kihonkeikaku/5basicplan_en.pdf.

Fukuzawa, N., & Ida, T. (2016). Science linkages between scientific articles and patents for leading scientists in the life and medical sciences field: the case of Japan. *Scientometrics, 106,* 629–644. https://doi.org/10.1007/s11192-015-1795-z.

Gingras, Y. (2014). Criteria for evaluating indicators. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact* (pp. 109–124). Cambridge: MIT Press.

Hu, X., & Rousseau, R. (2018). A new approach to explore the knowledge transition path in the evolution of science & technology: From the biology of restriction enzymes to their application in biotechnology. *Journal of Informetrics*, *12*(3), 842–857.

Marx, M., & Fuegi, A. (2020). Reliance on Science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal*. https://doi.org/10.1002/smj.3145.

Narin, F. (1991). Globalization of research, scholarly information, and patents—Ten-year trends. *The Serials Librarian, 21,* 33–44. https://doi.org/10.1300/J123v21n02_05.

Narin, F., & Noma, E. (1985). Is technology becoming science? *Scientometrics, 7*, 369–381. https://doi.org/10.1007/BF02017155.

Ogawa, T., & Kajikawa, Y. (2015). Assessing the industrial opportunity of academic research with patent relat-edness: A case study on polymer electrolyte fuel cells. *Technological Forecasting and Social Change, 90*(Part B), 469–475. https://doi.org/10.1016/j.techfore.2014.04.002.

Shibata, N., Kajikawa, Y., & Sakata, I. (2010). Extracting the commercialization gap between science and tech-nology—Case study of a solar cell. *Technological Forecasting and Social Change, 77,* 1147–1155. https://doi.org/10.1016/j.techfore.2010.03.008.

Tijssen, R. J. W. (2010). Discarding the 'basic science/applied science' dichotomy: A knowledge utilization triangle classification system of research journals. *Journal of the American Society for Information Science and Technology*, *61*(9), 1842–1852.

Tijssen, R. J. W., Buter, R. K., & Leeuwen, Th. N. (2000). Technological relevance of science: An assessment of citation linkages between patents and research papers. *Scientometrics, 47,* 389–412.

van Raan, A. F. J. (2017a). Patent citations analysis and its value in research evaluation: A review and a new approach to map technology-relevant research. *Journal of Data and Information Science, 2*, 13–50. https://doi.org/10.1515/jdis-2017-0002.

van Raan, A. F. J. (2017b). Sleeping beauties cited in patents: Is there also dormitory of inventions? *Scientometrics, 110*, 1123–1156. https://doi.org/10.1007/s11192-016-2215-8.

Veugelers, R., & Wang, J. (2016). Novel science for industry? *2016 IEEE international conference on management of innovation and technology (ICMIT), Bangkok, Thailand* (pp. 270–274). https://doi.org/10.1109/ICMIT.2016.7605046.

Warnke, P., Beckert, B., Bierwisch, A., Geserer, A., Kayser, V., Rung, S., & Schirrmeister, E. (2016). *OBSERVE horizon scanning methodology report*. Retrieved December 10, 2019, from https://www.horizon-obser vatory.eu/radar-wAssets/docs/WP-1_Deliverables/Deliverable-1-3_Methodology-Report.pdf.

Yamashita, Y. (2018). Exploring characteristics of patent-paper citations and development of new indicators. In M. Jibu & Y. Osabe (Eds.), *Scientometrics* (pp. 151–172). London: IntechOpen. https://doi.org/10.5772/intechopen.77130.

Yamashita, Y. (2019). Identification of technologically relevant papers based on their references. In *Proceedings of the 17th international conference on scientometrics and informetrics (ISSI2019)* (pp. 2145–2156). Sapienza University, Italy.