Impact of multidisciplinary research on innovation

David Campbell [1] Brooke Struck [2] Chantale Tippett [3] Guillaume Roberge [4]

[1] *david.campbell@science-metrix.com*
Science-Metrix (Canada)

[2] *brooke.struck@science-metrix.com*
Science-Metrix (Canada)

[3] *chantale.tippett@science-metrix.com*
Science-Metrix (Canada)

[4] *guillaume.roberge@science-metrix.com*
Science-Metrix (Canada)

Abstract

Governmental initiatives capitalising on multidisciplinary (or interdisciplinary) research are growing in number. They are motivated by the increasingly shared view among academics and policymakers that this mode of research, by favouring innovation in firms, will fuel the economic competitiveness of nations through job creation and increased revenues. However, there is an obvious lack of empirical evidence supporting the connection between multidisciplinary research and innovation. This study partly fills this gap by addressing the following question: Is the knowledge disclosed in a scientific publication more likely to be taken up in innovation (i.e. cited in the patent literature) as its multidisciplinarity index increases? The results thus obtained clearly show, in the aggregates, that multidisciplinarity increases the odds of research results being useful to innovation, thereby supporting existing R&I policy interventions or paving the way for new ones. However, because uptake in innovation, as measured through patent citations to scientific articles, remains a relatively rare phenomenon, multidisciplinary research is not an effective predictor of an individual article being useful to innovation; science is experimental, and the innovation outcome of an individual project cannot be guaranteed ahead of time merely through the extent of disciplinary mixing among the participating researchers.

Conference topic

Science policy and research assessment, Citation and co-citation analysis, Patent analysis, Knowledge discovery and data mining

Introduction

The past decade has seen a significant change in the organisation and management of scientific research. Although research has traditionally been organised into specialised disciplines, more and more governmental initiatives are emerging that aim to break the disciplinary silos (Van Rijnsoever & Hessels, 2011; Allmendinger, 2015). Various approaches are reported in the literature to operationalise this mode of research, the most common being interdisciplinary research (IDR) (Sonnenwald, 2007; Wagner et al., 2011).

The underlying rationale motivating IDR initiatives resides in an increasingly shared view among academics and policymakers: by uncovering innovative solutions residing outside the context in which they emerged, partnerships integrating multiple scientific cultures and bodies of knowledge help foster new lines of thought (i.e. the emergence of new disciplines) and help tackle and solve today's increasingly complex problems (Allmendinger, 2015; Blackwell et al., 2010; Mainzer, 2011). Blackwell et al. (2010) refer to the former application of IDR as being curiosity driven, and the latter application as being outcome driven.

By helping companies to stay ahead, or at least abreast, of the most recent developments in a rapidly changing business environment, outcome-driven IDR is perceived as having the potential to boost the competitiveness of firms and the economic well-being of nations. While the number of initiatives capitalising on outcome-driven IDR to fuel innovation in firms is increasing, there is a glaring lack of empirical evidence to support the notion that IDR is an effective mechanism to spur innovation and longer-term job creation and competitiveness in the knowledge economy. As laid out by Allmendinger (2015):

> *While there are plenty of data, insights and lessons on directed research programs and organized research units at universities, we have but next to no empirical evidence on how to best stage interdisciplinarity, about the added value it may produce, and what it may take universities and research organizations to effectively cross narrow disciplinary boundaries, perspectives, and interests. The ironic bottom line is that we need both more interdisciplinarity, and more organizational experiments, to advance it, and to learn more about what is conducive to it, what works and what does not.*

This study aims to partly fill this gap by investigating the type of added value that may be produced by IDR. More specifically, the following question is being addressed: To what extent is IDR positively associated with innovation, focusing on outcome-driven/applied research? Because current interventions rely mostly on assumptions derived from rational—but speculative—thinking, it is still pertinent to investigate the above question in furthering our understanding of IDR and its potential outputs/outcomes—this to further assess the relevance of existing and upcoming policy interventions. With this aim, this study's policy question was converted into the following data mining problem: the extent to which IDR is conducive to innovation is measured by investigating if, and to what extent, the knowledge disclosed in a scientific publication has greater odds of being taken up in the patent literature as its interdisciplinarity increases. Because the study focuses on a specific type of innovation—product/process innovations as disclosed in patents—the findings are not generalizable to the full innovation landscape.

Methods

Briefly, uptake in innovation is inferred by matching non-patent references (NPRs) in patents to a bibliographic database of peer-reviewed scientific literature focusing on science and engineering fields. This variable is referred to as the 'cited in patent' indicator throughout this paper. The scientific literature on measuring IDR has been blooming in recent years (Porter & Rafols, 2009; Rafols & Meyer, 2010; Campbell et al., 2015; Cassi, Mescheba, & de Turckheim, 2015; Calatrava Moreno, Auzinger, & Werthner, 2016). Most of these studies attempted to measure IDR using bibliometric data extracted from large bibliographic databases of peer-reviewed literature to quantify the diversity of integrated knowledge within individual research articles. This is typically done by computing the Rao-Stirling diversity index of the material cited in an article. This index integrates the number of different subfields cited, the balance between these subfields, and the intellectual distance between them (Porter & Rafols, 2009).
Because one of the main eligibility criteria for researchers applying to funding programmes targeting IDR is almost always the requirement that the project team includes researchers from diverse disciplines—and because IDR measured from co-referencing of multiple disciplines in an article (as in Porter & Rafols, 2009) can result from the work of a single author—it was decided that the Rao-Stirling diversity index would be applied to papers' contributing disciplines as revealed by the departmental affiliations of their authors, instead of their cited subfields. To differentiate this indicator from those based on cited subfields, as per a paper's references, it is referred to as the 'multidisciplinarity' (MDR) throughout this paper.

Data sources

Two sources of primary data are necessary to match NPRs in patents to individual scientific publications to create the cited in patent indicator: a patent database and a database of peer-reviewed scientific publications. For this study, PATSTAT was selected as the patent database, limited in this instance to a dataset of USPTO patents. The Web of Science (WoS), produced by Clarivate Analytics, was selected as the database for peer-reviewed scientific publications. The WoS also includes the addresses of all authors on a publication, enabling the computation of the MDR.

Only articles published in the domains of Natural Sciences and Engineering (NSE) and Health Sciences (HS) were retained for this study (as defined by Science-Metrix' classification, see Archambault, Caruso, & Beauchesne, 2011). This is because materials from the Social Sciences and Humanities are unlikely to be taken up in innovation as disclosed in patents. Also, although reviews might be cited in patents, they typically do not disclose original research contributions. Conference papers were not available to perform this analysis.

Cited in patent indicator

The cited in patent indicator is a binary indicator, computed at the paper level. It takes the value of 0 if the publication is not cited in the patent literature and a value of 1 otherwise. A detailed discussion of the matching procedure that was implemented in this study to link the patents' NPRs in the USPTO to scientific articles in the WoS is beyond the scope of the present paper. For a detailed presentation of this procedure, see Campbell et al. (2016).

Multidisciplinarity (MDR)

Recall that the MDR is measured with the Rao-Stirling diversity index of the disciplines represented among a paper's author addresses in the WoS. The department names appearing in the author addresses of all papers in the WoS were harmonized to 129 distinct forms representing the disciplines used in computing the MDR. The pairwise similarity matrix between the 129 disciplines is computed using the cosine similarity between the distribution vectors, across scientific subfields as per Science-Metrix' journal-based classification,[1] of any two disciplines. The distribution of a given discipline across subfields is obtained by counting the papers, in each subfield, that include the corresponding discipline among their author addresses. The index varies from 0 for monodisciplinary papers to 1 for highly multidisciplinary papers. Because it was not feasible to assign a cleaned department name to all author addresses in the WoS, potential biases might arise from this limitation in the data. Although the characterisation of this indicator revealed that robust inference can be performed even in cases where not all addresses have been classified (for more details, see Campbell et al., 2016), the analyses were performed using only those articles that had all their addresses classified by department. This was done to reduce the size of the dataset, and thus allow the computation of the analysis to run (see below).

Analyses

The question to be addressed in this paper can be re-formulated as follows: Is the knowledge disclosed in a scientific publication more likely to be taken up (i.e. cited) in the patent literature as its MDR increases? To address this question, it was decided that a logistic regression would be appropriate for the type of data we are dealing with: logistic regression supports multiple discrete and/or continuous predictors (MDR is a continuous predictor and additional controls are to be added to the model) and is suitable for binary outcomes (a paper being or not being cited in patents).

Three additional variables were created: the subfield of an article, the number of contributing authors, and the number of contributing countries on an article. These were inserted in the model to control for other networking effects that could supplant the effect associated with the

[1] http://www.science-metrix.com/en/classification

crossing of disciplinary boundaries in research teams. For example, an increased citation rate in patents could be due simply to the increase in the number of researchers or countries involved in a research project, rather than to the diversity of disciplines represented among them. Another factor that likely influences the citation in patent outcome is related to inter-sectoral cooperation. For example, is the publication resulting from a public–private partnership? At the time of submitting this paper, this variable had not yet been prepared and included in the study's model. Further, due to space limitations in the current paper, the additional materials related to this analysis will be added to the oral presentation.

In the context of this study, it is important to note that the event to be modelled—the citation of an article in patents—is rare. For instance, only 2% of 2008 articles in the NSE and HS, as indexed in the WoS, have been cited in USPTO patents (see Table 3 in *Results* section). With dozens of times fewer 1s than 0s (i.e. cited in patents vs. not cited in patents), most logistic regression tools would likely underestimate the probability of citation in patents even in the presence of sample sizes in the thousands (King & Zeng, 2001). An algorithm that provides an accurate estimate of the binary logistic regression's coefficients in such circumstances is Firth's biased reduced logistic regression using penalised likelihood (Firth, 1993). The R implementation of this algorithm was selected for this study (see 'logistf' package created by Heinze and colleagues, 2015).

Since the algorithm used to estimate the parameters of the constructed regression model is computationally intensive and could not run on the entire dataset, the dataset was reduced using three consecutive steps. First, the analysis was restricted to the most recent publication year of scientific articles (in the WoS) for which a sufficiently long time window was available to capture their citations in US patents (in the USPTO); this was to ensure that the analysis would reveal contemporary effects. Since it can take, on average, about 3.5 years for patents to be granted from their application date at the USPTO, and because relevant references to scientific articles can be added by patent examiners in this period, the time window over which citations can be captured should minimally extend over a five-year period—that is, four years to account for the granting process at the USPTO, and one year to account for the fact that papers can be published late in a given year. As complete data on patents were available up to 2014, this means that the citation windows of articles published after 2010 (citation window of five years; 2010 through 2014) would be too short. To balance data accuracy with the need to focus on contemporary effects, it was decided that a minimum window of seven years would be used, restraining the analysis to 2008 articles. Articles published in earlier years could have also been used, but this would have necessitated including an additional control variable to the model (i.e. publication year) to account for the fact that older papers have had the chance to accumulate citations over a longer period. Since the model was already taking a long time to run, the focus was placed on 2008 articles only. This led to a dataset of 402,916 articles times five variables (subfield, MDR, number of authors, number of countries and citation in patents) for a total of about 2 million data points. Following this restriction, the dataset was still too large to run the selected analysis method.

As a second step, limiting the analysis to only those articles with all their addresses classified by discipline (or department) reduced the size of the dataset by an additional 37% (from 402,916 to 255,372 articles). Filtering using such a criterion also carries the benefit of eliminating a potential source of biases in the measurement of multidisciplinarity at the individual paper level. Conversely, such a filter may bias the population towards papers with fewer authors, and hence towards papers that are less multidisciplinary; the more addresses there are on a paper, the more difficult it is to clean all addresses. To assess the robustness of the conclusions drawn in this study, the analyses presented in this paper could be run again using a random sample of papers that have at least 5 addresses with a cleaned department or that have a cleaned department assigned to at least 50% of their addresses when they have fewer than 10 addresses. That said, since the analysis was possibly focused on a portion of the population of papers that is less multidisciplinary, it is anticipated that the results presented here are conservative.

With the above filter applied to the dataset, the computer (4 cores/8 threads Intel Xeon CPU E3-1240 v2 processor (3.40GHz) running a 64-bit version of Windows Server 2008 R2 with 32GB of RAM) crashed after a day of computing. Accordingly, as a third step, this restricted dataset was randomly downsampled to 100,000 articles so the analysis could run successfully.

Global analysis: In performing the analysis for all 132 NSE and HS subfields combined, the subfield of scientific articles was added to the model as a control (dummy) variable. This was important to control for differences that prevail in the citation practices across technological fields (e.g. some fields generally include more references to the scientific literature than other fields do). As the global model was statistically significant (p-value < 0.05), the model was then run again for each subfield separately.

Analysis by subfield: Out of the 132 NSE and HS subfields, 44 (one third) that had at least 30 papers cited in patents were retained for analysis. The subfields were filtered since a small number of cases on the rarer of the two outcomes in binary logistic regression can lead to underestimation of the odds ratios. These 44 subfields accounted for 65% of all papers in the combined subfields. All articles with 100% of their addresses classified by discipline (or department) were kept in the analysis (i.e. 164,972 articles out of 255,372; the data were not downsampled). Only the subfields with a statistically significant relationship between the multidisciplinarity of peer-reviewed scientific articles and the event of being cited in patents are reported in the results section.

Results

Global analysis

When all 132 NSE and HS subfields were combined, the global model was statistically significant (p-value < 0.05); its associated parameters are shown in full in Table 1. Note that the lower and upper odds ratios provide the 95% confidence interval of the odds ratio for a given predictor. When the confidence interval does not overlap with 1, it means that the odds ratio is significant (p-value of the odds ratio < 0.05).

The odds ratio of a predictor typically shows the multiplicative factor by which the odds of citation (of an article, by a patent) increase with each full unit change in the predictor. For multidisciplinarity only, the odds ratio was re-scaled for a magnitude of change of 0.1 unit since a full unit change for MDR is highly unlikely; in contrast to the number of authors and countries, the MDR score of a paper is bounded between 0 and 1. Also, a variation of 0.1 is commonly observed across all papers; 42% of articles have a deviation from the median MDR score that is greater than or equal to 0.1. For the number of authors and countries, the odds ratio is reported for a 1-unit change, which is more commonly observed for the former relative to a 0.1-unit change for MDR (78% of papers deviate by one or more units from the median), but less commonly observed for the latter (27% of papers). How to interpret the MDR score, to digest what a 0.1-unit increase means in a practical context, is explored at length in Campbell et al. (2016).

As detailed in Table 1, the MDR of an article has a statistically significant effect on the likelihood that its results will be taken up in innovation; for instance, when the MDR score of an article increases by 0.1 unit, its odds of being cited in the patent literature increases by 12% (odds ratio of 1.12).

Table 1. Relationship between the multidisciplinarity, as well as the number of authors and countries, of scientific articles (2008) and their citation in the patent literature (patents issued in 2008–2014).

Model variables (Subfield as a dummy)	Odds Ratio	Lower Odds Ratio	Upper Odds Ratio
Model p-value = 0.000			
Multidisciplinarity ǂ	1.12	1.09	1.16
Number of authors	1.09	1.07	1.11
Number of countries	0.92	0.84	1.00

Note: ǂ The odds ratio was re-scaled for a magnitude of change of 0.1 unit instead of 1 unit (recall that the MDR can only take values from 0 to 1).

This analysis of odds ratios was also conducted on the number of authors listed for a paper and the number of countries participating in the collaboration, to determine whether either of these features exert an effect that could suppress the influence of multidisciplinarity on the uptake of scientific knowledge in innovation. The meaning of such a change is simple to interpret for these predictors: Does having one more author, or one more participating country, involved in publishing a paper increase its likelihood of being cited in the patent literature?

The odds ratio for having an additional author is statistically significant and equals 1.09, meaning that having one more author on a paper increases the odds of citation in the patent literature by 9% (Table 1). Thus, a larger number of collaborating authors does have a positive effect on the uptake of scientific knowledge in innovation and this effect does not override the effect linked to the multidisciplinarity of a research team. International collaboration does not appear to exert a statistically significant effect on the uptake of scientific knowledge in patents.

Analysis by subfield

Now that a positive and statistically significant effect of MDR and the number of collaborating authors on the odds of being cited in the patent literature has been detected at the global level, it becomes relevant to assess if this effect holds equally well across subfields or if there are any meaningful disparities. The results are only presented for the 44 subfields with at least 30 papers cited in patents; as previously noted, subfields were filtered in this way since a small number of cases on the rarer of the two outcomes in binary logistic regression can lead to an underestimation of the odds ratios.

A summary of results at the subfield level is presented in Table 2. Of the 44 retained subfields, the model was statistically significant in 29 cases (66%). Of those 29 subfields, the odds ratio for MDR scores is statistically significant in 16 subfields (55%) and the effect is always positive in these cases. For the number of authors on a paper, the odds ratio is statistically significant in 22 of the 29 subfields (76%) for which the model is significant and the effect is positive in 95% of these cases. As noted above, international collaboration does not have a statistically significant connection with innovation, a finding that is borne out here at the subfield level: the odds ratio for number of countries is statistically significant in 9 of the 29 subfields (31%) for which the model is significant and the effect is positive in only 44% of these cases.

Focusing on only those subfields where the connections to innovation (positive or negative) are statistically significant for each predictor taken separately, we can once again look at the effects of MDR, number of authors and number of countries. On average across statistically significant subfields for MDR, an increase of 0.1 in MDR score is associated with a 26% increase in the odds of being cited in the patent literature (avg. odds ratio of 1.26, Table 2). A paper with one more author (a 1-unit change for number of authors) is associated, on average, with a 13% increase in odds of being cited by a patent. The inclusion of an additional country in international collaboration is associated, on average, with a 14% increase in the odds of being cited in the patent literature. Although this is similar to the score for the number of authors, the range of the statistically significant odds ratios for the number of countries is broader, with stronger scores on the negative side effect (i.e. odds ratio below 1). The highest (significant) odds ratio for the number of countries is 1.72, but the lowest is 0.62. These values are more extreme than those for MDR (1.13 to 1.54) and the number of authors (0.92 to 1.13), meaning that there is much stronger variation from one subfield to the next in terms of the connection between international collaboration and citation in the patent literature. This finding is unsurprising given that findings presented above suggested that the connection was much weaker between international collaboration and patent citation (in the aggregate and across subfields). Only for multidisciplinarity are the effects consistently on the positive side across subfields where the predictor has a statistically significant effect—that is, multidisciplinarity never appears to lead to important decreases in the odds of citation in patents.

Table 2. Comparative analysis of the magnitude of effect of the multidisciplinarity, number of authors and number of countries of scientific articles (2008) on their odds of being cited in the patent literature (patents issued in 2008–2014).

	Multi-disciplinarity‡	Number of authors	Number of countries
	The results concern 66% (29) of the 44 retained subfields with a statistically significant model		
Percentage of subfields with a significant odds ratio	55%	76%	31%
Percentage of subfields with an odds ratio greater than 1 (i.e. positive effects) among those with a significant odds ratio	100%	95%	44%
Average odds ratio across subfields where the odds ratios are significant	1.26	1.13	1.14
Minimum odds ratio across subfields where the odds ratios are significant	1.13	0.92	0.62
Maximum odds ratio across subfields where the odds ratios are significant	1.54	1.45	1.78

Note: ‡ The odds ratios were re-scaled for a magnitude of change of 0.1 unit.

In summary, multidisciplinarity is most consistently connected to patent citation, compared to the other two predictors. This finding holds both in the aggregate and at the subfield level. However, the number of authors has a significant effect in a larger set of subfields with an effect which is nearly always positive. Thus, both multidisciplinarity and the number of authors are significant factors to the uptake of articles in patents.

For the 16 subfields in which the multidisciplinarity–innovation link is statistically significant, the odds ratios range from a 13% to a 54% increase in the odds of being cited in patents for a 0.1-unit increase in MDR score. In-depth results for each of these subfields are presented in Table 3. More specifically, this table provides information on the number of cited and uncited articles in each subfield (along with the percentage of cited articles), their respective odds ratios for a 0.1-unit increase in MDR score, the upper and lower bounds of those odds ratios, and the share of articles deviating by at least 0.1 unit from the median MDR score in the given subfield. The coefficients of other variables included in the model are not shown here.

Table 3. Relationship between the multidisciplinarity of scientific articles (2008) and their citation in the patent literature (patents issued in 2008–2014).

Subfield	N			Binary logistic regression				
	Uncited	Cited	Percent Cited	Model (p-value)	$OR_{X \rightarrow X+0.1}$	Lower $OR_{X \rightarrow X+0.1}$	Upper $OR_{X \rightarrow X+0.1}$	Share of articles with a dev. from the median multidisc. ≥ 0.1
Global model (subfield as dummy; the odds ratio [OR] is only shown for multidisciplinarity)								
All 132 NSE & HS subfields	98 013	1 987	2.0%	0.000	1.12	1.09	1.16	42%
Model by subfield (the odds ratio [OR] is only shown for multidisciplinarity)								
Orthopaedics	1 859	34	1.8%	0.001	1.54	1.21	1.99	46%
Pathology	1 803	45	2.4%	0.001	1.44	1.13	1.87	37%
Nuclear Medicine & Medical Imaging	4 231	106	2.4%	0.000	1.40	1.21	1.62	33%
Gastroenterology & Hepatology	3 507	50	1.4%	0.001	1.36	1.08	1.74	33%
General Chemistry	3 354	119	3.4%	0.000	1.36	1.21	1.54	50%
Ophthalmology & Optometry	1 825	41	2.2%	0.023	1.28	1.03	1.61	42%
Polymers	3 258	108	3.2%	0.000	1.28	1.11	1.48	38%
Organic Chemistry	4 580	167	3.5%	0.000	1.22	1.12	1.34	57%
Biomedical Engineering	1 980	106	5.1%	0.005	1.19	1.04	1.38	52%
Neurology & Neurosurgery	11 441	206	1.8%	0.000	1.19	1.07	1.32	38%
Applied Physics	8 466	265	3.0%	0.000	1.17	1.08	1.28	38%
Analytical Chemistry	3 733	109	2.8%	0.030	1.17	1.04	1.32	41%
Oncology & Carcinogenesis	9 026	350	3.7%	0.000	1.16	1.06	1.26	35%
Nanoscience & Nanotechnology	2 467	243	9.0%	0.000	1.15	1.05	1.26	38%
Pharmacology & Pharmacy	4 041	118	2.8%	0.005	1.14	1.02	1.29	43%
Cardiovascular System & Hematology	5 825	118	2.0%	0.001	1.13	1.00	1.30	38%

Note: $OR_{X \rightarrow X+0.1}$ = odds ratio re-scaled for a change of 0.1 unit in the MDR score of papers.

Out of the 16 subfields where a positive and statistically significant connection exists between multidisciplinarity and the uptake of scientific knowledge in innovation, 10 (63%) are related to the Health Sciences (Table 3), where the extent of multidisciplinarity research is more pronounced (data not shown). In Table 3, the odds ratio varies from a low of 1.13 in Cardiovascular System & Hematology to a high of 1.54 in Orthopaedics. In Orthopaedics, this means that a 0.1-unit change in MDR translates into a 54% increase in the odds of an article being cited in patents. Of all the articles in this subfield, 46% deviate from the median MDR score by 0.1 unit or more. This indicates that some papers stand out by a sufficient margin in terms of MDR, significantly increasing the odds of their results being taken up in patents. The subfields with the most disparate spreads of MDR scores are Organic Chemistry, Biomedical Engineering, and General Chemistry, each of which has at least 50% of its articles with an MDR score 0.1 unit or more away from the median. In turn, it is reasonable to assume that a policy aiming to promote multidisciplinarity as a way to catalyse innovation could lead to a sufficient change in the MDR of supported articles in these subfields to achieve a noticeable increase in the odds of their findings being taken up in innovation.

Until now, the odds ratios have only been analysed for 0.1-unit changes in MDR. Although the odds ratio when moving from an MDR of 0.1 to 0.2 is the same as when moving from an MDR of 0.5 to 0.6, it is worth noting that the relationship between an odds ratio and the magnitude of change (e.g. 0.1 unit, 0.2 unit, …, 1 unit) in the predictor (i.e. MDR) is exponential, making it relevant to investigate the magnitude of an effect that can be achieved when doubling, tripling or even quadrupling the MDR of a paper. In doing so, the analysis focused specifically on the subfield of Orthopaedics, which has the largest odds ratio—1.54 for a 0.1-unit change in MDR. Figure 1 shows the change in odds ratio as a function of the magnitude of change in the MDR score of a 2008 article in Orthopaedics. The odds of being cited in patent for articles with an MDR score of 0.4 is slightly more than five times as large as it is for monodisciplinary papers (MDR = 0) in Orthopaedics (odds ratio = 5.57); the magnitude of this effect is 3.6 times larger than for an MDR of 0.1 (odds ratio = 1.54). Taking any baseline, an increase of 0.4 of a unit in the MDR score of an article (e.g. from 0 to 0.4, 0.1 to 0.5, 0.2 to 0.6) will lead to a 5.57 times increase in the odds that the knowledge it contains will be taken up in innovation. This is a non-negligible effect when we note that close to a fifth of 2008 articles in Orthopaedics have an

MDR score of at least 0.4, while the mode of the distribution is at MDR = 0 (see inner chart in Figure 1). For an MDR score of 0.67, the highest score observed in Orthopaedics in 2008, the odds of being cited in patents are nearly 18 times larger than for monodisciplinary papers.

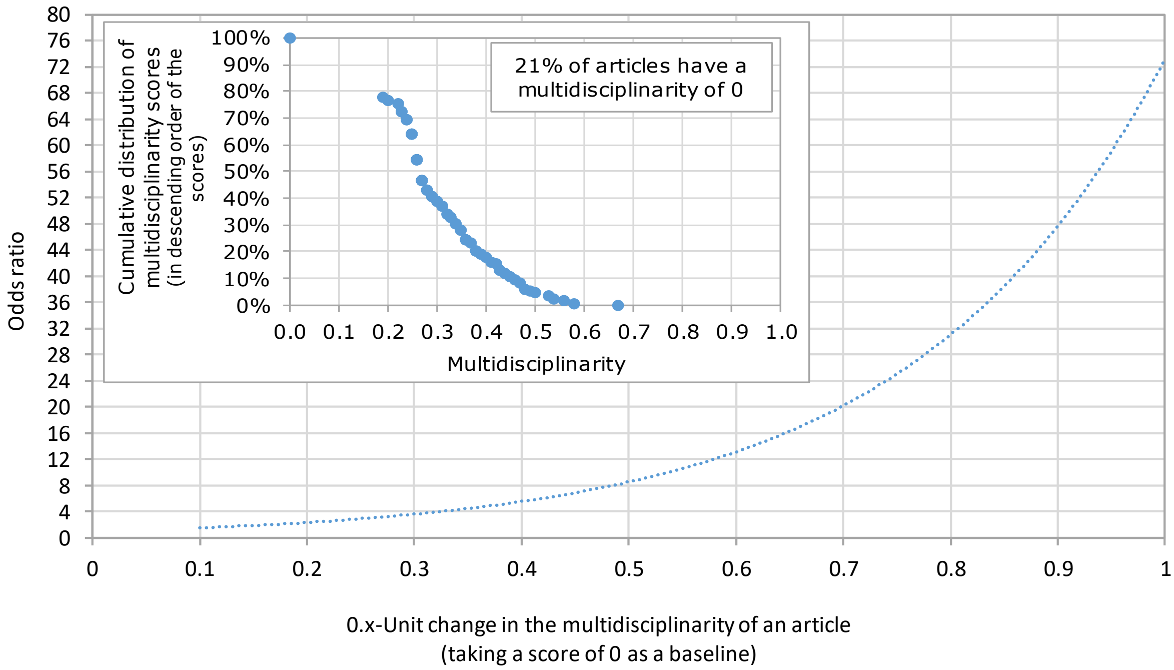

Figure 1. Relationship between the increased odds (odds ratio) of an article being cited in the patent literature and a given change in its multidisciplinarity score (for 2008 articles in Orthopaedics, cited in patents issued in 2008–2014).

The preceding analyses of MDR's effect on the odds of uptake in the patent literature, for the subfield of Orthopaedics, has shown that the relationship is exponential, with the odds of being cited in a patent growing more and more rapidly as the magnitude of change in MDR scores increases. But an important piece of context is still absent: the relative increases in odds of uptake in patents has been discussed, but the odds themselves, as a function of MDR scores, have so far not been quantified. Let us turn to that point now. Because citation by a patent is a relatively rare event, individual articles needed to be sorted into suitably large bins to enable the signal to emerge clearly from the background noise. For the present analysis, Orthopaedics articles were sorted into three bins in ascending order of their MDR scores: one for the ~22% of articles with an MDR score of 0, and then two equally sized bins of ~39% each, each accounting for exactly half of the remaining articles. More bins could not be produced without reducing the number of cited articles found in each bin, which would introduce unwanted noise into the analysis.

The average multidisciplinarity of articles, as well as their odds of being cited in patents, was computed for each bin, producing the data points shown in Figure 2. In this figure, the results are showing the actual odds of being cited, not the odds ratio (which quantifies the relative change in odds for a given change in MDR, but not the odds themselves). As detailed in Figure 2, the observed odds of citation in patents for the three bins prove to be a relatively strong fit to the expected odds of being cited in patents as predicted by the logistic regression model (see the exponential prediction line in the graph), suggesting the overall adequacy of the model. However, because of the rarity of article references in patents, the predictive power of this model for any individual article will be very low; instead, the model is better suited to predicting overall patterns in the aggregate.

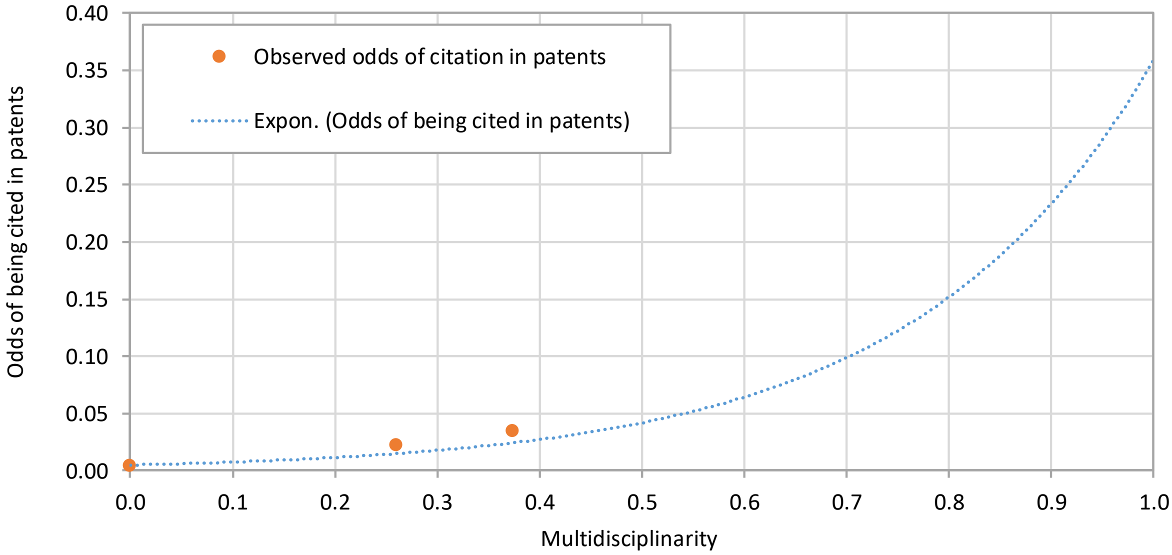

Figure 2. Relationship between the increased odds (odds ratio) of an article being cited in the patent literature and a given change in its multidisciplinarity score (for 2008 articles in Orthopaedics, cited in patents issued in 2008–2014).

Note: The model was fitted using three explanatory variables: multidisciplinarity, number of authors and number of countries. In this graph, the number of authors and countries were held constant using the average number of authors (4.94) and countries (1.17) across all Orthopaedics papers. This is acceptable given the small variability observed in these quantities across the three bins (no. of authors ranges from 4.47 and 5.19; no. of countries ranges from 1.17 to 1.20) and given the non-significance of the odds ratio for these two variables. The effect of these variables is here embedded in the model intercept.

Comparing Figure 1 to Figure 2 can provide a valuable perspective on the effects of MDR on uptake in innovation. For a paper with an MDR score of 0, the odds of being cited by a patent is 0.005—roughly a 1:200 chance. Increasing the MDR score to 0.4, which is a significant increase in multidisciplinarity, increases the odds more than fivefold. However, even a fivefold increase of such a small value results in a still relatively small value: a paper with an MDR score of 0.4 has about an odd of being cited by a patent of 0.027, still representing only a little better than 1:40 chance.

Accordingly, while it is appropriate to conclude that multidisciplinarity increases the odds of research being taken up in innovation, its effect is not large enough to predict with certainty that a given piece of work will contribute to new innovations (i.e. be cited in patents) solely on the basis of the disciplinary diversity of the research team. Far more factors are likely to influence such an outcome, and these would have to be introduced into the model to precisely predict the innovation outcome of an individual research article (or project). Still, the relationship that has been uncovered in this study suggests that, in some scientific (sub-)fields, promoting multidisciplinary research can significantly increase the odds of fuelling new innovations, thereby paving the way for evidence-based R&I policy intervention. Additionally, it should be noted that a rather short citation window has been used, which likely underestimates the longer-term odds of a patent citation. Consequently, policy interventions promoting multidisciplinary research may have more leverage on innovation than is suggested here.

Discussion

The present study has assessed the connection between research that crosses disciplinary boundaries and contributions to innovation through the measurement of multidisciplinarity (assessing the number, balance and diversity of disciplines or departments integrated into a research team) and the resulting odds of articles published by these research teams being taken up in the patent literature, as tracked through citations. The conclusions reached are that multidisciplinarity certainly does contribute to increased odds of uptake in innovation; this finding holds across the aggregated subfields of the NSE and HS, as well as in about 1 in 3 subfields individually. Multidisciplinarity seems to have a more robust connection to innovation than the number of countries (for international collaboration). The number of authors also appears to contribute positively to citation in patents.

Even so, this study has shown that MDR is only valuable as a predictor of general tendencies within a set of cases and cannot effectively predict whether an individual article will or will not be cited by a patent. If one assumes that uptake in innovation is not the result of a chancy system, then the factors that most strongly determine uptake for individual articles have yet to be identified. Further study would be needed to fill out this picture and hopefully identify those strongly determining elements. One such idea would be to examine the sectors in which the collaborating authors are working, seeking especially to see whether collaborations involving private-sector co-authors have a higher odd of providing the foundation for subsequent patents. Furthermore, it must be acknowledged that even once these determining factors are identified, the large majority of peer-reviewed publications are never cited by a patent. Uptake in innovation thus remains a relatively rare phenomenon.

Integrating these research findings back into the policy context, then, what lessons may be extracted? First, this study was initially driven by the need for some evidence to support a fundamental assumption underlying a growing number of R&I policy interventions: that crossing disciplinary boundaries helps to fertilise the ground for innovation, tilling the soil for job creation, economic growth and increased competitiveness. The evidence discovered in this study suggests that this baseline assumption is indeed borne out by facts.

It is important to note that the present study has only been a pilot test. It was conducted using a single year of peer-reviewed publications (2008) and a restricted range of years for patents (2008–2014). The cleaning of author addresses by discipline (or departments) could be improved to remove additional noise from the analysis, and the algorithm matching NPRs to scientific articles in the WoS should be improved to increase its recall while maintaining its precision. Currently, it is likely that the citation rate of scientific articles in patents is underestimated, diluting the effects measured in this project. For example, the 7-year citation window used is rather short in the context of patents, possibly underestimating the longer-term odds of a patent citation. Policy interventions promoting multidisciplinary research may thus have more leverage on innovation than is suggested in this paper. Furthermore, only 44 of 132 subfields could be analysed, as the others did not meet a minimum threshold of cited papers to facilitate robust analysis. Thus, while this study's findings are valuable and shed light on a fundamental policy assumption in need of evidential support, they are also extremely preliminary, requiring much more robust assessment before they should be considered sufficient for the basis of future policymaking—robust both in the sense of repeating these analyses on larger samples and in the sense of approaching these phenomena from other angles.

Acknowledgments

This study was performed as part of a larger project on the development and application of a methodological framework guiding the design and implementation of data mining projects in the R&I policy context. The project was commissioned by the European Commission, DG Research and Innovation, Unit A4 – Analysis and monitoring of national research policies (Framework Contract: No 2010/S 172-262618). Any use, right of dissemination or copyright belongs to the European Commission. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Commission.

References

Archambault, É., Caruso, J., & Beauchesne, O. (2011). Towards a multilingual, comprehensive and open scientific journal ontology, in B. Noyons, P. Ngulube, and J. Leta (Eds.), *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics (ISSI)* (pp. 66–77*)*. Durban, South Africa.

Allmendinger, J. (2015). *Quests for interdisciplinarity: A challenge for the ERA and HORIZON 2020*. Policy Brief by the Research, Innovation, and Science Policy Experts (RISE), DG Research and Innovation. doi 10.2777/499518.

Blackwell, A. F., Wilson, L., Street, A., Boulton, C., & Knell, J. (2010). *Creating values across boundaries*. NESTA Research Report, http://www.nesta.org.uk/sites/default/files/creating_value_across_boundaries.pdf.

Campbell, D., Deschamps, P., Côté, G., Roberge, G., Lefebvre, C., & Archambault, É. (2015). Application of an “interdisciplinarity” metric at the paper level and its use in a comparative analysis of the most publishing ERA and non-ERA universities. *Presentation at the 20th International Conference on Science and Technology Indicators (STI)*, Lugano, Switzerland.

Campbell, D., Tippett, C., Struck, B., Lefebvre, C., Côté, G., St-Louis Lalonde, B., Ventimiglia, A., Roberge, G., & Archambault, É. (2016). *Data mining on key innovation policy issues for the private sector: application report.* Report prepared for the European Commission, 228 pages (see Section 3.1). Because the report has not yet been published, it must be accessed through a request made to DG Research and Innovation, Unit A4.

Calatrava Moreno, M. D. C., Auzinger, T., & Werthner, H. (2016). On the uncertainty of interdisciplinarity measurements due to incomplete bibliographic data. *Scientometrics*, *107*(1), 213–232.

Cassi, L., Mescheba, W., & de Turckheim, E. (2015). My interdisciplinarity à moi: An analysis of neuroscience research in French universities, 2008–2012. *Presentation at the 20th International Conference on Science and Technology Indicators (STI)*, Lugano, Switzerland.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38.

Heinze, G., Ploner, M., Dunkler, D., & Southworth, H. (2015). Package ‘logistf’. http://cemsiis.meduniwien.ac.at/en/kb/science-research/software/statistical-software/fllogistf/.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, *9*, 137–163.

Mainzer, K. (2011). Interdisciplinarity and innovation dynamics: on convergence of research, technology, economy, and society. *Poesis Prax*, *7*(4), 275–289.

Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, *81*(3), 719–745.

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, *82*(2), 263–287.

Sonnenwald, D. H. (2007). Scientific collaboration: A synthesis of challenges and strategies. *Annual Review of Information Science Technology*, *41*(1), 643–681.

Van Rijnsoever, F. J., & Hessels, L. K. (2011). Factors associated with disciplinary and interdisciplinary research collaboration. *Research Policy*, *40*(3), 463–472.

Wagner, C., Roessner, J., Bobb, K., Thompson Klein, J., Boyack, K. W., Keyton, J., Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, *5*(1), 14–26.