Research Policy 38 (2009) 306–317

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Research collaboration and research output: A longitudinal study of 65 biomedical scientists in a New Zealand university

Zi-Lin He[a,*], Xue-Song Geng[b], Colin Campbell-Hunt[c]

[a] Centre for Innovation Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
[b] Rotman School of Management, University of Toronto, 105 St. George Street, M5S 3E6 Toronto, ON, Canada
[c] Department of Management, University of Otago, P.O. Box 56, Dunedin, New Zealand

A R T I C L E I N F O

Article history:
Received 4 April 2005
Received in revised form
14 November 2008
Accepted 18 November 2008
Available online 3 January 2009

Keywords:
Research collaboration
Research output
Longitudinal study

A B S T R A C T

Collaborative research has been increasingly celebrated by the science community, but the hypothesized positive relationship between research collaboration and research output is more assumed than rigorously tested. In this paper, we identify three methodological gaps in the literature: (a) hierarchical coding based on the ISI Web of Science database causes severe loss of information on local collaboration, (b) the relationship between research collaboration and research output is likely to be confounded by a common latent variable such as a scientist's ability, and (c) the lack of longitudinal analysis prevents causal inferences from being made. To address these methodological concerns, we constructed a longitudinal dataset of 65 biomedical scientists at a New Zealand university and coded collaboration variables by hand checking each of their publications in a period of 14 years. We found that at article level, both within-university collaboration and international collaboration are positively related to an article's quality and that, at scientist-year level, only international collaboration is positively related to a scientist's future research output.

1. Introduction

Research collaboration has been a phenomenon of growing importance for scientists, research organizations, and policy makers (Narin et al., 1991; Wagner and Leydesdorff, 2005; Cummings and Kiesler, 2007; Bammer, 2008). It is widely assumed among scientists and politicians that collaboration in research is "a good thing" and that it should be encouraged. Policy makers sometimes take for granted that collaboration will increase the quantity and quality of research, and numerous initiatives have been launched to promote collaboration among individual researchers. This optimistic view in some cases even leads to "a positive valuation of collaboration for its own sake" (Duque et al., 2005, p. 756). For example, Jager (2006) has shown in a case study that some German universities already regard co-publication as a separate performance dimension (cited from Schmoch and Schubert, 2008).

There are many obvious reasons why there should be a positive relationship between research collaboration and research output, and it is intuitive to believe that the benefits of collaboration should be greater than the costs of coordinating collaborative research projects. A large number of studies have demonstrated a positive correlation between co-authorship, especially international co-authorship, and an article's quality as measured by number of citations it receives from other articles (Narin et al., 1991; Katz and Hicks, 1997; Glanzel and Schubert, 2001). As shown in a recent large sample study by Wuchty et al. (2007), co-authored articles receive more citations than sole-authored ones, and such "teamwork advantage" has been increasing over time.

Despite these associations, the relationship between research collaboration and research output is more assumed than investigated because of three methodological gaps in existing studies. The aim of this paper is to address these methodological concerns. First, most studies rely on affiliation data in the ISI Web of Science database to code research collaboration, but this database does not provide information on individual author-institution affiliations. When an article has affiliation addresses from more than one country, it is regarded as an internationally co-authored article. Then, when an article does not have affiliation addresses from more than one country, but has more than one affiliation address from the same country, it is regarded as a case of domestic collaboration. Finally, if an article has only one affiliation address but more than one author, it is regarded as a case of research collaboration within an organization. Such hierarchical coding is problematic because an internationally co-authored article may well have more local collaborators than international ones (e.g., a Dutch scientist wrote an article with four other scientists from the same Dutch university and another scientist from a British university). It is obvious that the true contribution of local collaborators would be understated

* Corresponding author. Tel.: +31 13 4663260; fax: +31 13 4668354.
E-mail addresses: Z.L.He@uvt.nl (Z.-L. He), xuesong.geng@rotman.utoronto.ca (X.-S. Geng), ccampbell-hunt@business.otago.ac.nz (C. Campbell-Hunt).

0048-7333/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.respol.2008.11.011

or patently ignored in such hierarchical coding. This neglect of local collaboration on an internationally co-authored article calls into question the widely held conclusion in existing studies that international collaboration is related to superior research quality.

Second, the relationship between research collaboration and research output is likely to be confounded by their common dependence on a third latent variable such as a scientist's ability. Collaboration involves a bilateral selection process whereby an able scientist is more attractive for research partnership than her less able peers. By "able" here we mean everything that can contribute to the success of a research project, from the ability to get funding support, industry connections to collect data, to rich experience with the editorial process of a prestigious journal. Therefore, the relationship between research collaboration and research output can be overstated to the extent that a scientist's "ability" drives both research collaboration and research output. An acceptable research design should try to control for the focal scientist's ability in search for a positive relationship between research collaboration and research output.

The third and probably the most important gap in the literature is that few studies have explicitly investigated through a longitudinal analysis whether individual scientists can gain from collaboration. Most existing studies employ a cross-sectional design to correlate variations in collaboration and variations in research output *between* scientists. A cross-sectional study cannot answer the question whether research collaboration leads to more research output or the other way around. A rigorous test would require longitudinal data that trace a sample of scientists and record their collaboration activities and research output each year for a number of years. We are aware of only a few longitudinal studies in this area, such as McFadyen and Cannella (2004) and Singh (2007), but all of them directly rely on the ISI Web of Science database and thus have the problem of understating local collaboration.

In this paper, we distinguish between international collaboration, domestic collaboration, and within-university collaboration, and we examine how they are related to the quality of individual articles and the annual output of individual scientists. We collected data for a panel of 65 biomedical scientists over a 14-year period. To avoid understating within-university collaboration, we painstakingly coded collaboration variables based on co-authorship information by visual inspection of every publication of these 65 scientists.

In article level analyses, we controlled for scientists' latent ability by including a full set of scientist dummies and accounted for non-independence in observations (because each scientist contributes more than one observation to the sample) by employing the generalized estimating equations (GEE) method. We found that both within-university collaboration and international collaboration are positively related to an article's quality as proxied by the impact factor of the journal in which it was published or number of citations it received in a 2-year window after publication.

In scientist-year level analyses, we used fixed-effects models to account for unobserved heterogeneity among these scientists and examine how research collaboration and research output covary over time within scientists. We found that the relationship between research collaboration and research output depends on the choice of dependent variable. When research output is measured by quantity, within-university collaboration is significantly related to a scientist's future research output. However, when research output is adjusted by both quality and authorship, international collaboration is significantly related to a scientist's future research output.

In Section 2, we review the literature and make a number of hypotheses. In Section 3, we describe the data sources and coding procedures in detail, and define the variables. We then provide the results of article level analyses and scientist-year level analyses. The last section offers a summary of the results and our conclusion.

2. Literature review and hypotheses

A common assumption in science policy is that collaborative research has benefits of various kinds and thus should lead to increased research output. Among numerous benefits of research collaboration, often cited in the literature are sharing knowledge and techniques, cross-pollination of ideas, pooling research resources and sharing expensive instruments, increasing visibility and recognition, and accelerating research progress.

2.1. Why a positive relationship between research collaboration and research output?

There are at least three theoretical reasons for a positive relationship between research collaboration and research output. First, the perspective of knowledge recombination predicts that knowledge creation is often enhanced by combining different expertise and know-how from a wide variety of sources. As early as in 1934, Schumpeter stated that new knowledge is mostly created by recombining existing knowledge: "Development in our sense is then defined by the carrying out of new combinations" (Schumpeter, 1934, p. 66). According to this perspective, when partnering scientists bring together complementary knowledge and skills into a research project, the resulting research output should be of higher quality than it would be otherwise. Moreover, by working together, research collaborators form an internal quality control (or internal refereeing) process to improve research quality by rigorously selecting out unpromising combinations (Rigby and Edler, 2005).

Second, collaboration provides a learning experience for a scientist to acquire skills and techniques from partners for her future research activities. Scientific knowledge is commonly regarded as a public good to which every scientist, regardless of her location or field, has free and open access. However, some important knowledge in the actual conduct of scientific research is inextricably tacit and stored in the minds of individual scientists, not in published materials. The acquisition of tacit knowledge between scientists is best achieved when they jointly experience problem-solving and spend time together discussing and reflecting. The importance of tacit knowledge in conducting scientific research and how it can be transferred between scientists were demonstrated in an earlier study by Collins (1974) who reported that, in the early construction of the TEA laser, no laboratory succeeded in building a working TEA laser equipment without the participation of someone from another laboratory that had already put a device of such type into operation (cited from Cowan et al., 2000, p. 215). Collaboration allows scientists to keep abreast of the latest development in the field and take prompt advantage of the latest scientific advances. Reading published works by other scientists is not enough because these latest developments often embody a large amount of tacit knowledge that has not appeared in written forms.

Third, collaboration provides scientists with social networks where they can capture valuable information on research opportunities and expose themselves to future research collaboration which leads to future research output. Production of scientific knowledge is deeply embedded in social structures and practices among scientists (Crane, 1972; Katz and Martin, 1997). Through collaboration, scientists build, expand, and maintain their social capital that helps uncover novel research questions and facilitate future research collaboration. In a recent study, McFadyen and Cannella (2004) used number of co-authors and frequency of co-publishing with the same co-author to approximate two dimensions of social capital: number of relations and strength of relations. They found that a scientist's future research output increases with social capital.

2.2. Why this relationship is not always positive?

One should not assume that research collaboration will lead inexorably to higher or better research output. Collaboration also entails various costs, including costs of finding and assessing research partners, costs of establishing an agreement to organize collaboration and allocate the credit of expected research output, and costs of coordination among collaborating scientists, among others. Time must be spent clarifying roles and responsibilities from the outset and continuously updating them as a collaborative research project evolves (Katz and Martin, 1997).

These costs of research collaboration are rarely examined in the literature. One notable exception is Landry and Amara (1998) who applied the concept of transaction costs to investigate why a certain governance structure is adopted in organizing collaborative research. Landry and Amara (1998, p. 904) regard collaboration research as mini-joint ventures where collaborating scientists exchange resources and skills to generate and share expected research output. Because of bounded rationality, no one could exhaust all the contingencies of a collaborative research venture, no one is absolutely sure what research findings will be produced in the future, and no one is fully aware of the costs of implementing a specific part of the collaborative project. This impossibility of designing complete cooperative contracts creates room for opportunistic behaviours such that a scientist may strategically misrepresent information to secure more resources or credit for her contribution to the final research output. According to Landry and Amara (1998), this is why collaborating scientists have to spend time coordinating, monitoring, enforcing, and sometimes renegotiating contractual promises of inputs from other partnering scientists, while such a contract is often not explicitly put down in a written document.

2.3. Hypotheses

We have reviewed three theoretical perspectives that predict a positive relationship between research collaboration and research output, as well as reasons why research collaboration will not lead inexorably to higher or better research output. While no one can disentangle all the abovementioned causal mechanisms and cost factors, we propose that by comparing how collaboration at different geographical scales (international, domestic, and within-university) is related to research output at article level and scientist-year level, we can find evidence in support of one or more perspectives.

First, it seems that the knowledge recombination perspective operates mainly at article level because more co-authors on a given article indicate more sophisticated and greater scales of recombination from different scientists. Whereas the other two perspectives, the learning and networking perspectives, indicate research collaboration should be positively related to a scientist's future research output. Therefore, if we find research collaboration is positively related to research output at article level (as found in many existing bibliometric studies), but not at scientist-year level, we would conclude that learning and networking are not driving the relationship between research collaboration and research output.

Second, at article level, collaboration with partners at different distance ranges may have different impacts on research output in terms of an article's quality. According to the knowledge recombination perspective, international collaboration can be hypothesized to be more positively related to research output than domestic or within-university collaboration because distant partners are more likely to bring different experience and diverse ideas to expand the scope of combinatorial search. And there has been a general consensus in the literature of social psychology on group creativity that diversity rather than conformity is more likely to generate novel and high quality outcomes (De Dreu and West, 2001).

However, international collaboration incurs additional costs relative to local collaboration, especially in terms of coordinating research carried out at geographically dispersed locations and possibly extra time for travelling and visiting. One would also expect transaction costs in international collaboration to be larger than those in local collaboration. In contrast, when a scientist collaborates with her local colleagues for research, co-location and frequent day-to-day contact facilitate coordination. Therefore, while international collaboration expands the scope of combinatorial search by bringing together diverse ideas and inputs from distant partners, within-university collaboration reduces the cost and thus increases the intensity of combinatorial search through frequent face-to-face interaction with co-located colleagues. However, and especially for small countries such as New Zealand, the intermediate category of domestic collaboration lacks the diversity of one and the intensity of the other. Hence, we have the following competing hypotheses:

Hypothesis 1a. At article level, international collaboration is more positively related to an article's quality than is domestic or within-university collaboration.

Hypothesis 1b. At article level, within-university collaboration is more positively related to an article's quality than is domestic or international collaboration.

Third, at scientist-year level, learning and social networking may operate differently at different distance ranges to influence a scientist's future research output. In the knowledge management literature, co-location is often highlighted as the most important factor in effective transfer of knowledge, especially tacit knowledge which is the focus of the learning perspective of research collaboration (Brown and Duguid, 1991). To the extent that a scientist is likely to spend more time interacting with her local collaborators, co-location is expected to improve the effectiveness of learning, holding constant the amount of knowledge to be learned.

However, local collaboration may not significantly contribute to a scientist's learning because of its limited scope for knowledge transfer. Compared with local partners from within the same university, partners from afar, especially those from another country, are more likely to possess ideas and techniques that are novel and non-overlapping for a scientist to learn (Burt, 1992). Although many of these ideas and techniques are tacit, distance per se is not a barrier to acquiring tacit knowledge from research partners because tacitness is not an intrinsic property of knowledge stock, but a property of knowledge flow (Breschi and Lissoni, 2001). Tacitness is also relative, and a scientist can always make investment to reduce the tacitness of knowledge with some partners at some locations (von Hippel, 1994). Due to this investment, together with the internet, video-conferencing, and regular visit, tacit knowledge can be learned between scientists over long distances as long as there are sufficient levels of mutual understanding and commitment.

Similarly, the ability of a scientist to develop network contacts is thought to vary with distance. Sociology research has long established that spatial proximity increases the probability of informal communication which, in turn, leads to network relationships. In fact, Katz (1994) found that the likelihood of co-authorship decreases exponentially with the distance separating pairs of partners. However, this does not necessarily meant that collaboration at a local scale has a stronger impact on future research output than collaboration over long distances. On the one hand, local collaboration allows a scientist to be embedded in a densely interconnected local network that is characterized by high levels of social capital such as trust, shared beliefs, mutual obligations and expectations, and cooperative norms which, in turn, enhance her productivity in future research (Coleman, 1988). On the other hand, international collaboration can "plug" a scientist into a much wider network of global science and greatly expand her network advantage for her future research.

Taken together, we have the second pair of competing hypotheses:

Hypothesis 2a. At scientist-year level, international collaboration is more positively related to a scientist's future research output than is domestic or within-university collaboration.

Hypothesis 2b. At scientist-year level, within-university collaboration is more positively related to a scientist's future research output than is domestic or international collaboration.

Finally, it is important to investigate the flow of causality between collaboration and research output. There is always potential for joint determination between collaboration and research output because past quality research can bring more chances for collaboration through various mechanisms, such as increased visibility and stronger funding support. We are not aware of any study that has established whether research collaboration induces research output or results from past research output. The concern here is that if we find a causal direction from research output to research collaboration but not the other way around, our ambitious policies to encourage research collaboration will be less justifiable because we might have been mistaking means for ends. Specifically, we make the following competing hypotheses:

Hypothesis 3a. At scientist-year level, research collaboration causes research output more than research output causes research collaboration.

Hypothesis 3b. At scientist-year level, research output causes research collaboration more than research collaboration causes research output.

3. Data and methods

3.1. Using co-authorship to measure research collaboration

Since the pioneering work of Price and Beaver (1966), co-authorship has been widely used as a direct measure of research collaboration at individual, organization, regional and country levels (McFadyen and Cannella, 2004; Cockburn and Henderson, 1998; Luukkonen et al., 1993). However, like any measure, co-authorship is no more than a partial indicator of research collaboration. There are two concerns in using co-authorship as a measure of research collaboration (Melin and Persson, 1996):

(a) Research collaboration does not always lead to co-authored articles. For example, a researcher may provide a key idea for an article but, for some reason, does not appear as a co-author.
(b) Co-authorship can arise without research collaboration. A researcher may be listed as a co-author simply by providing experiment materials or performing a routine test. A person may be listed as a co-author not by virtue of collaboration but because of position or prestige ("honorary co-author").

Situation (a) implies the risk of understating research collaboration if we focus on co-authorship alone. As for situation (b), we think it may not pose a serious threat to our empirical design because co-authorship in such cases suggests at least some level of collaborative relationship and possibilities of future collaboration.

Despite various limitations, objective data like co-authorship have four key advantages: verifiability, stability over time, unobtrusiveness, and ease of measurement (Katz and Martin, 1997). In our case, these advantages are crucial because our research questions required us to collect longitudinal data over a sufficiently long period. Self-reported collaboration measures based on surveys or interviews were not an option for us. One would not expect a scientist to accurately and consistently remember with whom she collaborated 14 years ago, 13 years ago, and so on.

In scientist-year level analyses, we were able to alleviate the concern that co-authorship on articles underestimates the extent of collaboration by including articles, reviews, research notes, book reviews, letters, editorial materials, etc. (but not corrections) when coding collaborative activities. Our research output variables were based on the more restrictive definition of articles, reviews (but not book reviews) and research notes. Our approach improves on previous studies that have used the same set of publications to code both dependent and independent variables. In the following sections we refer to all items included in research outputs as "papers", and to the wider set of items used to identify collaborations as "publications" or "documents". We used the ISI Web of Science database to identify both papers and publications.

3.2. Sample

Our sample consists of 65 biomedical scientists from a New Zealand university. A single institution was chosen for this study for a number of reasons. First, as previously mentioned, to accurately code different types of collaboration, one has to check the full text copy of each publication and identify each co-author's exact location because the ISI Web of Science database does not provide one-to-one correspondence between authors and affiliations. We are not aware of any study that has coded co-authorship data based on visual inspection of each publication included in the study sample, but doing so for many scientists from multiple institutions would be prohibitively expensive in time and resources. Second, two of us were working at this university when this study was conducted and we had access to personnel information of these biomedical scientists, such as age, gender, title, department affiliation, promotion, administrative position, etc. These demographic variables must be included to isolate a reliable relationship between research collaboration and research output. Third, the proximity with these biomedical scientists allowed us to contact them or their departments for clarification whenever a suspicious case arose. For example, we found a paper co-authored by one of the 65 scientists, but his university address did not appear on that paper. By contacting this scientist, we confirmed that it was his paper but the corresponding author had incorrectly put him under another address. This verification process was crucial to ensure the accuracy and completeness of our data. Fourth, by focusing on one university we were able to rule out some confounding effects caused by different policies and sizes of in-house expertise across different universities. Nevertheless, as pointed out by one of the referees, the results of this study may not generalize to other institutions in different contexts, especially those outside New Zealand, a small country geographically isolated from the other world centres of science. While our narrowly defined sample, as can be seen below, was intended to maximize internal validity (causal inference), future research should apply our research design to large samples of different countries and examine our findings' external validity. With this admonition in mind, we turn to the construction of our sample.

We first downloaded from the ISI Web of Science all publications between 1990 and 2003 for which at least one author's address was at the university. Then, using the *University Calendar* which was published every year, we compiled a list of biomedical scientists of the university for the 14-year period. We carefully adjusted for spelling variants because the same scientist may appear in slightly different names in the ISI Web of Science. For example, Andre M. van Rij, one of the 65 scientists in our sample, published most of his papers under "van Rij AM", but also appeared as "vanrij AM" or "vanrij A" on a few other papers. In such case, these papers were assembled under one scientist, and we contacted the scientist directly for clarification if we were not sure.

Next, we used the following four criteria to select biomedical scientists into the sample: (a) full-time faculty member of the university's biomedical schools; (b) appointment with the university was either confirmed or on confirmation path (similar to tenured or tenure-track at North American universities), and no joint appointment with any other institutions; (c) took up continuous employment for 10 years or longer during the period from 1990 to 2003, excluding partial years of participation; (d) published 10 or more papers indexed in the ISI Web of Science database during this period. Criteria (a) and (b) were necessary because if a scientist took a part time or joint appointment with the university, her scores of collaboration could be artificially inflated by her striding across two institutions or even two countries. Criterion (c) was desirable for us to track a scientist's research collaboration continuously for a number of years. Criterion (d) ensured that a scientist demonstrated at least a threshold level of research activity.

One may challenge that these requirements are unduly restrictive and make the sample less representative, but we believe that these restrictions are necessary to code the research collaboration variables consistently across scientists and over time. This consistency is crucial for us to draw any causal inferences from longitudinal data. Although we cannot claim we have a representative sample, we are confident that our results are not driven by a selective sample of "star scientists" because research output varied enormously among these 65 scientists, ranging from over 100 papers to just 10 during the period of 14 years.

Seventy biomedical scientists fulfilled the above four criteria. However, three of them published a few papers which had an abnormally large number of co-authors (for example, more than 100 co-authors), and thus were dropped from the sample.[1] Although such giant collaboration of "big science" may represent a unique structure for collaborative research that deserves a close investigation, this topic is beyond the scope of this paper and we did not have enough data to explore this issue. A further two scientists with very common surnames (Taylor and Zhang) were also dropped because identification of their papers and citations to these papers could not be done with high accuracy.

The resulting final dataset had 65 biomedical scientists who had 2244 publications among which 1860 are papers (1670 articles, 97 reviews, and 93 research notes).[2] We then aggregated these data to scientist-year level to construct a longitudinal dataset. Because not all the scientists worked for the university throughout the 14-year period, we had an unbalanced panel of 850 usable scientist-year observations.

As pointed out by Hood and Wilson (2003), electronic databases like the ISI Web of Science have errors of many kinds and, to ensure the integrity of data, scrutiny procedures must be followed. Thanks to the very large collections of biomedical research journals at the university, we were able to retrieve the full text version for 90% of these 2244 publications, and we also obtained the full text version for most of the remaining 10% from our friends at other universities. We carefully hand checked each of the 2244 publications to correct any obvious errors.[3]

3.3. Variables

Table 1 presents variables and measures for article level analyses. Most of these variables are self-explanatory, so we only briefly mention a few key issues here. First, we used two measures of paper quality: one is 5-year (1999–2003) average impact factor of the journal in which a paper was published, and the other is number of citations a paper received in a 2-year window excluding self-citations. We also experimented with a 3-year citation window, but the results reported in the next section were not affected mostly because the 2-year and 3-year measures are highly correlated with each other (the correlation between the two is 0.97 in our sample); hence we chose to use a 2-year citation window that is consistent with the ISI Web of Science algorithm to calculate journal impact factors. Total citation counts cannot be used as a measure of quality across papers because everything else being equal, papers published earlier in time would naturally receive more citations. Self-citations must be excluded because a multi-authored paper can accumulate more citations simply by each of the co-authors subsequently publishing a separate paper that cites their joint paper.[4] Failure to correct for self-citations may therefore produce spurious relationships between research collaboration and research output. Impact factor data before 1998 were not available to us. We are aware that average impact factor in recent years can be a noisy measure of quality for a paper published 10 years back in time, but we hope number of citations can better operationalize this construct.

Second, in coding research collaboration variables, we used strict count of co-authors, not addresses as in other studies. As previously mentioned, achieving this required us to painstakingly check every publication of these 65 scientists by visual inspection. There were no other less labour intensive choices and we believed this was absolutely necessary to overcome the shortcomings of hierarchical coding. It is worth mentioning that a co-author may give more than one institutional address because she has a joint appointment at two institutions or on sabbatical outside her home institution. In such cases, the best solution was to code this author as "multiple co-authors" because it is highly uncertain in terms of which type of research collaboration this co-authorship should reveal. On the other hand, a co-author of multiple institutional addresses may well bring more resources and more learning and networking opportunities into research collaboration.[5]

Variables for scientist-year level analyses are summarized in Table 2. The key difference here, as previously mentioned, is that collaboration variables were coded using all the 2244 publications of various types, not just the 1860 papers of these 65 scientists.[6] We measured research output of a scientist in a year in three different

[1] Among all the publications included in our sample, the highest number of authors on a publication is 16, and only 8 publications have 10 or more authors.

[2] "Non-paper" publications by these 65 scientists include 48 editorial materials, 109 letters, 225 meeting abstracts, one discussion, and one item about an individual.

[3] Among these 2244 publications, we found that the ISI Web of Science database has mistakes for 55 of them. Sample errors include but are not limited to (a) Auckland becomes a part of Australia, (b) Dunedin becomes a city in Scotland, (c) a letter is misclassified as an article, (d) a title like "Ralph Barnett Professor of Surgery" is recorded as a co-author.

[4] Following Noyons et al. (1999, p. 116) and Wuchty et al. (2007, p. 1039), we define a self-citation as any citation where at least one author appears on both the cited and the citing publications. More specifically, we carefully identified self-citations in the following steps: (1) when there were two or more cases of full name match (identical surnames and initials) between the cited and the citing publications, it was accepted as an instance of self-citation; (2) when there was one case of full name match and the shared surname was not a common one (e.g., Barbezat), it was accepted as an instance of self-citation; (3) when there was one case of full name match and the shared surname is a common one (e.g., Smith), further search was conducted by comparing addresses and research themes between the pair of publications; (4) when there was a case of partial name match (identical surnames but not all initials were identical, e.g., Hurst PR and Hurst P), the same search described in step 3 was conducted to confirm the instance of self-citation. This stepwise approach was necessary to avoid two mistakes: (a) incorrectly accept a non-self-citation, where the authors share the same surnames and initials but actually they are different people, as an instance of self-citation; and (b) incorrectly reject an instance of self-citation where the authors are the same people but use slightly different initials in different publications. A similar procedure was used by Meyer (2006, p. 1652).

[5] But note that a co-author with two or more department addresses of the same institution was not coded as "multiple co-authors".

[6] We are aware that journal papers are no more than a partial indicator of a scientist's total research output. Other forms of research output like books, book chapters,

Table 1
Variables for article level analysis.

Variable	Description	Remarks
Paper Quality1	5-year (1999–2003) average impact factor of the journal in which a paper was published	Dependent variable Source: computed from the ISI *Web of Science*
Paper Quality2	Number of citations a paper received in a 2-year window excluding self-citations	Dependent variable Source: computed from the ISI *Web of Science*
International collaboration	Number of co-authors from outside New Zealand on a paper	Source: The ISI *Web of Science* and visual inspection of each paper
Domestic collaboration	Number of co-authors from within New Zealand but outside the University on a paper	Source: The ISI *Web of Science* and visual inspection of each paper
Within-university collaboration	Number of co-authors from the University on a paper excluding the focal scientist herself	Source: The ISI *Web of Science* and visual inspection of each paper
Number of references	Number of references cited by a paper	Source: The ISI *Web of Science*
Page count	Number of pages of a paper	Source: The ISI *Web of Science*
Paper type dummies	Article, Review, Note	Article is used as the base category Source: The ISI *Web of Science*
Administrative	A dummy variable indicating whether the focal scientist of a paper took an administrative position at the year of publication of the paper (1 = "yes", 0 = "no")	Only department head and school dean are counted. This variable is time-varying because most heads and deans did not hold their positions throughout the period from 1990 to 2003. No scientist in the sample ever took a university level position during this period. Source: *University Calendar*
Overseas PhD/MD	A dummy variable indicating whether the focal scientist of a paper had a doctoral or MD degree from outside New Zealand (1 = "yes", 0 = "no")	Source: *University Calendar*
Academic rank dummies	Dummy variables indicating whether the focal scientist of a paper was a Professor, Associate Professor, Senior Lecturer, or Lecturer at the year of publication of the paper	Lecturer is used as the base category Source: *University Calendar*
Gender	A dummy variable indicating the gender of the focal scientist of a paper (1 = "male", 0 = "female")	Source: *University Calendar* and department web pages
Department dummies	Dummy variables indicating department affiliation of the focal scientist of a paper: Anatomy and structural biology, Biochemistry, Dentistry and oral sciences, Medical and surgical sciences, Microbiology and immunology, Pharmacy, Physiology, Others	"Others" is used as the base category "Dentistry and oral sciences" covers a number of related departments, including Oral diagnostic and surgical sciences, Oral rehabilitation, Oral sciences, and Stomatology. "Others" covers the remaining departments with a small number of scientists included in this study, including General practice, Pathology, Pharmacology and toxicology, Preventive and social medicine, Women's and children's health. Source: *University Calendar*
Year dummies	Dummy variables indicating the year of publication of a paper	Source: The ISI *Web of Science*
Scientist dummies	Dummy variables indicating the focal scientist of a paper	Source: The ISI *Web of Science* and *University Calendar*

ways. Research Output1 is a direct count of papers, but a "note" was only counted as a half paper. This dependent variable only captures the quantity of research output. Research Output2 and Research Output3 further capture the quality of research output and account for authorship sequence.[7] Suppose a focal scientist had m papers in a certain year.

$$\text{Research Output2} = \sum_{i=1}^{m} W_i I_i A_i$$

where $W_i = 1$ if paper i is an "article" or "review", 0.5 if paper i is a research "note", I_i is the impact factor of the journal in which paper i is published, A_i is authorship index and takes the value of 1 if the focal scientist is the first author or corresponding author, 0.5 if the second author (but not corresponding author), 0.25 if the third author (but not corresponding author), and 0.1 in other cases.[8]

$$\text{Research Output3} = \sum_{i=1}^{m} C_i A_i$$

papers in journals not indexed in the ISI Web of Science, and working papers etc., were not included for two reasons: (1) it is almost impossible to objectively evaluate the quality of these types of research output; (2) as far as we know, journal papers, especially papers in journals indexed in the ISI Web of Science, are the most important form of research output for biomedical scientists in universities.

[7] As suggested by one of the referees, a note is poorly defined document type compared to an article or review. We recalculated these three measures of research output either excluding notes all together or including notes as full papers. We found nearly identical results as reported in the next section. This is not surprising given the small percentage of notes in the sample (5%).

[8] Three schemes have been suggested in the literature to calculate research output: straight count, normal count, and adjusted count (Lindsey, 1980). Straight count only considers first-authored or sole-authored papers; normal count includes all papers published by a scientist and gives full credit to her regardless of number of authors on a paper; adjusted count includes all papers published by a scientist but each is divided by number of authors on a paper. None of these three schemes would allow us to achieve reasonable accuracy in calculating a scientist's yearly research output, which is critically important because we are analyzing research output at individual scientist level. Straight count ignores many non-first-authored papers; normal count greatly inflates a scientist's research output; adjusted count does not utilize information on authorship sequence. Our approach here is similar to the scheme used by the Shanghai Group in their *Academic Ranking of World Universities* (Liu et al., 2005, p. 103) except that the Shanghai Group allocates only 50% to the first author. In biomedical sciences and other natural sciences, reprint (corresponding) author is often the senior author who oversees the project as principal investigator. However, reprint author is not always the last author, and vice versa.

Table 2
Variables for scientist-year level analysis (fixed-effects model)[a].

Variable	Description	Remarks
Research output1	Number of papers published by a scientist each year	Dependent variable While an "Article" or "Review" is counted as one paper, a "Note" is counted as a half paper. Source: computed from the ISI *Web of Science*
Research output2	Impact factor weighted and authorship adjusted papers published by a scientist each year	Dependent variable Source: computed from the ISI *Web of Science*
Research output3	Number of citations to a scientist's papers of year t in a 2-year window (year $t+1$ and year $t+2$), excluding self-citations and adjusted for authorship	Dependent variable Source: computed from the ISI *Web of Science*
International collaboration	Number of co-authors from outside New Zealand on a publication	Aggregated to scientist-year level Source: computed from the ISI *Web of Science* and visual inspection of each publication
Domestic collaboration	Number of co-authors from within New Zealand but outside the University on a publication	Aggregated to scientist-year level Source: computed from the ISI *Web of Science* and visual inspection of each publication
Within-university collaboration	Number of co-authors from the University on a publication excluding the focal scientist herself	Aggregated to scientist-year level Source: computed from the ISI *Web of Science* and visual inspection of each publication
Administrative	A dummy variable indicating whether a scientist took an administrative position in a year (1 = "yes", 0 = "no")	Only department head and school dean are counted. This variable is time-varying because most heads and deans didn't hold their administrative positions for the whole period from 1990 to 2003. No scientist in the sample once took a university level position during this period. Source: *University Calendar*
Promotion	A dummy variable indicating whether a scientist was promoted to a higher academic rank in a year (1 = "yes", 0 = "no")	Source: *University Calendar*
Year dummies	A full set of year dummies are included in the scientist-year level analysis	

[a] Fixed-effects models control for unobserved heterogeneity between scientists by including a separate intercept for each scientist. Therefore, no time-invariant variables, such as Overseas PhD/MD, gender, and department dummies as shown in Table 1, can be included in fixed-effects models. No scientist in our sample changed their department affiliation during the 14-year period except department name changes (e.g., from Department Microbiology to Department of Microbiology and Immunology) and mergers of a few small departments to form large departments (e.g., Department of Medical and Surgical Sciences), which are not regarded as changes of department affiliation.

where A_i is similarly defined as in Research Output2, and C_i is number of citations received by paper i in a 2-year window, excluding self-citations.

Different types of paper were not distinguished for Research Output3 because there should be no material differences between citations made to an article, a review and a research note. One may argue that "non-paper" publications also receive citations, though much less frequently than "paper" publications. In unreported additional analyses, we found that including those citations did not introduce any material change to our results. In the following section, we will only report the results using "paper-based" Research Output3.

To better isolate the impact of research collaboration, we controlled for two time-varying scientist characteristics: "administrative" and "promotion". "Administrative" was included because a scientist holding an administrative position may have less time for research but may also have a larger social network for possible research collaboration. "Promotion" was included to capture the effect of a scientist's career path. This variable takes the value of 1 when a focal scientist is promoted along the ranks of Lecturer, Senior Lecturer, Associate Professor, and Professor. For example, when a focal scientist is a Senior Lecture in year t and Association Professor in year $t+1$, this variable takes the value of 1 for year t.

Another issue in scientist-year level analyses is related to the life cycle theory of research productivity (Levin and Stephan, 1991; Gonzalez-Brambila and Veloso, 2007), which predicts individual scientists' research productivity to follow an inverted U curve. But in fixed-effects models, any variable whose change across time is constant (like age) cannot be included as regressors together with a full set of year dummies. We elected to include a full set of year dummies because they can control many factors that vary over time and affect all scientists in the sample, such as government and university policy changes that may cause aggregate fluctuations in their research output. Since we did not observe the whole life-cycle of these scientists, year dummies should be a stronger set of controls than scientist age to isolate the relationship between research collaboration and research output.

4. Results

4.1. Article level analysis

Table 3 presents the analysis at article level using Paper Quality1 as the dependent variable. In Model 1, the OLS regression includes three collaboration variables, all the control variables, and a full set of year dummies, but not scientist dummies. It shows that both international collaboration and within-university collaboration are positively ($p < 0.001$) related to a paper's quality as measured by impact factor. In Model 2, a full set of scientists dummies are included, but time-invariant dummies, including overseas PhD/MD, gender, and department dummies have to be dropped otherwise the regression would be overcome by multicollinearity.

Table 3
Article level analysis using Paper Quality1 as the dependent variable[a].

	Model 1 (OLS)	Model 2 (OLS)	Model 3 (GEE)
Intercept	1.325**(0.480)	2.678**(0.819)	1.868***(0.488)
International collaboration (α_1)	0.271***(0.038)	0.176***(0.038)	0.197***(0.046)
Domestic collaboration (α_2)	0.048(0.064)	0.080(0.066)	0.068(0.099)
Within-university collaboration (α_3)	0.284***(0.045)	0.226***(0.046)	0.239*(0.102)
Paper characteristics			
Number of references	0.021***(0.003)	0.019***(0.003)	0.020***(0.004)
Page count	−0.085***(0.018)	−0.074***(0.018)	−0.076**(0.027)
Review	0.427(0.368)	0.297(0.362)	0.310(0.295)
Note	−0.543 (0.314)	−0.557(0.309)	−0.546(0.341)
Scientist characteristics			
Administrative	−0.266(0.253)	−0.053(0.311)	−0.055(0.257)
Overseas PhD/MD	−0.124(0.146)		−0.161(0.273)
Professor	0.059(0.235)	−0.793(0.513)	−0.301(0.361)
Associate professor	0.435(0.242)	0.131(0.396)	0.349(0.296)
Senior lecturer	−0.122(0.206)	−0.047(0.278)	−0.032(0.229)
Gender (1 = male)	0.605(0.330)		0.383(0.361)
Department dummies			
Anatomy and structural biology	1.008***(0.280)		0.737*(0.348)
Biochemistry	2.449***(0.216)		1.912**(0.603)
Dentistry and oral sciences	−0.579*(0.247)		−0.881**(0.290)
Medical and surgical sciences	0.126(0.235)		−0.005(0.298)
Microbiology and immunology	0.213(0.242)		0.031(0.282)
Pharmacy	−0.078(0.246)		−0.353(0.256)
Physiology	0.567*(0.285)		0.380(0.385)
Year dummies	Included	Included	Included
Scientist dummies	Not included	Included	Not included
Test for $\alpha_1 = \alpha_2$	F = 9.47**	F = 1.74	F = 1.56
Test for $\alpha_1 = \alpha_3$	F = 0.07	F = 0.97	F = 0.27
Test for $\alpha_2 = \alpha_3$	F = 10.41**	F = 3.87*	F = 3.39
R-squared	0.21	0.29	
Adjusted R-squared	0.19	0.25	
Wald Chi-square			270.43
Number of observations	1860	1860	1860

***$p < .001$, **$p < .01$, *$p < .05$, two-tailed tests.
[a] Standard errors are in parentheses.

With this set of dummies to better control for differences between scientists, international collaboration and within-university collaboration are still significantly related to Paper Quality1 ($p < 0.001$) but with smaller coefficients than in Model 1. In Model 3, the generalized estimating equations (GEE) method is used to account for the clustered nature of observations (i.e., observations are clustered around these scientists). GEE regressions can achieve more reliable estimates by taking into account the possibility that observations on a given scientist are more correlated than those between different scientists. There are no material changes to the coefficients of collaboration variables in Model 3.

Table 4 presents the analysis at article level using Paper Quality2 as the dependent variable. We address the discrete and non-negative nature of citations by adopting a negative binomial specification, which is a generalized form of Poisson model. We followed Cameron and Trivedi (1998) to perform a likelihood ratio test of over-dispersion and the assumption of equal mean and variance in the dependent variable was rejected, suggesting preference for a negative binomial specification. The specifications in Models 4–6 are similar to those in Models 1–3 in Table 3, and the GEE method is also used in Model 6 to account for the clustered nature of observations. Similar results are obtained for Paper Quality2.

Overall, the results in Tables 3 and 4 consistently show that both international collaboration and within-university collaboration are positively related to the quality of a paper published by a scientist regardless of the choice of dependent variable and model specification, but the coefficient of domestic collaboration is never significant. In all six regressions, the coefficient size of international collaboration is smaller than within-university collaboration, but Wald tests for coefficient equality do not confirm that the difference is statistically significant. Therefore, Hypotheses 1a and 1b are both partially supported.

4.2. Scientist-year level panel data analysis

As discussed in Section 2, we want to have an estimate for the relationship between last year's collaboration and this year's research output within scientists, not between scientists. Fixed-effects models achieve this purpose by including scientist fixed-effects to account for unobserved heterogeneity between scientists. Another reason why fixed-effects models are preferred here is that fixed-effects models allow for arbitrary correlation between unobserved heterogeneity and explanatory variables, which is very likely in our case. For example, a scientist's personality and proficiency in a foreign language are not observed in our study, but they may be correlated with her preferences in the form and intensity of research collaboration. We ran a Hausman test and found preference for fixed-effects models over random-effects models in all regressions. We assumed a 1-year lag between our regressors and dependent variables to avoid too much loss of degrees of freedom. A longer lag structure is investigated below in Section 4.3.

We included lagged research output as an additional explanatory variable to control for path dependence of a scientist's research output over time. A lagged dependent variable also gives another advantage because it helps control for unobserved factors that vary across time and within scientists. This is important because we only have two time-varying control variables ("administrative" and

Table 4
Article level analysis using Paper Quality2 as the dependent variable[a].

	Model 4 (negative binomial)	Model 5 (negative binomial)	Model 6 (GEE)
Intercept	−0.464(0.270)	−0.450(0.450)	−0.104(0.359)
International collaboration (α_1)	0.150***(0.021)	0.109***(0.020)	0.121***(0.027)
Domestic collaboration (α_2)	0.016(0.035)	0.023(0.037)	0.030(0.051)
Within-university collaboration (α_3)	0.190***(0.023)	0.125***(0.024)	0.177***(0.056)
Paper characteristics			
Number of references	0.016***(0.002)	0.013***(0.002)	0.014***(0.001)
Page count	−0.032**(0.010)	−0.014(0.011)	−0.028*(0.012)
Review	−0.237(0.188)	−0.202(0.183)	−0.190(0.193)
Note	−0.228(0.169)	−0.281(0.160)	−0.319(0.210)
Scientist characteristics			
Administrative	−0.215(0.135)	−0.023(0.162)	−0.144(0.156)
Overseas PhD/MD	0.098(0.076)		0.049(0.190)
Professor	−0.154(0.127)	0.017(0.270)	−0.158(0.216)
Associate professor	−0.236(0.132)	0.062(0.211)	−0.107(0.171)
Senior lecturer	−0.299**(0.110)	0.002(0.152)	−0.184(0.156)
Gender (1 = male)	0.425*(0.178)		0.276(0.296)
Department dummies			
Anatomy and structural biology	0.500***(0.146)		0.337(0.368)
Biochemistry	0.815***(0.113)		0.673**(0.244)
Dentistry and oral sciences	−0.244(0.137)		−0.337(0.265)
Medical and surgical sciences	0.257*(0.129)		0.165(0.297)
Microbiology and immunology	0.150(0.131)		0.100(0.303)
Pharmacy	−0.296*(0.134)		−0.329(0.191)
Physiology	−0.150(0.155)		−0.223(0.253)
Year dummies	Included	Included	Included
Scientist dummies	Not included	Included	Not included
Test for $\alpha_1 = \alpha_2$	F = 11.34***	F = 4.71*	F = 2.36
Test for $\alpha_1 = \alpha_3$	F = 2.13	F = 0.37	F = 1.31
Test for $\alpha_2 = \alpha_3$	F = 19.12***	F = 6.54*	F = 8.79**
Log likelihood	−3952.59	−3837.90	
Wald Chi-square			684.65
Number of observations	1860	1860	1860

***$p < .001$, **$p < .01$, *$p < .05$, two-tailed tests.
[a] Standard errors are in parentheses.

"promotion"),[9] but scientist fixed-effects only control for stable scientist characteristics such as gender, personality, "imprint effect" of her PhD or MD training, and year dummies only control for exogenous factors that influence all scientists. Finally, squared lagged research output was also included to control for non-linearity between past performance and current performance. We often see that after a surge of research output in year $t-1$, a scientist may decide to "take a break" in year t.

The results of fixed-effects regressions are presented in Table 5. In Model 7 where quantity of research output is the dependent variable, only within-university collaboration is positively related to next year's research output. In Model 8 where Research Output2, the impact factor weighted and authorship adjusted research output, is used as the dependent variable, the positive relationship between international collaboration and next year's research output is highly significant ($p < 0.001$). When Research Output3, the citation weighted and authorship adjusted research output is used as the dependent variable in Model 9, similar results are found except a much stronger non-linear relationship between past performance and current performance. Hypothesis 2a is supported in both Model 8 and Model 9, where Wald tests for coefficient equality suggest that international collaboration is more positively related to future research output than is domestic or within-university collaboration.

4.3. Causality analysis

As shown in Table 5, among the three collaboration variables, only international collaboration is significantly related to a scientist's future research output that is not purely measured by quantity, so we only try to detect the possible reverse causality between international collaboration and research output. Using Research Output2 or Research Output3 does not materially change the results of causality analyses, and thus we will only report the results based on Research Output3.

In Table 6, we compare the regression with international collaboration as the dependent variable (Model 11) and the regression with research output as the dependent variable (Model 10), with all explanatory variables lagged by just 1 year. Note that fixed-effects negative binomial regression is used in Model 11 because international collaboration is a count variable. We find that international collaboration has a strong impact on future research output in Model 10 but that research output does not have any significant impact on future international collaboration in Model 11.

In Table 7, a similar comparison is performed with a 3-year window, i.e., all explanatory variables are the sum of corresponding values in previous 3 years (from year $t-3$ to year $t-1$) except lagged dependent variables. Similar results are obtained: while international collaboration in previous 3 years has a strong impact on current research output in Model 12, research output in previous

[9] If there are any other scientist-specific time-varying factors which significantly affect both research collaboration and research output, our results may suffer from omitted variable bias. For example, sometime during the 14-year period of this study, a scientist shifted her aspiration level which made her both more dedicated to research and more enthusiastic about collaboration. Still, another possibility is that she redirected her research to a sub-field that has more potential for high quality output and more opportunity for collaboration. If these are the case, our results will overstate the impact of research collaboration on research output. A lagged dependent can help reduce such bias but will not completely rule out alternative explanations for the observed relationship between research collaboration and research output.

Table 5
Scientist-year level analysis (fixed-effects model)[a].

	Model 7 Research output1	Model 8 Research output2	Model 9 Research output3
Research output$_{t-1}$	−0.140(0.092)	0.130*(0.066)	0.341***(0.063)
Research output$_{t-1}$ squared	0.002(0.007)	−0.003*(0.001)	−0.002***(0.0004)
International collaboration$_{t-1}$ (β_1)	0.013(0.021)	0.193***(0.048)	0.524***(0.107)
Domestic collaboration$_{t-1}$ (β_2)	0.025(0.038)	−0.094(0.089)	−0.375(0.198)
Within-university collaboration$_{t-1}$ (β_3)	0.079***(0.021)	0.033(0.037)	−0.009(0.074)
Administrative$_{t-1}$	−0.338(0.365)	−0.762(0.892)	−0.897(1.984)
Promotion$_{t-1}$	0.369(0.247)	1.538*(0.601)	2.635*(1.334)
Year dummies	Included	Included	Included
Test for $\beta_1 = \beta_2$	F = 0.08	F = 7.51**	F = 14.96***
Test for $\beta_1 = \beta_3$	F = 5.99*	F = 6.43*	F = 14.93***
Test for $\beta_2 = \beta_3$	F = 1.58	F = 1.57	F = 2.65
R-squared within	0.05	0.06	0.11
R-squared between	0.33	0.57	0.70
Number of observations	785	785	785

***p < .001, **p < .01, *p < .05, two-tailed tests.
[a] Standard errors are in parentheses.

Table 6
Causality analysis I[a].

	Model 10 (fixed-effects) dependent variable: research output3	Model 11 [b] (fixed-effects negative binomial) dependent variable: international collaboration
Research output$_{t-1}$	0.341***(0.063)	0.001(0.003)
Research output$_{t-1}$ squared	−0.002***(0.0004)	
International collaboration$_{t-1}$	0.524***(0.107)	0.083***(0.026)
International collaboration$_{t-1}$ squared		−0.002*(0.001)
Domestic collaboration$_{t-1}$	−0.375(0.198)	0.029(0.023)
Within-university collaboration$_{t-1}$	−0.009(0.074)	0.026***(0.008)
Administrative$_{t-1}$	−0.897(1.984)	0.358(0.243)
Promotion$_{t-1}$	2.635*(1.334)	0.331*(0.171)
Year dummies	Included	Included
R-squared within	0.11	
R-squared between	0.70	
Log likelihood		−1000.84
Number of observations	785	762 [c]

***p < .001, **p < .01, *p < .05, two-tailed tests.
[a] Standard errors are in parentheses.
[b] The fixed-effects negative binomial regression can include time-invariant variables because "fixed effects" in a negative binomial specification apply to the distribution of the dispersion parameter, not to explanatory variables. In unreported analyses, the main results in Model 11 do not change materially when time-invariant variables (department dummies, gender, Overseas PhD/MD) are included.
[c] Compared with Models 10, 23 observations are lost due to all zero outcomes of 2 scientists. That is, these 2 scientists have a zero count of international collaboration for each year during the period under investigation.

Table 7
Causality analysis II[a].

	Model 12 (fixed-effects) dependent variable: research output3	Model 13[b] (fixed-effects negative binomial) dependent variable: international collaboration
Research output3$_{t-1}$	0.432***(0.077)	
Research output3$_{t-1}$ squared	−0.002***(0.0004)	
International collaboration$_{t-1}$		0.048(0.029)
International collaboration$_{t-1}$ squared		−0.001(0.001)
Research output3 $_{(t-3,\, t-1)}$		0.002(0.002)
International collaboration$_{(t-3,\, t-1)}$	0.195**(0.072)	
Domestic collaboration$_{(t-3,\, t-1)}$	−0.271(0.142)	0.020(0.015)
Within-university collaboration$_{(t-3,\, t-1)}$	−0.014(0.046)	0.014**(0.004)
Administrative$_{(t-3,\, t-1)}$	−0.563(1.019)	0.118(0.106)
Promotion$_{(t-3,\, t-1)}$	0.043(1.091)	0.037(0.139)
Year dummies	Included	Included
R-squared within	0.09	
R-squared between	0.72	
Log likelihood		−816.34
Number of observations	655[c]	636[d]

***p < .001, **p < .01, *p < .05, two-tailed tests.
[a] Standard errors are in parentheses.
[b] The fixed-effects negative binomial regression can include time-invariant variables because "fixed effects" in a negative binomial specification apply to the distribution of the dispersion parameter, not to explanatory variables. In unreported analyses, the main results in Model 13 do not change materially when time-invariant variables (department dummies, gender, Overseas PhD/MD) are included.
[c] Compared with Models 10 in Table 6, 130 observations are lost due to further lagged independent variables.
[d] Compared with Model 12, 19 observations are lost due to all zero outcomes of 2 scientists. That is, these 2 scientists have a zero count of international collaboration for each year during the period under investigation.

3 years does not have any significant impact on current international collaboration in Model 13. The results in Tables 6 and 7 indicate that the causal direction should be from (international) collaboration to research output.

We also ran the Granger test to obtain a formal answer to the causality question (Granger, 1969). Generally, in a regression of Y on other variables and its own lagged values, if we include lagged values of X and X significantly improves the predication of Y, then we say that X causes Y. We performed the Granger causality test in two steps. In the first step, based on the specification in Model 10 in Table 6, we added another two lagged values of research output, research output$_{t-2}$ and research output$_{t-3}$, and found that international collaboration$_{t-1}$ still causes research output ($F = 25.41$, $p < 0.001$). In the second step, based on the specification in Model 11 in Table 6, we added another two lagged values of international collaboration, international collaboration$_{t-2}$ and international collaboration$_{t-3}$, and found that research output$_{t-1}$ does not cause international collaboration ($F = 0.27$, $p = 0.603$). Taken together, Hypothesis 3a is supported and we are reasonably assured that international collaboration causes research output, not the other way around.[10]

5. Conclusion

There has been a growing awareness of the importance of research collaboration in the science community and among policy makers. The benefits of research collaboration to individual collaborating scientists are commonly believed to outweigh its costs. While the literature has proposed various theoretically satisfying explanations for a positive relationship between research collaboration and research output, this hypothesis has not been rigorously tested due to three methodological barriers: (a) loss of information on local collaboration in hierarchical coding, (b) confounding effects of common latent variables such as a scientist's ability, and (c) the lack of longitudinal analysis. In this paper, we have tried to bridge these three gaps in the literature by examining the influence of international collaboration, domestic collaboration and within-university collaboration on research output at both article level and scientist-year level. By meticulously hand checking each publication, we have overcome problem (a). By including a full set of scientist dummies or using the GEE method, we believe we have sufficiently addressed problem (b). By constructing a longitudinal dataset of 65 biomedical scientists over a period of 14 years and performing fixed-effects panel data analyses and causality analyses, we have addressed concern (c).

Our analyses generate the following two headline results. First, both international collaboration and within-university collaboration are significantly related to the quality of an individual paper, even after including a full set of scientist dummies or controlling for the clustered nature of observations. In contrast to commonly held assumption, our results do not show that international collaboration is more strongly related to paper's quality than is within-university collaboration. We therefore suspect that the importance and contribution of local collaboration may be severely understated in existing studies.

Second, our longitudinal data analyses have shown that while within-university collaboration is related to future quantity of research output, only international collaboration is related to "real" research output of a scientist that takes quality and authorship into account. Our causality analyses also show that research collaboration causes research output more than research output causes research collaboration. Although our results must be interpreted with caution because we only had a small sample of 65 scientists, these results in general lend support to government and university initiatives that promote international research collaboration with the hope to improve future research output. We understand that impact factor and citation counts cannot measure quality perfectly and our allocation of credit to different authorship positions is somewhat arbitrary. However, one would not deny that this is better than not controlling for quality and authorship at all.

Combining the article level analyses and the scientist-year level analyses, we may conclude that probably all the three perspectives of scientific collaboration discussed in Section 2 are driving the relationship between research collaboration and research output. While the results of the article level analyses can be explained by the knowledge recombination perspective, however, we could not empirically distinguish between the learning explanation and social networking explanation because they are both consistent with the results that international collaboration is positively related to future research output in the scientist-year level analyses. We were unable to address this issue with the data available to us. Moreover, although our results support the widely held belief that research collaboration normally brings greater benefits than costs, our empirical design did not allow us to directly compare benefits and costs of research collaboration. We bring up these issues as a suggestion for future research that may follow from our study.

A peripheral finding of our study is that the influence of domestic collaboration is significant in none of the regressions, neither in the article level analyses nor in the scientist-year level analyses. We must interpret this result with caution and within the context of New Zealand. Most of these domestic collaborative relationships involve a government research institute and sometimes a private company in New Zealand. In New Zealand, government research institutes are the major channel of knowledge transfer of public research to private sectors. Unfortunately, it is very difficult to estimate such spillover benefits from these university researchers to government research institutes and private sectors. The conclusion should not be drawn, and it is NOT drawn here, that domestic collaboration does not generate benefits.

Finally, we do not believe that our results should be used to favour international collaborations over those within a scientist's own university. The latter have been shown here to be as powerful in lifting the quality of an individual paper. Also, our investigation of the causal structure between collaboration and research output reported in Models 11 and 13 indicates that those who learn collaboration at home are also more active in forging international collaborations. As Griffith and Miller (1970) pointed out nearly 4 decades ago, "individual scientists may be reluctant at one extreme, to travel seventy-five feet to utilize another person's store of knowledge but, at the other extreme, would willingly travel hundreds or thousands of miles to communicate with other persons" (cited from Katz, 1994, p. 41). Our own experience and stories from our colleagues in other universities tell us this is still the case today. Is it time to correct the under-appreciation of local collaboration?

Acknowledgements

We thank three anonymous referees and the editor for their thoughtful comments. Jasjit Singh, Ewan McComb, and seminar participants at Otago University, Auckland University and Tilburg University provided helpful comments on earlier drafts of this paper. All errors and omissions remain our own.

[10] It is, however, impossible to completely rule out the possibility of reverse causality that international collaboration is a consequence of research output. Granger causality tests have a strict statistical meaning of observational precedence, which may not be the same thing as theoretical causality. An interesting spurious Granger causality used by Ron Smith (Birkbeck College, University of London) is that weather forecasts can be shown to Granger cause the weather (cited from Athreye and Cantwell, 2007, p. 216).

References

Athreye, S., Cantwell, J., 2007. Creating competition? Globalisation and the emergence of new technology producers. Research Policy 36 (2), 209–226.

Bammer, G., 2008. Enhancing research collaborations: three key management challenges. Research Policy 37 (5), 875–887.

Breschi, S., Lissoni, F., 2001. Localized knowledge spillovers vs. innovative milieux: knowledge tacitness reconsidered. Papers in Regional Science 80 (3), 255–273.

Brown, J.S., Duguid, P., 1991. Organizational learning and communities of practices: towards a unified view of working, learning and innovation. Organization Science 2 (1), 40–57.

Burt, R.S., 1992. Structural Holes: The Social Structure of Competition. Harvard University Press, Cambridge, MA.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge.

Cockburn, I.M., Henderson, R.M., 1998. Absorptive capacity, coauthoring behaviour, and the organization of research in drug discovery. Journal of Industrial Economics 46 (2), 157–182.

Coleman, J.S., 1988. Social capital in the creation of human capital. American Journal of Sociology 94 (Suppl.), S95–S120.

Collins, H.M., 1974. The TEA set: tacit knowledge and scientific networks. Science Studies 4 (2), 165–186.

Cowan, R., David, P.A., Foray, D., 2000. The explicit economics of knowledge codification and tacitness. Industrial and Corporate Change 9 (2), 211–253.

Crane, D., 1972. Invisible Colleges. University of Chicago Press, Chicago, IL.

Cummings, J.N., Kiesler, S., 2007. Coordination costs and project outcomes in multi-university collaborations. Research Policy 36 (10), 1620–1634.

De Dreu, C.K.W., West, M.A., 2001. Minority dissent and team innovation: the importance of participation in decision making. Journal of Applied Psychology 86 (6), 1191–1201.

Duque, R.B., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D.B.S., Shrum, W., 2005. Collaboration paradox: scientific productivity, the Internet, and problems of research in developing areas. Social Studies of Science 35 (5), 755–785.

Glanzel, W., Schubert, A., 2001. Double effort = double impact? A critical view at international co-authorship in chemistry. Scientometrics 50 (2), 199–214.

Gonzalez-Brambila, C., Veloso, F.M., 2007. The determinants of research output and impact: a study of Mexican researchers. Research Policy 36 (7), 1035–1051.

Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37 (3), 424–438.

Griffith, B.C., Miller, A.J., 1970. Networks of informal communications among scientifically productive scientists. In: Nelson, C.E., Pollock, D.K. (Eds.), Communications Among Scientists and Engineers. Health Lexington Books, Lexington, MA.

Hood, W.W., Wilson, C.S., 2003. Informetric studies using databases: opportunities and challenges. Scientometrics 58 (3), 587–608.

Jager, M., 2006. Leistungsorientierte Budgetierung: Analyse der Umsetzung an ausgewählten Universitäten und Fakultäten/Fachbereichen. Hochschulinformationsystem A1/2006.

Katz, J.S., 1994. Geographical proximity and scientific collaboration. Scientometrics 31 (1), 31–43.

Katz, J.S., Hicks, D., 1997. How much is a collaboration worth? A calibrated bibliometric model. Scientometrics 40 (3), 541–554.

Katz, J.S., Martin, B.R., 1997. What is research collaboration? Research Policy 26 (1), 1–18.

Landry, R., Amara, N., 1998. The impact of transaction costs on the institutional structuration of collaborative academic research. Research Policy 27 (9), 901–913.

Levin, S., Stephan, P., 1991. Research productivity over the life cycle: evidence for academic scientists. American Economic Review 81 (1), 114–132.

Lindsey, D., 1980. Production and citation measures in the sociology of science: the problem of multiple authorship. Social Studies of Sciences 10 (2), 145–162.

Liu, N.C., Cheng, Y., Liu, L., 2005. Academic ranking of world universities using scientometrics—a comment to the “Fatal Attraction”. Scientometrics 64 (1), 101–109.

Luukkonen, T., Tijssen, R.J.W., Persson, O., Sivertsen, G., 1993. The measurement of international scientific collaboration. Scientometrics 28 (1), 15–26.

McFadyen, M.A., Cannella, A.A., 2004. Social capital and knowledge creation: diminishing returns of the number and strength of exchange relationships. Academy of Management Journal 47 (5), 735–746.

Melin, G., Persson, O., 1996. Studying research collaboration using co-authorships. Scientometrics 36 (3), 363–377.

Meyer, M., 2006. Are patenting scientists the better scholars? An exploratory comparison of inventor-authors with their non-inventing peers in nano-science and technology. Research Policy 35 (10), 1646–1662.

Narin, F., Stevens, K., Whitlow, E.S., 1991. Scientific cooperation in Europe and the citation of multinationally authored papers. Scientometrics 21 (3), 313–323.

Noyons, E.C.M., Moed, H.F., Luwel, M., 1999. Combining mapping and citation analysis for evaluative bibliometric purposes: a bibliometric study. Journal of the American Society for Information Science 50 (2), 115–131.

Price, D.J.D., Beaver, D.D., 1966. Collaboration in an invisible college. American Psychologist 21 (11), 1011–1918.

Rigby, J., Edler, J., 2005. Peering inside research networks: some observations on the effect of the intensity of collaboration on the variability of research quality. Research Policy 34 (6), 784–794.

Schumpeter, J.A., 1934. The Theory of Economic Development. Harvard University Press, Cambridge, MA.

Schmoch, U., Schubert, T., 2008. Are international co-publications an indicator for quality of scientific research? Scientometrics 74 (3), 361–377.

Singh, J., 2007. External Collaboration, Social Networks and Knowledge Creation: Evidence from Scientific Publications. INSEAD, Mimeo.

von Hippel, E., 1994. “Sticky information” and the locus of problem solving: implications for innovation. Management Science 40 (4), 429–439.

Wagner, C.S., Leydesdorff, L., 2005. Network structure, self-organization, and the growth of international collaboration in science. Research Policy 34 (10), 1608–1618.

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. Science 316 (5827), 1036–1039.