

The organization and role of international collaboration in research production

A thesis submitted to the University of Arizona College of Medicine -- Phoenix
in partial fulfillment of the requirements for the Degree of Doctor of Medicine

David Hsieh
Class of 2013

Mentor: Kerr Whitfield, PhD

Abstract

The prevalence of multi-national and cross-disciplinary collaborative in the production of knowledge defines modern science as a social enterprise that extends beyond political, social, and geographic boundaries. The purpose of this study was to assess global trends in the composition and impact of multinational research teams. By examining the bibliometric data of 3.7 million primary research articles published from 1975 to 2005, it was ascertained that the frequency and scale of international collaborations has increased globally. Of note, the publications of many countries associated with lower research output were more often consistently affiliated with other nations across the time frame studied. By analyzing the number of times a publication is cited, it was discovered that multinational research studies have a greater research impact than research without an international presence, although the number of affiliated nations does not strictly correlate with citations. Taken together, this study provides insight into the dynamics of research teams which may better inform us how scientific partnerships between countries may be fostered and which collaborations may be advantageous.

Table of contents

Introduction	1
Materials and methods	4
Results	5
Discussion	18
Future directions	21
Conclusions	22
References	23

List of figures

Figure 1. The rise of International collaboration and authorship.

Figure 2. Quantifying the change in the composition of national affiliations.

Figure 3. Heat map representation of collaborative research output by country.

Figure 4. Heat map representation of normalized collaborative research output by country.

Figure 5. Heat map representation of the fractional collaborative research output by country.

Figure 6. Correlation between citation count and the number of national affiliations of a publication.

Figure 7. Citation advantage and disadvantage of multinational collaborations.

Introduction

Scientific discoveries are traditionally recognized as the feat of individuals or small teams. For example, internationally renowned science prizes such as the Nobel and Lasker awards are bestowed to no more than three persons per achievement under conventional provisions. Archival evidence also supports the dominant role of individuals in early scientific research, given that nearly all scientific papers published before the 1920s had no more than one author (1-3). However, since the second half of the 19th century, the number of one-author manuscripts has declined exponentially (3, 4). Beginning in the 1950s, multi-author and multi-institute research papers have emerged as the primary products of scientific research in both the natural and social sciences with hundreds of papers currently published each year having more than a hundred authors and even a handful with more than a thousand authors. This paradigm shift in scientific publishing suggests that scientific research has evolved from a singular enterprise into an expanding social endeavor (5).

If contemporary scientific research is defined as the cumulative effort of the scientific community, it can be inferred that all scientists must lie within a global social network which allows them to integrate their own research into the general body of knowledge. Derek de Solla Price first applied the term *invisible college* to describe the informal exchange of information and personal interactions among scientists that enable them to directly or indirectly affect one another's work (6, 7). This system of communication was proposed to underlie formal ties and collaborations that could be discerned through co-authorships in published works. As previously mentioned, multiple studies have demonstrated an increasing number of total authors and thus, collaborations in science publications. This trend is perhaps anecdotally unsurprising given that research collaboration has been widely upheld as an advantageous or essential process with few detriments in the advancement of scientific knowledge (1, 5, 8-10). Indeed, much evidence supports a positive impact of collaboration in science: increased levels of research collaboration between individuals, disciplines or institutes are correlated with greater manuscript quality, scientific output, citation number, and rates of manuscript acceptance (2, 5, 11-14).

Several caveats exist in these assertions which undermine a straightforward interpretation of the effect and relevance of collaboration in the scientific process. First, collaborations are associated with higher citations when collaborations are with higher tier research institutes (ranked by the number of citations), but less or oppositely so with lower tier institutes (5). In addition, among ecology publications, international collaborations are more common in European countries than in the United States, but lead to fewer citations (15). Second, when citations are divided by the number of co-authors in a research paper (fractional count), collaboration is not significantly correlated to publishing productivity (16). The average number of science papers per scientist across the natural sciences has not increased in the last few decades, suggesting that science output is a function of total authors rather than collaboration (17). Third, while manuscript acceptance is associated with author number, an affiliation with English speaking or economically developed countries is also correlated with more favorable scientific review (18). Collectively, when confounding factors such as geography or the fractional count are accounted for, it becomes less clear to what extent collaboration positively or negatively impacts science, possibly due to the inequitable and unpredictable nature of collaborations.

Further studies are warranted to define strict conditions and contexts where a collective effort is beneficial, efficient, and synergistic rather than additive in impact. Such findings would be crucial because the risks of collaborations, including opportunity costs (the value of an alternative investment that was forgone), diminishing marginal returns and increasing management and administrative expenses, have not been well elucidated and are likely not to be appreciated, even though they may be significant. For example, incentivizing international collaboration may impede internal development by discouraging domestic partnerships and promoting specialization over diversification of research in effort to be appealing affiliates. A corollary of such policies may be the formation of exclusive cliques, which is supported by an analysis of normalized co-authorship network structures of collaborative publications showing that most multi-country partnerships are established within a 14-member core group that has shrunk during the period from 1990 to 2005 (19).

Recent grant initiatives, including the United States National Institute of Health's Clinical and Translational Science Award program, European Union's Seventh Framework Programme, and Germany's Deutsche Forschungsgemeinschaft Mercator Programme, have been launched worldwide to promote research collaboration between countries. Such funding schemes support joint efforts between labs, nations, and academia/industry, and demonstrate the perceived worth of international research teams. However, it is not clear whether political and financial policies are effective in guiding research strategies or advancing scientific knowledge. In fact, the number of new molecular targets and drug approvals in the past decades has dramatically fallen behind the rate of spending in translational research, demonstrating that common assumptions about investments or policies needed to promote scientific progress may not be valid under certain circumstances (20, 21). Nonetheless, given the importance of research collaboration in the agenda of scientists and governments globally, the purpose of this study is to define networking trends within authorship structures of primary research articles with a focus on country-level relationships. I hypothesize that collaborations between countries have increased globally over time in part due to the increasing scale of research studies being conducted that require a wide scope of expertise in different subjects and the development of technologies that allow rapid communication worldwide. The citation impact of international collaborations over time is also investigated. By elucidating how international collaboration in research and its impact has changed over the last few decades, insights may be gained into the evolving behavior and utility of scientific international relationships.

Materials and methods

Data collection

Bibliometric records of primary research articles published from 1975 to 2005 were downloaded from the Thomson Reuters Web of Science (WOS) database between March 2012 and December 2012. To query primary research articles, WOS publication results were refined using Boolean operators to include only English language entries coded as article document types without anonymous authors. Publications coded in the WOS as article document types are distinguished from non-primary research articles if they include more than 100 references and contain the words “review” or “overview” within the title. Records with missing or unreliable data in any record field were excluded from this study.

To extract the country of origin, unique text strings corresponding to a country’s name were ascertained within the author address and reprint address field tags to ensure that only one country was recognized from each address. Data from countries which have changed their names or composition since 1970, such as the USSR, Yugoslavia, Germany, and Czechoslovakia, were incorporated into the publication counts of modern sovereign states which are politically and geographically representative of the previous entity or succeeded the previous state identity.

Statistical analysis

The Student’s t-test was used to determine a significant difference between mean citation counts between publications published in 1975 and 2005.

Results

By examining 3,730,067 primary research articles published across 7 years from 1975-2005, it was determined that the number of publications with authors associated with two or more countries has increased rapidly especially between 1990 and 2000 (Fig. 1). This may be explained by the fact that the number of authors in each publication has been increasing over time as seen in Figure 1, which may increase the coincidence of additional country affiliations. However, it is apparent that the rate of increase in the mean number of authors per publication is subtle and is therefore not likely to contribute significantly to the surge in international collaborations.

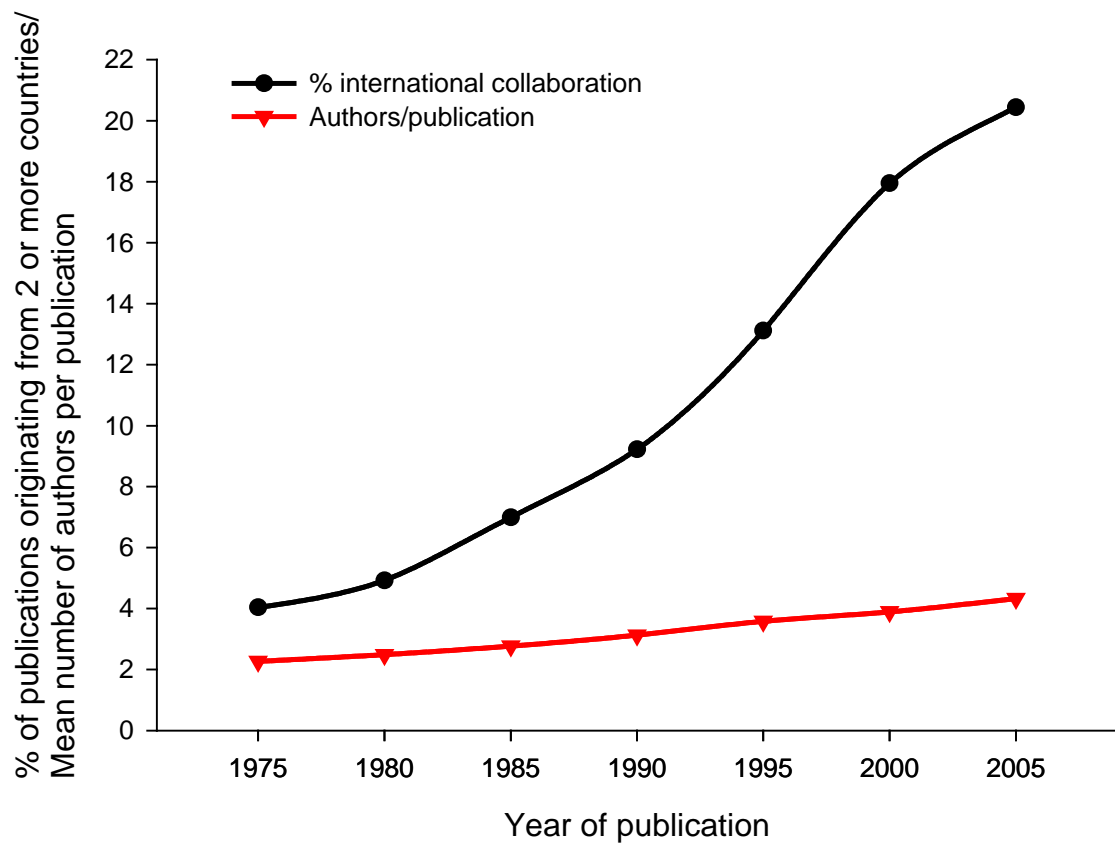


Figure 1. The rise of international collaboration and authorship. The number of publications affiliated with at least two countries was divided by the total number of publications published each year and plotted over time. The number of authors was similarly divided by the total number of publications each year and plotted over time.

Next, the scale of multinational collaboratives was assessed by classifying publications according to the number of nationalities represented. As seen in Figure 2, the number of publications composed of 2, 3-5, and 6 or more different national affiliations has increased from 1975 to 2005 nearly 4, 16, and 24 fold respectively. Interestingly, the rate of increase in the number of publications with greater than 5 national affiliations is greater than that of publications with 3-5 national affiliations which in turn is greater than that of publications with 2 affiliated countries (Fig. 2). Taken together, these results indicate that the rate of increase in paired relationships in international research is diminishing over time due to the nearly exponential rise of large scale collaboratives. However, it should be pointed out that the number of publications with 2 national affiliates is still the largest category, suggesting that the barriers to collaboration may be lowest when it is with only one other country.

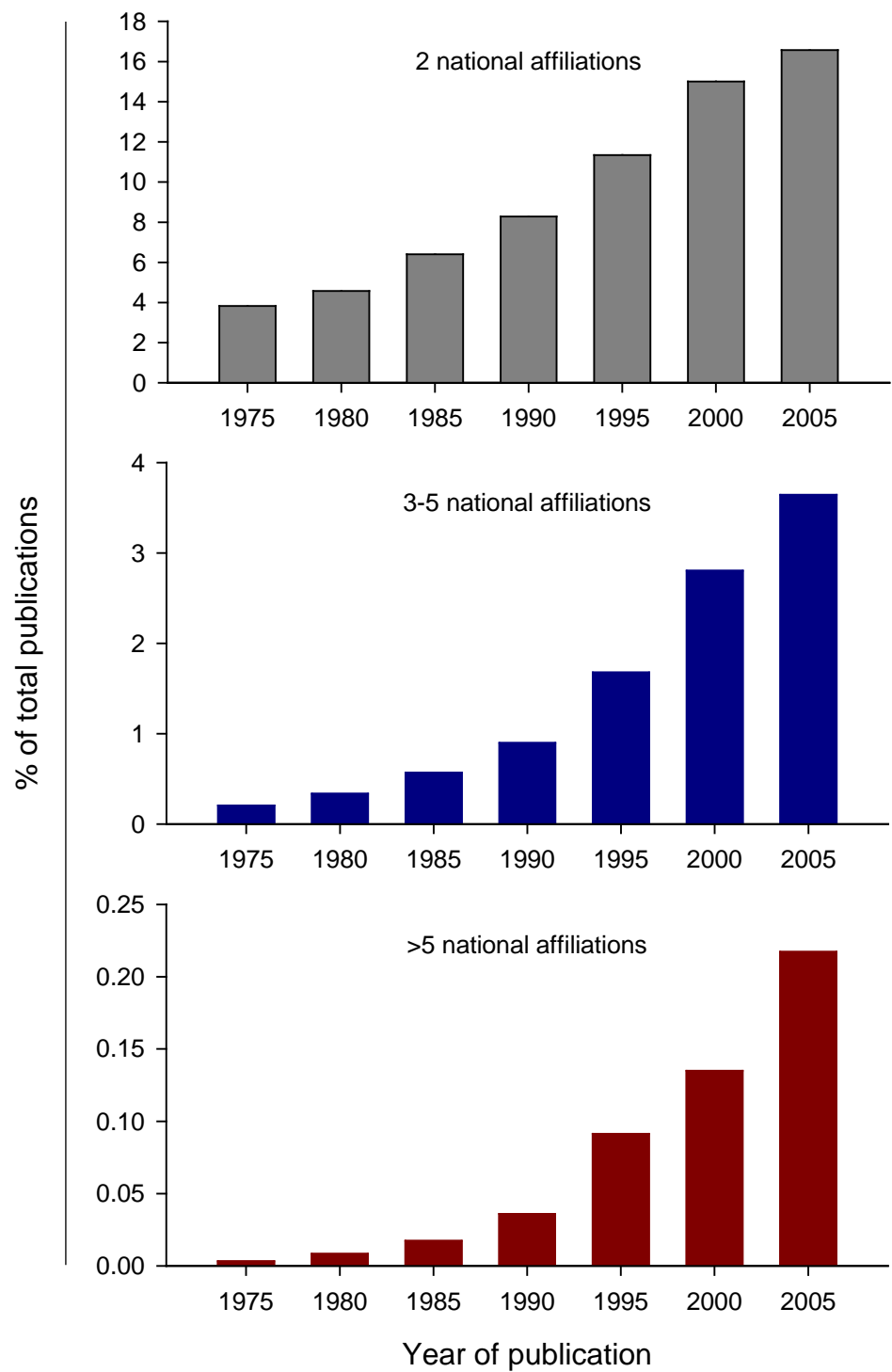


Figure 2. Quantifying the change in the composition of national affiliations. Publications were stratified into three groups based on the number of associated countries and their sum was plotted over time.

To examine individual collaborative behaviors, the number of publications with multiple national affiliations was quantified for each country and portrayed as a heat map (Fig. 3). Because the same scaling of color intensity must be applied to all countries, it becomes difficult to grossly assess collaborative trends due to the majority of research production being concentrated in a small group of states. However, by setting the maximal value for scaling at 1000 publications, it becomes apparent that only a few countries consisting of many developed states, including the US, Canada, UK, Switzerland, France and Italy, contribute substantially to the total number of collaborative papers. By normalizing data within each country (row), it is possible to discern how individual collaborative behaviors have changed over time (Fig. 4). Consistent with a global increase in research connectedness, Figure 4 demonstrates how nearly all countries uniformly show increased rates of collaboration over time. Certain exceptions involve countries with a small research output and thus their data may not be appropriately generalized in these results.

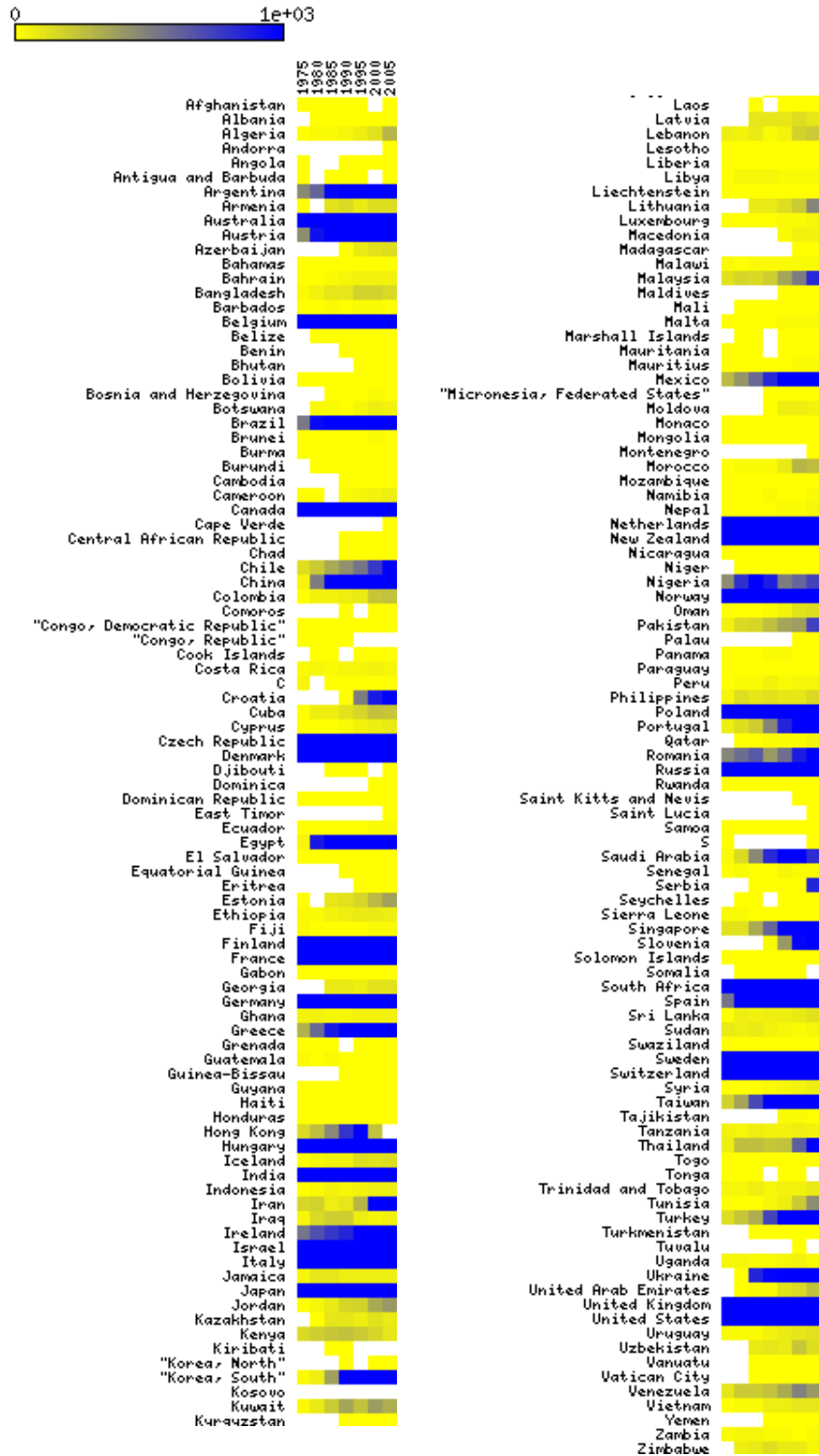


Figure 3. Heat map representation of collaborative research output by country. The number of publications with two or more national affiliations was quantitated by country. Color intensity was based on a 0-1000 scale. White boxes represent a lack of publications for that year.



Developed countries have a critical advantage in science because they are able to commit large investments into research and develop the infrastructure necessary for specialized studies. This may translate into greater collaborative opportunities for developed countries because they are the most attractive candidates for joint efforts since they are more likely to provide indispensable resources and experience. However, when the fractional collaborative research output was determined for each country, the greatest percentage of multinational efforts is seen among developing countries and nations with relatively lower research output, such as Belize, Chad, Congo, Ecuador, Gabon, and Mongolia (Fig. 5). This finding was consistently observed throughout the years analyzed in this study which suggests that while most countries are becoming increasingly linked through research over time, developing countries are the most reliant on collaborations for knowledge production.



Figure 5. Heat map representation of the fractional collaborative research output by country. For each country, the number of publications with two or more national affiliations divided by the total number of publications. White boxes represent a lack of publications for that year.

The relative impact of international collaboration in research over the span of the present study was determined by quantitating citations in 1975 and 2005. As seen in Figure 6, there is a correlation between the citation impact of publications and the number of international affiliations. Interestingly, there was a significant citation advantage ($p < 0.001$) of papers with one or two national affiliates published in 1975. However, papers with three national associations had a dramatic citation advantage ($p < 0.001$) if they were published in 2005 as opposed to 1975 (Fig. 6). The mechanism behind this phenomenon is unclear, although it may be influenced by temporal changes in the nature and efficiency of joint research ventures.

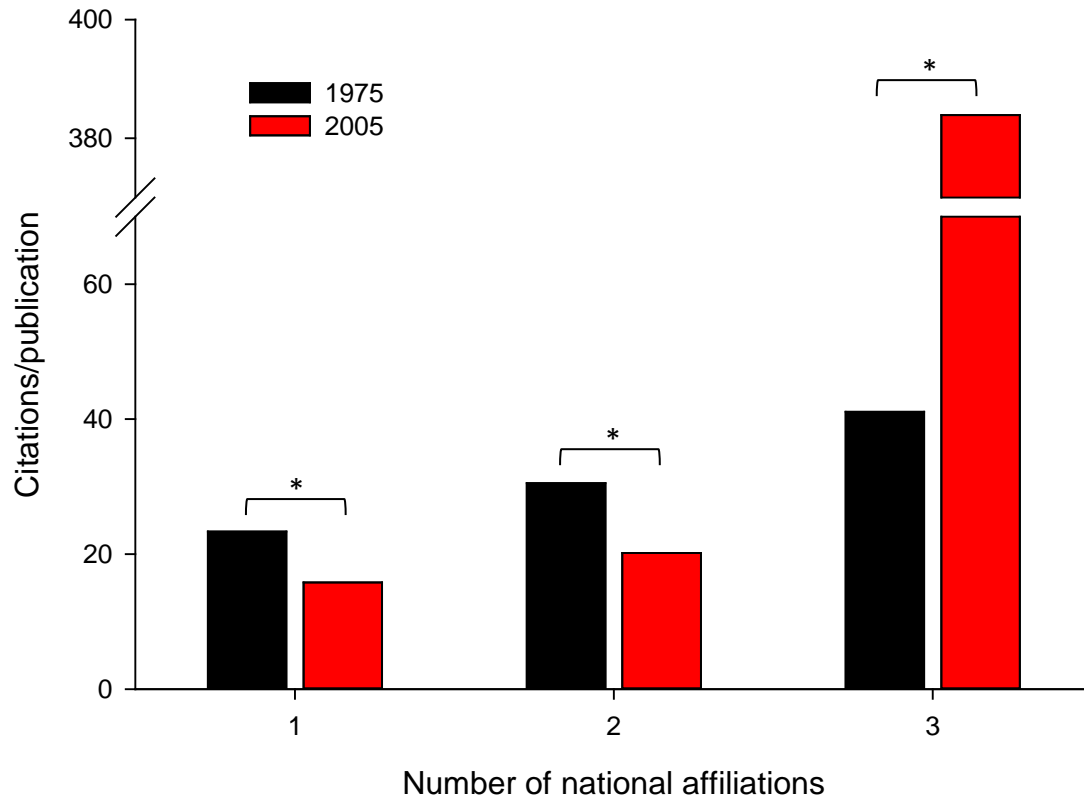


Figure 6. Correlation between citation count and the number of national affiliations of a publication. Citations of publications published in 1975 and 2005 were quantitated and compared using the Student's t-test. Asterisks denote $p < 0.001$.

Anecdotally, it seems logical that collaboration may foster creativity, increase utilization of resources, and thus yield high-impact research studies, but this concept has not been wholly scrutinized. As previously noted in Figure 6, the number of national affiliates appears to be a conserved positive regulator of research impact. However, when the mean citation of publications in 1975 with more than three national affiliates is assessed, the positive correlation between citation count and the scale of international collaboration appears tenuous (Fig. 7). Similar findings are observed with publications from other years which may indicate a diseconomy of scale when collaborations may become too large and thus inefficient. Collectively, these results imply that scientific collaboration may benefit research impact, but there may be a threshold after which adding more international partners may create diminishing returns.

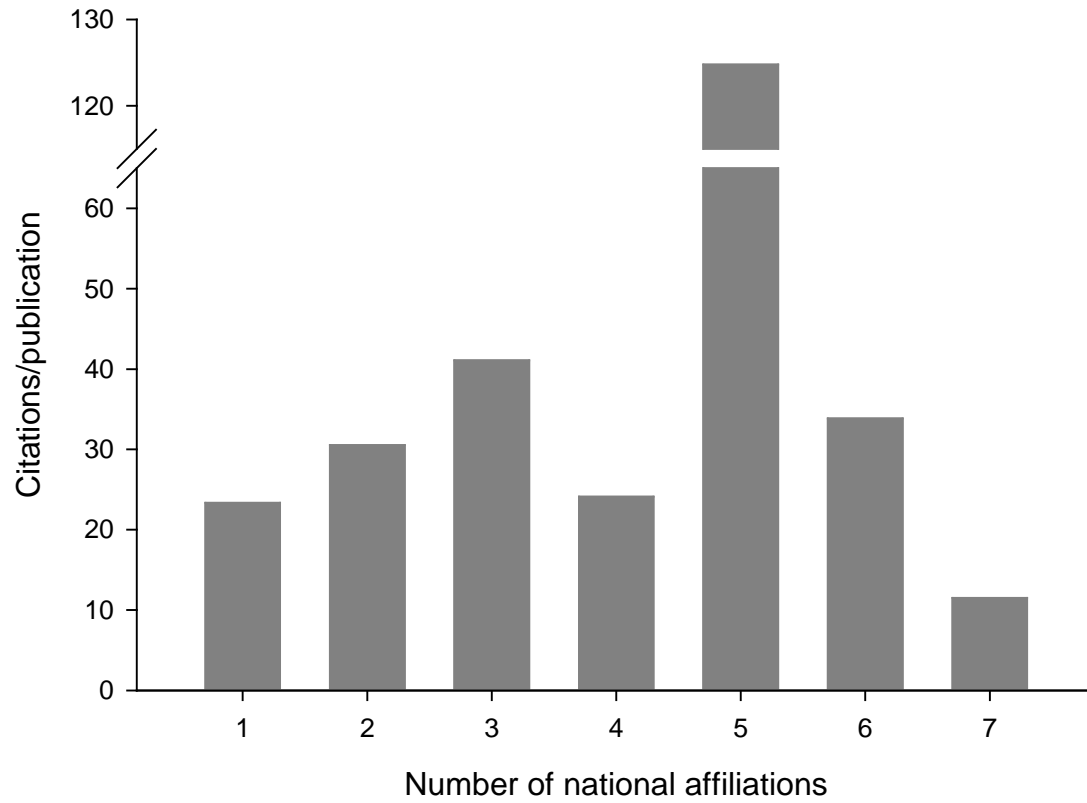


Figure 7. Citation advantage and disadvantage of multinational collaborations. Publications from 1975 were segregated by the number of national affiliations and their citations averaged.

Discussion

In this study it was demonstrated that in the last four decades, most countries became more intimately linked through scientific coauthorships. In addition to the increasing frequency of international collaborations, it was observed that large scale relationships involving three or more countries were the fastest growing collaborative structures. This may indicate that collaborative ventures are currently more accessible and feasible given the substantial technological improvements made since the 1970s. Of particular importance, it was discovered that collaborative publications constitute a considerable portion of the total research output of developing countries rather than that of developed nations. This demonstrates how collaboration may be a key resource for knowledge production to allow certain countries to serve as critical nodes linking other countries in scientific research. It would be of great interest for future investigations to assess the dyadic partners of collaborations for each country in order to clarify how and why scientists in developing nations are becoming intricately connected to their international peers.

This study provides evidence that is contrary to the widely held assumption that collaborative research efforts lead to higher impact publications. Although it was indeed observed that publications with up to 3 national affiliations receive more citations, the citation benefit did not appear to be consistently additive in larger collaborative structures. It is important to note that an assumption of this study is that article citation and research impact increase in parallel. This has been the subject of much controversy, given that citations may be influenced by open access, the availability of raw research data, the country of origin of authors, and co-authorship networks (22-25). Further studies may be able to control for such confounding factors to accurately determine how citation measures or other metrics may be used to appraise the value of international collaborations in science.

Whether co-authorship is a suitable assessment of a collaborative relationship is contentious given the significant drawbacks in the use of authorship as a surrogate indicator of research efforts. First, authorship may not accurately portray qualitative or quantitative differences in the contribution of each author. Although it is often customary that the first (lead) author and last author are credited with making the greatest overall input, this is not an

absolute practice (26-29). Instead, author positions are often arranged in a manner dictated by personal preference, norms in the research specialty, or institutional policies (30, 31). Consequently, the weight or characteristics of an author's contribution or collaboration may not be determined by author position (i.e. an author placed in the second author position may or may not have been the second most important collaborator). Second, a universal standard for authorship criteria is lacking. The International Committee of Medical Journal Editors (ICMJE) has established uniform requirements for authorship to remedy this dilemma, but a large proportion of journals either do not have explicit authorship policies or do not use ICMJE guidelines. It is also impractical for journals to enforce authorship requirements due to a lack of transparency and consensus in authorship information (32, 33, 8, 34-38). Unfortunately, the increasing average number of total authors in science publications may actually be inflated by "honorary" authorships in many journals irrespective of impact factor or field of study. Considering all these factors, collaborations and co-authorships cannot be fully quantified by authorship lists. Nevertheless, co-authorship is a readily accessible and archived component of science publications that remains a popular and widely employed index to measure collaboration based on its enduring, quantifiable and objective utility in systematic analysis. In this study, it is presumed that authorship designates a collaborator worthy of acknowledgment who is jointly responsible for the integrity of the study without regard to the actual nature of the contribution (39).

This study provides a macro-level analysis of collaborative behaviors in science. However, the precise mechanism underlying the willingness of a country to adopt international relationships in its scientific endeavors is likely partly derived from a micro-level cost-benefit analysis. As reviewed by Bukvova (40), many studies indicate that collaboration is rewarding for scientists by allowing them access to other expertise, resources, and skills. Interestingly, in large research groups, publishing endurance over long periods of time is dependent on rapid turnover rates, but in small research groups, stable relationships are needed (41). This suggests that the active publishing lifespan of individual scientists is dependent on the dynamics of their collaboration with others. In addition, because collaborations are viewed by many as corresponding with scientific productivity, mutual relationships may be sought after as a means

of attaining funding, promotion, hiring, and reputation. Taken together, these results suggest that research collaboration have both micro- (scientist) and macro- (country) level consequences, although the scale dependent effects of collaboration may be difficult to distinguish and are not within the scope of this study.

Future directions

Authorship dynamics and team compositions vary between fields of research suggesting that collaborative behaviors may also be different among scientific disciplines. These disparities may be influenced by the type and quantity of resources needed to conduct investigations pertaining to a subject field. Thus, it may be instructive to assess how nationality and research disciplines may be linked to international collaborations and their outcomes. This may be crucial in determining how governing policies may impact creativity, efficiency, and productivity differentially among scientists of different backgrounds and interdisciplinary research.

One limitation of this study was that co-authorship positions were not taken into account, even though they may be important determinants in how research collaborations actually operate. Future studies elucidating how co-authorship structures are spatially organized and determining whether different weights may be attached to each author position may allow for the proper allocation of credit and thus permit an exact analysis of the nature of a collaborative effort.

Other forms of intellectual exchange including patents, book chapters, and industrial relationships may also contribute to knowledge production and their inclusion in a future analysis may provide a more comprehensive portrayal of the global scientific network.

Conclusions

This study demonstrates that multinational teams have an increasing role in the production of knowledge and are evolving into larger scale structures of three or more nationalities. It was also shown that developing countries are more often associated with international collaboratives when compared to developed countries. Lastly, the citation advantage of international collaborations diminishes with larger collaborations (three or more national affiliations) irrespective of the year of publication.

References

1. Whitfield, J., Collaboration: Group theory. *Nature* 455 (7214), 720-723 (2008).
2. Guimera, R., Uzzi, B., Spiro, J., & Amaral, L.A., Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308 (5722), 697-702 (2005).
3. Greene, M., The demise of the lone author. *Nature* 450 (7173), 1165 (2007).
4. Abt, H., The future of single-authored papers. *Scientometrics* 73 (3), 353-358 (2007).
5. Jones, B.F., Wuchty, S., & Uzzi, B., Multi-university research teams: shifting impact, geography, and stratification in science. *Science* 322 (5905), 1259-1262 (2008).
6. Price, D.J.D.S. *Little Science, Big Science*. Columbia Univ Press New York Amer Zool 15, 119- (Columbia University Press: 1963).
7. De Solla Price, D.J. *Little Science, Big Science... and Beyond*. (Columbia Univ. Press: 1963).
8. King, J.T., Jr., How many neurosurgeons does it take to write a research article? Authorship proliferation in neurosurgical research. *Neurosurgery* 47 (2), 435-440 (2000).
9. Wuchty, S., Jones, B.F., & Uzzi, B., The increasing dominance of teams in production of knowledge. *Science* 316 (5827), 1036-1039 (2007).
10. Weeks, W.B., Wallace, A.E., & Kimberly, B.C., Changes in authorship patterns in prestigious US medical journals. *Soc Sci Med* 59 (9), 1949-1954 (2004).
11. Katz, J.S. & Martin, B.R., What is research collaboration? *Research Policy* 26 (1), 1-18 (1997).
12. Gordon, M.D. A critical reassessment of inferred relations between multiple authorship, scientific collaboration, the production of papers and their acceptance for publication. *Scientometrics* 2, 193-201 (1980).
13. Figg, W.D. et al., Scientific collaboration results in higher citation rates of published articles. *Pharmacotherapy* 26 (6), 759-767 (2006).
14. Bridgstock, M., The quality of single and multiple authored papers; An unresolved problem. *Scientometrics* 21 (1), 37-48 (1991).
15. Leimu, R. & Koricheva, J. Does Scientific Collaboration Increase the Impact of Ecological Articles? *BioScience* 55, 438 (2005).
16. Lee, S. & Bozeman, B., The Impact of Research Collaboration on Scientific Productivity. *Social Studies of Science* 35 (5), 673-702 (2005).
17. Abt, H., The publication rate of scientific papers depends only on the number of scientists. *Scientometrics* 73 (3), 281-288 (2007).
18. Tregenza, T., Gender bias in the refereeing process? *Trends in Ecology & Evolution* 17 (8), 349-350 (2002).

19. Leydesdorff, L. & Wagner, C. International Collaboration in Science and the Formation of a Core Group. *Journal of Informetrics* 2, 317-325 (2009).
20. Dorsey, E.R. et al., Financing of U.S. Biomedical Research and New Drug Approvals across Therapeutic Areas. *PLoS One* 4 (9), e7015 (2009).
21. Moses, H., Dorsey, E.R., Matheson, D.H.M., & Thier, S.O., Financial Anatomy of Biomedical Research. *JAMA: The Journal of the American Medical Association* 294 (11), 1333-1342 (2005).
22. Eysenbach, G. Citation advantage of open access articles. *PLoS Biology* 4, e157 (2006).
23. Piwowar, H.A., Day, R.S. & Fridsma, D.B. Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2, 5 (2007).
24. Meneghini, R., Packer, A.L. & Nassi-Calò, L. Articles by Latin American Authors in Prestigious Journals Have Fewer Citations. *PLoS ONE* 3, 4 (2008).
25. Wallace, M.L., Larivière, V. & Gingras, Y. A small world of citations? The influence of collaboration networks on citation practices. *PLoS ONE* 7, e33339 (2011).
26. Lake, D., Who's on First? Listing Authors by Relative Contribution Trumps the Alphabet. *PS: Political Science & Politics* 43 (01), 43-47.
27. Tschardtke, T., Hochberg, M.E., Rand, T.A., Resh, V.H., & Krauss, J., Author Sequence and Credit for Contributions in Multiauthored Publications. *PLoS Biol* 5 (1), e18 (2007).
28. Hunt, R., Trying an authorship index. *Nature* 352, 187 (1991).
29. Laurance, W., Second thoughts on who goes where in author lists. *Nature* 442 (7098), 26-26 (2006).
30. Laband, D.N., Contribution, attribution and the allocation of intellectual property rights: economics versus agricultural economics. *Labour Economics* 9 (1), 125-131 (2002).
31. Weltzin, J.F., Belote, R.T., Williams, L.T., Keller, J.K. & Engel, E.C. Authorship in ecology: attribution, accountability, and responsibility. *Frontiers in Ecology and the Environment* 4, 435-441 (2006).
32. Flanagan, A. et al., Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. *JAMA* 280 (3), 222-224 (1998).
33. Mowatt, G. et al., Prevalence of honorary and ghost authorship in Cochrane reviews. *JAMA* 287 (21), 2769-2771 (2002).
34. Slone, R.M., Coauthors' contributions to major papers published in the *AJR*: frequency of undeserved coauthorship. *AJR Am J Roentgenol* 167 (3), 571-579 (1996).
35. Eisenberg, R.L., Ngo, L., Boiselle, P.M., & Bankier, A.A., Honorary authorship in radiologic research articles: assessment of frequency and associated factors. *Radiology* 259 (2), 479-486.

36. Shapiro, D.W., Wenger, N.S., & Shapiro, M.F., The Contributions of Authors to Multiauthored Biomedical Research Papers. *JAMA: The Journal of the American Medical Association* 271 (6), 438-442 (1994).
37. Ivaniš, A., Hren, D., Sambunjak, D., Marušić, M., & Marušić, A., Quantification of Authors' Contributions and Eligibility for Authorship: Randomized Study in a General Medical Journal. *Journal of General Internal Medicine* 23 (9), 1303-1310 (2008).
38. Hwang, S.S. et al., Researcher Contributions and Fulfillment of ICMJE Authorship Criteria: Analysis of Author Contribution Lists in Research Articles with Multiple Authors Published in *Radiology*¹. *Radiology* 226 (1), 16-23 (2003).
39. Berk, R., Irresponsible coauthorship. *Am. J. Roentgenol.* 152 (4), 719-720 (1989).
40. Bukvova, H., Studying Research Collaboration: A Literature Review. *Sprouts: Working Papers on Information Systems* 10 (3).
41. Palla, G., Barabasi, A.L., & Vicsek, T., Quantifying social group evolution. *Nature* 446 (7136), 664-667 (2007).