Temporary Colocation and Collaborative Discovery

Sen Chai
Harvard University
chais@nber.org

August 21, 2015

JOB MARKET PAPER

Draft: Please do not cite or circulate without author permission

Keywords: temporary colocation; conferences; collaboration; knowledge production and diffusion

Acknowledgements: I would like to thank the Chemical Heritage Foundation's Beckman Library for the travel grant to access the GRC's attendance data and the Research Program on the Economics of Knowledge Contribution and Distribution funded by the Sloan Foundation for supporting this work. I am especially indebted to Richard Freeman for his continual support and feedback throughout this project. I am appreciative of the comments from Sujin Jang, Matthew Lee, Anoop Menon, Matthew Marx, participants of the REER 2013 conference and 2014 HBS TOM Alumni Workshop as well as three anonymous conference reviewers for the Academy of Management 2014 Meeting. I am also grateful to Marissa Suchyta and Sifan Zhou for their research assistance. All errors remain my own.

Temporary Colocation and Collaborative Discovery

ABSTRACT

Our understanding of how collaborative relationships form remains relatively thin. I assess a specific vehicle that fosters the formation of collaborations by studying how temporarily colocating at conferences affects attendees' research trajectory. The lack of empirical evidence on the impact of conferences on participants has fueled a heated debate. On the one hand, researchers are advised to attend conferences to further their careers, but there are obvious trade-offs of diverted funding and potential productivity loss while away from the bench. I use difference-in-differences regressions on a sample of attendees from Gordon Research Conferences and most similar matched researchers, and several different cuts of the data to address endogeneity of better researchers selected to present, existing co-authors attending together and choosing to go to a conference. My results suggest that even after a transitory period being colocated, long-term collaborations between conference attendees increase with especially strong effects for those who have never published together beforehand. Conditional on collaborative ties forming, I find collaborative outputs between conference attendees draw more from the knowledge space of the conference and are also more highly cited. Conferences also enable attendees who have never been cited by other attendees to showcase their research as evidenced by increases in within-attendee citations. Given the cumulative nature of research, these findings imply that over time conferences can have a significant impact in steering the research path of attendees, from the works that they cite and build upon to the colleagues with whom they collaborate.

INTRODUCTION

Understanding knowledge production and diffusion continues to be a fervent area of research, as the ensuing scientific and technological advancements spur wealth creation and stimulate sustained economic growth (Mansfield, 1972, Rosenberg, 1974). A well-documented and widespread trend in the production of knowledge is the prevalence of teams, as patents and scientific papers in all domains have become increasingly collaborative (Wuchty et al., 2007, Adams, 1990) and more cited (Singh and Fleming, 2010). While the literature has extensively studied what drives this increase in collaboration (Jones, 2009) and how collaboration and different structures of teamwork affect the subsequent quality of knowledge created (Singh and Fleming, 2010, McFadyen and Cannella, 2004, Girotra et al., 2010), our understanding of how collaborative relationships form remains relatively thin. A potential cause is the over reliance on data from existing co-inventor and co-author collaborations without knowing how collaborators came together to conceive of a common idea.

Emerging literature in the area has shown that colocated researchers with neighboring laboratories are more likely to collaborate together (Catalini, 2012). However, close proximity is difficult and not always possible to attain. Some organizations and institutions have attempted to rearrange the physical layout and structure of their workspace, but such changes remain very expensive as it entails disruptions from moving and high setup costs. A more affordable and feasible alternative to permanently colocating is the organization of events designed to lead to collaboration and innovation, such as meetings or seminars. Boudreau et al. (2012) study such a transitory setup and find increased short-term collaborations on grant applications. However, it does not address sustained long-term collaborative effects, the quality and inventive direction of collaborative outputs as well as citation behavior. In addition, since the sample of researchers is all affiliated with the same institution with less than five miles of distance between one another, it is hard to gauge how much the close geographic proximity confounds their results. This work studies conferences where distant researchers come together for a transitory period. I further refine colocation into two categories – permanent vs. temporary colocation – and empirically show that temporary colocation has long-lasting impacts similar to permanent colocation. Specifically, I construct a novel dataset of conference attendance that enabled me to observe interactions between potential collaborators.

The first modern conferences evolved alongside scientific societies as early as the 1660s, and some of these original conferences still exist today (McKie, 1960). Despite a long historic tradition and over 25,000 yearly meetings organized within the United States alone, very little quantitative work has assessed the effect of temporarily colocating at conferences on participants' subsequent research trajectory. Lack of empirical evidence has fueled a heated debate on such meetings' impact on researchers. On the one hand, scientists are advised to attend conferences to further their academic careers, but at the same time there are obvious trade-offs of diverted funding and potential productivity loss while away from the bench. Moreover, opponents criticize the high cost of conference organization and argue that the marginal benefits of conferences, if any, could be surpassed if the organizers redistribute the funds spent on conferences to grant more awards to young researchers in the harsh scientific funding climate (Ioannidis, 2012). Critics also cite the environmental impact of conferences, as it is estimated that total attendees of a mid-sized conference utilize 10,000 metric tons of carbon from travel[1] (Green, 2008).

Notwithstanding the critiques, researchers surveyed by the Science Advisory Board report an upward trend in the attendance of conferences with an average of 3.7 conferences per researcher per year in 2008 – a 54% increase from 2002 (SAB, 2013). Thus determining a quantitative means to measure this impact is important for policymakers, researchers and managers especially in the current tight financial climate. Spending cuts on conferences and overall decrease in flexible grant funding from the US Federal government has led investigators to reallocate funding – such as that normally reserved for conference travel and registration – to essential laboratory costs. Anecdotal evidence indicates that researchers sometimes view conference attendance as nonessential compared to core operating and experimental costs of running a lab, partly because there is no clear indication of how funding conference attendance will affect them and are therefore more willing to cut it. The aim of this paper is to shed light on the issue and investigate how the attendance of a transitory meeting affects attendees' research trajectory.

Without a randomized experiment, the key empirical challenge in causally linking conference attendance to subsequent collaborative and citation behavior is the endogeneity of attending. I mitigate this identification concern as follows. Through hand-collection and digitization of over one thousand attendees from fifteen Gordon Research Conferences – a series of esteemed international conferences in the natural sciences – I carefully construct a sample of qualitatively similar researchers who did not attend – matched on observable dimensions such as research focus, prior productivity, collaborations, citations and experience – to address the selection bias that submissions by superior researchers are more likely to be accepted and presented at conferences. Using difference-in-differences models, I compare the attendee and matched samples before and after a conference event. Although this empirical methodology contrasts attendees and their most analogous peers, it suffers from the endogeneity of existing coauthors attending the same conference together to present a collective work. To solve this issue, I differentiate between attendees who have no prior collaborations with any other attendee of the conference and those who have, and focus on collaborations between new collaborators. However, even with this first cut of the sample,

[1] A passenger vehicle that averages 12,000 miles per year emits approximately 4 metric tons of carbon.

the decision to attend a conference is still endogenous. To address this concern, I separate attendees most similar to the average characteristics of conference attendees from those who are different, with the motive that attendance by more similar participants is more likely to be random.

My results suggest a substitution effect for collaborators who met at conferences: collaborations amongst conference attendees increase significantly more compared to collaborations between attendees and non-attendees, while the overall number of collaborations after the conference only rises slightly for attendees compared to non-attendees. I find especially strong results for attendees who have never collaborated with any other attendee from the conference. Conditional on a collaborative tie forming, collaborations between attendees are more highly cited and draw more from the knowledge space of the conference compared to collaborations between attendees and non-attendees. Exploring forward citations, I observe strong between-attendee citation effects for those who have never been cited by other attendees prior to the conference. Studying the differential benefits on junior versus senior attendees, I find greater collaboration and citation effects for junior attendees. Finally, I observe that the effects for non-presenters are stronger than for presenters.

Given path dependence of research, these results imply that over time the cumulative effect of conferences can have a significant impact in steering the research path of attendees, from the works that they cite and build upon to the colleagues they choose to collaborate with. These results also suggest that even when researchers colocate temporarily for a short period of time, they can reap similar collaborative benefits of permanent colocation.

The structure of this work is as follows. I begin by developing the theoretical basis that guides my empirical predictions. I then describe the setting from where I compiled my data, detail the empirical strategy and estimation methodology employed to conduct my analyses. Finally, I elaborate on my quantitative results with qualitative interviews of researchers and explore the implications of my findings.

THEORETICAL FRAMEWORK

Conferences and Collaborations

Recent studies show a continuing and increasing trend for teams to contribute in the production of knowledge in all scientific and technological domains (Wuchty et al., 2007, Adams, 1990). Individual inventors and researchers are burdened by the ever expanding body of knowledge, and have to narrow their expertise and work in teams in order to contribute at the frontier of science (Jones, 2009). A rich body of work has debated the merits of working in teams versus working alone. Teamwork permits the recombination and generation of more diverse ideas, and better selecting out of bad ideas through extensive critiquing between collaborators, thereby decreasing the likelihood of poorer outcomes (Singh and Fleming, 2010). However, even though teams bring greater collective knowledge and effort, there remains significant costs to increased teamwork such as coordination losses (McFadyen and Cannella, 2004) and groupthink (Janis, 1971). A compromise between these opposing views shows that working alone during certain stages of a collaborative project benefits from minimized idea suppression and social loafing (Girotra et al., 2010). While these literatures extensively study collaborative effects of established teams, an emerging branch has started to examine the mechanism through which collaborations are formed (Boudreau et al., 2012, Catalini, 2012).

A potential collaboration forming mechanism can be conceptualized as a matching process. Researchers reveal information about themselves, such as their research focus and interest, in order to reduce the search cost of finding potential collaborators. The information is exchanged between potential collaborators through various interactions so as to reduce asymmetric information and aid the decision to form a collaborative bond (Fafchamps et al., 2010). This helps collaborators find a common topic of interest where the inventive direction of the collaborative output is determined through the recombination of ideas stemming from each collaborator's knowledge space (Henderson and Clark, 1990, Fleming, 2001). Permanent colocation supports this matching process, as increased convenient opportunities in meeting and interacting in common areas foster information to be revealed and exchanged over time, thereby increasing the potential for collaborative relationships to be formed. But the average collaboration quality of neighboring labs are lower than that over distance and have greater variance (Catalini, 2012). Moreover, permanent close proximity is difficult and not always possible to attain, and attempts to rearrange the physical layout and structure of workspaces remain expensive due to disruptions from moving and high setup costs.

As an alternative to being permanently colocated, conferences provide an environment of temporary colocation where more distant researchers who otherwise would not have had the chance to meet can interact together. Thus, it provides a physically and temporally condensed platform for researchers to showcase their work and easily reveal information through posters and presentations, and exchange information about the latest research ideas and results between colleagues during coffee breaks and networking events. These interactions facilitate rapport and network building amongst attendees, and in turn foster new collaborative opportunities.

Once researchers decide to establish a collaborative relationship, sustaining the existing collaboration requires that benefits outweigh the costs of coordination inherent in collaborative work. Again geographic proximity and colocation play an important role in reducing this cost as studies have shown that collaborators tend to be more proximate geographically (Katz, 1994), although with the introduction of inexpensive communication tools, such as the Internet and Skype, geography's role is becoming less imperative (Agrawal and Goldfarb, 2008) and collaborations are becoming more international (Freeman et al., 2013). Thus once partnership links are formed at conferences, collaborators can turn to low cost means of communication if they are more distant or face-to-face interactions if they are more proximate to sustain the relationship.

Using an experimental design, Boudreau et al. (2012) show a positive effect of colocation on the short-term outcome of jointly applying to grants. However, since the sample of researchers all come from the same institution with less than five miles of distance between one another it is hard to gauge how much the close geographic proximity confounds their results. Furthermore, no study has explored the effect of temporally colocating at conferences on long-term collaborative behavior. This paper sheds light on this relationship of how conferences act as a platform for the formation of *sustained* collaboration, where attendees are able to acquire valuable information about and exchange information with potential collaborators through face-to-face exchanges. From these theoretical expositions, I posit that collaborations amongst conference attendees are more prevalent compared to similar researchers who did not attend the conference.

Inventive Direction and Quality of Conference Collaborations

Conferences also serve as a tool to expand one's scope of awareness, enabling attendees not only to gather new ideas but also to identify links between seemingly disparate streams of work and observations from other potential collaborators. Prompted by presentations and posters around a common topical focus, collaborators will not only draw from knowledge in their own repertoire of expertise but also borrow ideas embedded in the conference, and recombine them in different contexts (Fleming, 2001, Henderson and Clark, 1990). Thus conditional on collaborations forming amongst attendees of the same conference, the inventive direction of the subsequent collaborative output is directed by complementary inputs from the collaborators as well as the knowledge space of the conference.

Conferences provide a dynamic forum for creative thinking and brainstorming. Informal and critical feedback from peers in similar fields of interest can often lead to a redesign of experiments, and improvements in the strength of the research ultimately published. This exposure may incite new ideas from collaborators, thereby increasing the combinatorial (Weitzman, 1998) likelihood of producing higher quality research, and perhaps even instigating a breakthrough. Thus, I postulate that collaborative ideas conceptualized during or shortly after a conference benefit from inputs of collective brainstorming and critical feedback, and in turn tend to be more highly cited than collaborations between attendees and non-attendees.

Conferences and Citations

Studies show that despite difficulty of knowledge transmission between boundaries, it can be enhanced through mobility of individuals (Almeida and Kogut, 1999) and geographic colocation (Zucker et al., 1998). The structure of conferences – where attendees come together temporarily from distant locations to exchange ideas and return with an updated set of knowledge to their home institutions – breaks down institutional boundaries and combines the two mechanisms of mobility and colocation that enhances knowledge spillover. The topical organization of modern conferences also fosters a sense of community that provides social proximity (Freeman et al., 2013), which may in turn increases citations. Thus, these meetings also act as a platform for the diffusion of knowledge where attendees are more likely to cite amongst themselves.

Junior vs. Senior Attendees

The effect of collaboration after attending a conference should differ depending on the career stage of the attending researcher. Since junior researchers are relatively young and unknown, conferences can showcase their research interest and ability to the community. It is also a vehicle for the junior attendee to see others' work and match with potential collaborators beyond their immediate geographical and social networks. Senior researchers, on the other hand, have already established themselves in the community and are known amongst peers, thus reveal less information compared to juniors. Moreover, junior researchers are less embedded in existing collaborative relationships, and are more prone to the influence of a shaping event like a conference. Thus, I posit that the effects of attending a conference and establishing collaborative relationships for junior researchers will be more pronounced. A similar argument based on information disclosure and exchange is valid for citations, where junior attendees will garner more subsequent citations than senior participants by presenting their research interests and providing a signal of quality.

Presenter vs. Non-Presenter Attendees

Along the same reasoning of revealing and exchanging more information, presenters of formal talks and discussants are given a structured chance to showcase themselves and their research to other conference attendees. It enables others to more easily strike up a conversation with the speaker after the presentation by referring to the presented work, and thereby improving the chances of establishing collaborative links through increased interactions. Therefore, I posit that the effects of attending a conference and establishing collaborative and citation relationships are stronger for presenters than non-presenters.

METHODOLOGY

Setting – Gordon Research Conference

This work studies the Gordon Research Conference (GRC), considered as one of two major conference series in the natural sciences with the mission of fostering communication at the frontiers of science. The first GRC convened in 1931 in chemistry. Today, there are over two hundred GRCs that specialize in specific aspects of the natural sciences. GRCs are relatively small and specialized conferences, organized in a decentralized manner, in the fields of biological, chemical, and physical sciences, and their related technologies. They are weeklong and held at remote sites where attendees are isolated from potentially distracting environments. All GRCs are also single-track conferences in that there are no parallel sessions. It has a focus of building community among attendees, substituting additional talks with small group discussions and informal afternoon activities[2]. This setup is not only typical of small academic conferences; it is also similar to many global summits that companies hold. It ensures that attendees meet face-to-face and interact with one another, which may not be the case in larger and shorter timeframe conferences. Participation in GRCs is only possible through successful application. Unlike bigger conferences where attendance is difficult to track, the setup of these GRCs entails that a list of actual attendance is available. These characteristics of the GRCs provide a suitable setting to investigate my research questions that link temporary colocation of conference participants to subsequent collaboration and citation behaviors.

Dataset and Sampling

To determine how conferences affect subsequent collaborative and citation behavior of attendees, two sets of data are required: first, the list of attendees to a set of conferences; and second, publication data on the scientific productivity of attendees. I hand-collected and digitized attendance lists for fifteen biological GRCs between 1992 and 1995 from the Chemical Heritage Foundation's Beckman Library in Philadelphia, PA[3] and manually matched a total of 1,265 attendees onto the Author-ity database (Torvik and Smalheiser, 2009), a disambiguated dataset of all Medline publications in the life sciences, to obtain

[2] Since its initiation, GRCs do not publish their own report in scientific journals on proceedings. Furthermore, attendees are not permitted to cite conference proceedings in scientific journals. The GRC administration believes that this restriction encourages attendees to present and discuss new, unpublished, and innovative information more freely. These policies also ensure that my dependent variables of publications, collaborations and citations are unbiased, as attending the conference does not systematically inflate these measures.

[3] The fifteen conferences include the Meiosis, Mitochondria & Chloroplasts, Molecular Cytogenetics and Neuroendocrinimmunology conferences in 1992; the Matrix Metalloproteinases, Neurotrophins, Wound Repair and Calcium Signaling conferences in 1993; the Hormonal & Neural Peptide Biosynthesis, Oxygen Binding Proteins and Cellular Basis of Salinity Tolerance in Plants in 1994; and finally, the Angiogenesis and Microcirculation, Cell Death, Epigenetics, Human Molecular Biology in 1995.

peer-reviewed publication data for each individual. Using this initial group of conference attendees, I derived a second group of most analogous non-attendee peers. To ensure that I am comparing similar individuals between attendees and non-attendees, and minimize the selection bias where submissions from highly productive and cited researchers are more likely to be accepted for conference presentations, I matched individuals based on five observable dimensions: field of study, prior publication history, prior collaborations and citations, as well as experience.

My matching strategy was performed as follows. I exact matched the top three Medical Subject Heading (MeSH) keywords[4] from attendees and non-attendees' five-year prior publications. For each attendee, I obtained a set of non-attendees with comparable research focus. If this set of potential matched non-attendees is greater than one, I refined the matching by finding nearest neighbors to each attendee based on years of experience since first publication, as well as the number of five-year prior publications, collaborations, and forward citations received. Weighing all four dimensions equally, I employed a vector space analysis where I kept the two nearest neighbors defined by having the shortest Euclidian distance to the attendee being matched. Through this process, I identified for each attendee one to two closest researchers who did not attend the conference. This sample of matched researchers comprises of 2016 individuals.

If the matched sample truly includes individuals most analogous to attendees, there should be no significant difference on the observable dimensions in the five years prior to the conference. Two-sided t-tests on these dimensions in Table 1 confirm that both groups were not significantly different. To rule out the possible interpretation that post conference results are due to the continuation of a prior tendency that started well before the conference rather than the effect of attending the conference itself, I ensured that the year-by-year trends of the attendee and matched groups in the pre-period are comparable as depicted in Figures 1A, 1B and 1C. Focusing on the left of the vertical line at time zero, overall publications, collaborations, and forward citations reassuringly do not follow any discernibly different trends between the two groups.

[Insert Table 1 about here] [Insert Figure 1 about here]

One additioal issue is that attending a conference together is not exogenous, especially as collaborators are present at the same conference once a common project is selected. Thus, I needed to ensure that more frequent collaborations and citations between attendees after the conference is not driven by existing collaborations and citations before attending the conference. To address this issue, I isolated

[4] Instead of being assigned by authors, MeSH is a comprehensive controlled vocabulary for the purpose of indexing journal articles and books in the life sciences, and also serves as a thesaurus that facilitates searching. It is created and updated by the United States National Library of Medicine (NLM) and used by the MEDLINE/PubMed article database and by NLM's catalog of book holdings

attendees who never collaborated with one another in the five years prior to the conference from those who did collaborate, and ran my analyses separately for the two groups. For those with no prior collaborations within attendees (*n=641*), I was able to identify the pure effect of conferences. For those who did coauthor with another attendee prior to the conference (*n=624*), I further separated the data into collaborations formed with new collaborators met at the conference, and collaborations between old collaborators from before the conference. Similarly for between-attendee citations, I separated the data into attendees not having (*n=649*) received any prior citations from within the conference and those who have (*n=616*).

While this first cut of the sample addresses the endogeneity problem of collaborators attending the same conference together, the decision to attend is still endogenous. I isolated attendees most similar to the average characteristics of all conference attendees from those who are dissimilar, with the motive that attendance by similar participants is more likely to be random. Less similar participants are most likely to attend with a particular motivation, while similar attendees' participation is more likely to be random as they can just as easily chose another conference to interact with their community. I defined the conference's average based on six dimensions derived from the five-year prior publications of all attendees: average number of individual's MeSH terms that matches the conference's top 10% most frequent MeSH terms, average number of publications, average number of collaborations, average number of collaborators, average number of forward citations and average experience at the time of the conference. Each attendee and matched researcher in my sample is similar to the conference for a particular dimension if the difference between their individual value and the conference average falls within the middle 50% of the sample distribution. Repeating the same exercise for all six dimensions of each individual, the dissimilar sample (*n=1204*) is defined as researchers with three similar dimensions or less, while the similar sample is defined as those with four similar dimensions or more (*n=2077*).

Finally, to test the differential benefits of attending conferences for junior versus senior attendees, I classify researchers with less than 10 years of experience since first publication at the time of the conference as junior (*n=1566*) and the remaining as senior attendees (*n=1715*). I use ten years as an approximation of the time to tenure from first publication, assuming researchers first publish in the late stages of their doctorate and take an average of seven to eight years to tenure once in a faculty position. I also use seven years and twelve years from first publication to tenure as cutoffs for junior attendees as robustness checks. Similarly, I classify researchers who gave a formal presentation or was a discussant as a presenter (*n=1417*), and the remaining as non-presenters (*n=1864*).

Variables

Except for variables measuring inventive direction and quality of collaborations at the publication level, the main data is at the individual-year level set up in panel form with annual observations for five years before (including the year of the conference) and after conference attendance. Table 2 shows summary statistics including the sample size, mean, standard deviation, minimum and maximum for each variable used in the analysis. For count variables, I took the natural logarithm plus one whenever they entered the regression on the right-hand side of QML Poisson and logistic regressions to match count or indicator explanatory variables that underwent the same transformation in those models.

[Insert Table 2 about here]

Experience

Experience is the number of years since a researcher's first publication until the year of the conference. I also included an indicator variable for junior researchers (*junior*) defined as one if the researcher had ten years or less of experience.

Presenter

I identified whether the researcher formally presented a talk or was a discussant using the indicator variable (*presenter*).

Distance

To control for the confound that collaborations between attendees who are permanently colocated are more easily sustained, I calculated a measure of distance in miles from the researcher's primary affiliation at the time of the conference to the meeting venue at the individual level (*distance to conference*), as well as the average distance at the time of the conference between collaborators on the same publication who attended the same meeting (*average collaborative distance*).

Publication

I measured the quantity of knowledge produced using the number of peer-reviewed publications (*# publications*).

Collaboration

I developed two variables to explore the collaborative effect of attending a conference. I first counted the number of papers where researchers collaborated with one or multiple co-authors (*# collaborations*). I then focused on the collaborations formed between attendees. This measure is defined as the number of collaborative links formed between participants who attended the same GRC for attendees. Since the counterfactual sample consists of would-be participants in the same scientific domain, I took the number of collaborative links formed between the would-be attendees and participants of the conference. Thus (#

collaborations within attendees) has the number of attended-attended collaborations for the attendee group and attended-matched collaborations for the counterfactual group.

Inventive direction and quality of collaborative output

Conditional on collaborations forming between attendees after the conference, I measured how similar the collaborative output is to the authors' field of expertise prior to the conference. I used MeSH keywords as a proxy for summarizing the content of the collaborative output. Comparing MeSH keywords of the collaborative publication to the MeSH keywords of each collaborator in the dyadic relationship, I derived the fraction of MeSH keywords of the collaborative output that came from one or the other author (*MeSH fraction from one or other*) and both authors (*MeSH fraction from both*). I also defined the core knowledge space of the conference to be the aggregate knowledge of all attendees as proxied by their 5-year prior publications' MeSH terms. This allowed me to quantify the extent to which collaborative works draw from knowledge embedded in the conference itself, where I counted the number of MeSH keywords collaborations between conference attendees have in common with the conference's top ten percent most frequent MeSH keywords (*# MeSH in common with conference*). Moreover, I counted the number of forward citations each collaborative paper garnered in the ten-year period after its publication (*# citations for within-attendee collaborations*) to measure its quality. To gain better insight on the distribution of these citations, I used an indicator variable to identify collaborative outputs on the left tail with zero citations (*zero citation indicator*) and those on the right tail with citations in the top 90^{th} percentile of the citation distribution (*top90th citation indicator*).

Citations

Similar to collaboration measures I used two variables to capture the knowledge spillover trends of attending a conference. I counted the total number of forward citations *(# citations*) all papers of each researcher garnered in the five-year period after its publication. I also focused on the citations formed between attendees. This measure is the number of attended-attended citations for the attendee group and attended-matched citations for the counterfactual group received from another participant (*# citations within attendees*).

Empirical Approach and Regression Model Estimation

The empirical analysis employs three models. In the case of non-zero values before and after the conference, I employed a difference-in-differences (DiD) panel regression model to estimate the relationship between conference attendance and subsequent collaboration and diffusion. I identified the effect of attending a conference by measuring the difference between the average gain of attendees and the average gain of non-attendees. It removes potential biases in the post period comparisons between attended and non-attended researchers due to their permanent differences, and biases from comparisons over time between the pre and post periods from trends in the attended group.

$$Y_{i,s,t} = \alpha + \gamma\ attended_s + \lambda\ post_t + \beta(attended_s \cdot post_t) + \delta X_{i,t_0} + \varepsilon_{i,s,t}$$

The outcome variable is $Y_{i,s,t}$ for researcher *i* at time *t* for attended state *s*. Conference attendees are exposed to the treatment of attending a conference in the post period, while the group of non-attendees is not exposed to any treatment in the pre or post periods. *Attended* is the indicator of whether individual *i* has attended a conference at time t_0, and γ is the difference between attendees and non-attendees. *Post* is the indicator of the period after conference, and λ is the difference between the pre and post conference periods irrespective of the group. The DiD is captured by the interaction effect of $attended_s$ and $post_t$, where β is the coefficient of interest. For each individual *i* in the vector X_{i,t_0}, I included covariates for observables such as conference fixed effects as well as individual characteristics like experience, number of publications, collaborations, citations, and distance from primary affiliation to the conference venue. If the before values are zero for both groups such as in the cases where attendees have never collaborated or been cited within the conference, the DiD model simplifies to the following panel setup in the post period, where γ, the coefficient for the *attended* indicator, becomes that of interest:

$$Y_{i,s,t} = \alpha + \gamma\ attended_s + \delta X_{i,t_0} + \varepsilon_{i,s,t}$$

Since most outcome variables are non-negative and over-dispersed counts, I used quasi-maximum likelihood (QML) Poisson models with random effect robust standard errors clustered at the individual level instead of simple Poisson models to circumvent the assumption of equal mean and variance distribution to minimize estimation bias.

Conditional on collaborative ties forming between conference attendees, the level of analysis is at the publication level. The estimation model where γ, the coefficient for the *attended* indicator, is still the coefficient of interest is depicted as follows:

$$Y_{j,s} = \alpha + \gamma\ attended_s + \delta X_{j,t_0} + \varepsilon_{j,s}$$

for publication *j* and attended state *s*. QML Poisson, OLS and logistic regressions with robust standard errors are used when dependent variables are expressed respectively as count, percentage or indicator variables.

RESULTS

Table 3 summarizes my results of collaborations and citations within attendees compared to collaborations and citations between attendees and matched non-attendees, and the various samples used in my analysis. Below are detailed interpretations of these results.

[Insert Table 3 about here]

Conferences and Collaborations

Although temporary colocation at conferences should theoretically enhance the mechanisms of information revelation and exchange conducive to collaborations, it is important to document whether collaborations do form. But before delving into between-attendee collaborations, I first establish a baseline for the change in overall publications and collaborations of researchers in my sample. The first two regression models in Table 4A respectively demonstrate a modest significant increase in productivity of 3.2%[5] and in overall collaborations of 1.3% for attendees than non-attendees after the conference for the full sample. If I divide the sample to the dissimilar and similar sets of attendees, I find similar small effects on overall publications and collaborations as shown in Tables 4B and 4C. These effects on productivity however cannot be fully attributed to conference attendance alone as researchers who participate may tend to do so more when they have a late-stage – and most likely collaborative – work in the pipeline expected to be published shortly after the conference.

While attending a conference does not have strong effects on the overall productivity and collaborative behavior of attendees, it has a pronounced effect on whom the attendee chooses as collaborators. Model 3 in Table 4A depicts a 40.8% increase in between-attendee collaborations versus attended-matched collaborations in the full sample. Figure 2A graphically illustrates this outcome for the full sample. Comparing results from Models 2 and 3 in Table 4A, attendees substitute other collaborators with co-authors that they meet at conferences since their overall collaborations (3.2% increase) change much less than between-attendee collaborations after a conference (40.8% increase). Further decomposing the sample into dissimilar and similar attendees in Tables 4B and 4C, I find that the substitution effect is stronger for the similar sample than the dissimilar sample since attendees in the former are more likely to find other attendees with common research interests.

[Insert Figure 2 about here] [Insert Tables 4A, B & C about here]

However, the trend in Figure 2A and the significant strong effect for *attended* in Model 3 of Tables 4A, 4B and 4C indicate that attending together is endogenous as attendees collaborate more and

[5] QML coefficients interpreted in percentage as $e^{\text{coefficient of interest}} - 1 = e^{0.0307} - 1 = 3.2\%$ increase

more with one another prior to the conference. To tease out the confounding effect of existing collaborators attending conferences together, I further decompose the sample into researchers who had prior collaborations within the conference and those who have never collaborated with anyone attending the conference. For the subset of attendees without prior collaborations within the conference, I observe in Model 1 of Tables 5A, 5B and 5C very strong and significant effects that attending the conference increases collaborations with a fellow attendee by 10.0[6] to 14.5 times for the three samples. Figure 2B graphically depicts this result for the full sample. Interestingly, the effect is stronger for dissimilar attendees as they are less likely to have had prior collaborations with other attendees from the conference and therefore the change in within-attendee collaborations is higher after the conference.

For conference attendees with existing prior collaborations within the conference, I further divide the data into post conference collaborations between new and old collaborators. Conferences significantly increase attendee collaborations with new collaborators by 2.5 times versus an increase of 32.4% with old collaborators, as shown in Models 2 and 3 of Table 5A as well as graphically in Figure 2C and 2D for the full sample. Again similar results are found for the dissimilar and similar samples in Tables 5B and 5C. Similar attendees have stronger effects than dissimilar attendees for those with existing prior collaborations within the conference as common interests between similar attendees facilitate the likelihood to forge new collaborative relationships.

Decomposing Figure 2A that depicts overall collaborations into Figures 2B and 2C that show collaborations between new collaborators and 2D illustrating collaborations between old collaborators, the endogenous trends first observed on the effect of between-attendee collaboration in Figure 2A are driven mainly and unsurprisingly by collaborations between old collaborators in Figure 2D. Taken altogether, these results suggest strong evidence that temporarily colocating at conferences foster the formation of new collaborative links between attendees.

[Insert Tables 5A, B & C about here]

To ensure that being permanently colocated does not confound the increase in sustained collaborations amongst attendees, I control for the distance between the conference venue and the researcher's primary affiliation in all regression models. Moreover in collaborative outputs between conference attendees, the average collaborative distance between conference collaborators is greater for collaborations between attendees than those between attendees and matched non-attendees as shown in Figure 1D, which implies that the increase in sustained collaborations amongst attendees is not driven by closer permanent geographic proximity.

[6] QML coefficients interpreted in number of times increase as $e^{\text{coefficient of interest}} = e^{2.303} = 10.0$ times

Inventive Direction and Quality of Conference Collaborations

Conditional on collaborations forming after the conference and sustained through publication[7], the ten-year citation count of collaborations is significantly positive for collaborations amongst conference attendees compared to collaborations between attendees and matched non-attendees. In Table 6A, collaborative outputs amongst attendees are 41.2% more cited as depicted in Model 1 and 56.2% more cited for collaborative outputs amongst collaborators with prior collaborations within the conference in Model 3. Further comparing the citation distribution, I find that collaborations between conference attendees are more skewed toward the right tail. Collaborative outputs between conference attendees have 43.8%[8] to 62.7% less odds of receiving no citations as depicted in Table 6B Models 1 to 3, and have 66.2% and 79.3% more odds of citations being in the top 90^{th} percentile of the distribution in Table 6B Models 4 and 6. The number of citations (Table 6A Model 2) and top 90th percentile citations (Table 6B Model 5) for collaborative outputs amongst attendees with no prior collaborations within the conference have negative effect sizes but are insignificant. Taken together, these results suggest that overall collaborative outputs amongst attendees are more likely to be more impactful and less likely to be not cited at all.

[Insert Tables 6A, B about here]

Conferences also slightly steer the inventive direction of these outputs towards the conference's core knowledge space. The number of MeSH terms in common between collaborative outputs and the conference's aggregate top ten percent most frequent MeSH terms is 11.2% higher for within-attendee collaborations as illustrated in Table 7 Model 1 compared to collaborations between attendees and matched non-attendees. Besides being influenced by topics of the conference, coauthors also recombine ideas from their unique field of expertise. Models 2 and 3 in Table 7 suggest that the collaborative outputs tend to be significantly more complementary, where outputs' MeSH keywords are 2.6% more from one or the other author and 1.0% less from overlapped MeSHs of the two authors. These results shed light on whom collaborators choose to match with and suggest that attendees prefer to partner more with collaborators with complementary knowledge rather than those with more redundant expertise.

[Insert Table 7 about here]

Conferences and Citations

Having documented in the prior sections that temporary colocation from conferences has a positive effect on forming and sustaining collaborations and steering their inventive direction, I now turn my analysis to

[7] In this section, all analysis is at the publication level.
[8] Logistic coefficients interpreted odds decreases as $e^{\text{coefficient of interest}} - 1 = e^{-0.576} - 1 = 43.8$ times

evaluating its role as a vehicle for knowledge diffusion. Again I establish a baseline effect on overall forward citations, as Figure 1C illustrates. Prior to attending the conference forward citations of all three groups are very similar in insignificance and trend, with an increase of 4.7% for attendees after the conference as shown for the full sample in Model 1 of Table 8A. Citations between attendees in Model 2 increase insignificantly by 15.1% for the full sample (Table 8A) and by 16.8% for the dissimilar sample (Table 8B), while the similar sample increases significantly by 22.0% (Table 8C). Comparing results from Models 1 and 2 suggest that attendees also substitute other potential works to be cited with those written by fellow conference attendees.

[Insert Tables 8A, B & C about here]

However, attendees tend to cite each other more to begin with, as both the strong and significant effect on *attended* in Model 2 of Tables 8A, 8B and 8C and the graphical trends in Figure 3A suggest. Decomposing the citations between attendees into citations between researchers who have and have not been previously cited by another attendee prior to the conference, I isolate the effect of conferences on establishing new citation links from existing citation behavior. Attending the conference strongly and significantly increases citations by new citers 4.2 times for the full sample in Model 1 of Table 9A, while the effects are stronger for the dissimilar sample than the similar sample with respective increases of 7.5 (Table 9B) times versus 3.7 times (Table 9C). Dissimilar attendees benefit more as their work is less likely to be known to other attendees prior to the conference.

The effect on getting cited by existing citers is much less and also insignificant in Model 2 for all three samples. Since the cited researcher's work is already previously known, going to the conference reveals less information for these relationships. Figures 3B and 3C illustrate these results visually and decompose Figure 3A that illustrates the overall between-attendee citation patterns for the full sample. Given research path dependence and the cumulative nature of knowledge, these citations patterns between attendees is another indication of how conferences guide the inventive direction of attendees toward the conference's knowledge space.

[Insert Figure 3 about here] [Insert Tables 9A, B & C about here]

Junior vs. Senior Attendees

Impacts of attending conferences are also likely to be heterogeneous depending on the tenure of attendees. Decomposing my full sample into junior and senior researchers, I find, by doing model-to-model comparisons between junior researchers in Table 10A and senior researchers in 10B, stronger effects for between-attendee collaborations for junior researchers. Similarly comparing forward citations in Table 11A and 11B, the beneficial effect is stronger for junior than senior researchers. Junior researchers are less embedded in existing collaborative relationships, less known amongst peers, reveal more information compared to senior researchers, and benefit more from attending conferences in terms of finding potential collaborators and getting cited.

[Insert Tables 10A, B and 11A, B about here]

Presenters vs. Non-Presenters

Surprisingly, I find stronger or similar effects on collaborations and citations for non-presenters than presenters as depicted respectively in Tables 12A and 12B as well as Table 13A and 13B. These results suggest that non-presenters and presenters behave differently at conferences, which is driven by several potential explanations. Delving into the composition of presenters, I observe a mild positive correlation (0.2221) between senior attendees and presenters. Thus the weaker effects for senior attendees compared to junior attendees also partly explain the weaker effects for presenters. Also, even though non-presenters do not have a structured and formal way to present their work, they either attend the conference with the aim of working with existing collaborators or they compensate for the lack of being able to formally showcase their work by networking more with other attendees. Moreover, because non-presenters do not need to prepare for any talks, they also have more time to network and establish new connections.

[Insert Tables 12A, B and 13A, B about here]

Qualitative Interviews

Aside from quantitatively exploring the effect of attending conferences on attendees' research trajectory, I gathered qualitative evidence through interviews with several (*n=18*) life science researchers to obtain a more nuanced understanding of the actual processes and mechanism at work. The following quotes are most representative. The first set illustrates how participants gather information and update their information sets when attending conferences, which is one of the crucial steps in establishing both collaborative and citation links with other attendees.

> *"[Conferences] provide a pretty easy way of keeping up to date with the field, staying current, [be]cause it's easier to sit back and hear a series of talks. And again, you are finding out information that's usually unpublished, then to wait and find that information in journals."* – Molecular Geneticist, Canada
>
> *"The most important thing at conferences is what you hear in the halls and in the coffee breaks. For example, we heard about microRNAs way in advance before there were publications, in a train station on our way to a conference."* – Epigeneticist, France

The narratives described by attendees also suggest that informal exchanges of information help in generating new ideas and doing more effective research, and explicitly show how conferences affect the inventive direction of attendees. Put in the context of my quantitative findings, the actions described show conferences as a forum for participants to brainstorm, interpret results and develop together. These interactions and discussions influence the research direction of attendees and are also vital in eventually turning into collaborative or citation ties.

> *"Both of you will hear a talk. You can discuss what you think are the reasons, what's really happening there, to what extent you think it's going to be reproducible, to what extent is this really going to change the way people think, are there other explanations. All these things you can do between sessions, and also talk to people about some surprising thing that you're finding and get input and be able to test ideas with."* – Biologist, US

> *"And it only requires you going along to one conference. We've been clearly influenced. I had a theory, I didn't have any confidence in it, and this guy from Harvard shows up and talked about something utterly different, and you think that's worth doing a few experiments."* – Molecular Biologist, UK

> *"I think it's important to, once in a while, go to a conference outside your immediate field where people are presenting work in different areas, [be]cause somebody might mention something that's a new connection you would not be exposed to that otherwise."* – Geneticist, UK

Surprisingly out of the full sample of interviews, no interviewee specifically mentioned looking for new collaborators as a primary goal or purpose when attending conferences. This suggests that perhaps collaborations happen more organically and form less consciously, as the quotes describe various stages in the matching process that may eventually lead to collaborative relationships[9]. Moreover, none of my interviewees explicitly discuss how attending a conference affects who they cite. However, some of the previous quotes suggest that, for instance, if indeed the Harvard researcher were influential his work would most likely be cited.

Robustness Checks

Collaborations vs. Collaborators and Citations vs. Citers

Since the distribution of how frequently researchers collaborate with one another is not necessarily uniform where some researchers may collaborate with the same individuals multiple times while others may have a broader set of infrequent collaborators, I performed the same set of analyses on unique collaborators instead of collaborations. The same argument applies for citations and citers. However, high correlations between collaborations and collaborators as well as between forward citations and citers as shown in Table A1 of Appendix A indicate that results should be fairly analogous. For ease of comparison between collaborations and collaborators, I show both measures as outcome variables in

[9] A more systematic method to understand why participants go and hope to accomplish at conferences is to survey them, but care must be taken in order not to prime the respondents into answering one way or another.

Table A2, and as expected I observe comparable effects between the two. Again findings are alike for forward citations and citers, and shown in a similar fashion in Table A3.

Matched-Matched Counterfactual Sample

In the analysis of between-attendee collaborations and citations, I used the counterfactual group of attended-matched links to compare to attended-attended links in the DiD. A second plausible comparison measure is the link formed between matched would-be participants themselves, in other words matched-matched links. Table B1 in Appendix B depicts these results for collaborations and collaborators using both attended-matched and matched-matched counterfactuals to facilitate comparison between the two counterfactual groups. Comparing the DiD coefficient between Models 1 and 2, the bigger effect in Model 2 implies that collaborations between attended-attended are greater than attended-matched, which are in turn greater than matched-matched ($Collaborations_{AA} > Collaborations_{AM} > Collaborations_{MM}$). The same trend is also observed for collaborators in Models 3 and 4. I also persistently observe similar tendencies for citations and citers where $Citations_{AA} > Citations_{AM} > Citations_{MM}$ as shown in Table B2. These findings suggest that collaboration and citation links in the matched-matched counterfactual group are even less likely than between the attended-matched group, and can be explained by how matched researchers were determined. Since I matched each participant individually, the matched researchers are closer in knowledge space and other observables to attendees than between themselves.

Other Definitions of Junior Attendee

I include two other definitions for junior attendees – seven and twelve years from first publication to tenure – and find very similar results to the ten-year definition where junior attendees benefit more than senior attendees in terms of subsequent within conference collaborations and citations. In the interest of space, I do not show results herein but they can be obtained from the author.

Panel Regressions with Individual Fixed Effects

Finally, I ran the regressions using individual fixed effects QML models and find as expected very similar effect sizes and significance levels. However, these individual fixed effects also make the *attended* indicator drop out, which only leaves me with interpretable results for DiD models. Again, I do not show results herein but they can be obtained from the author.

DISCUSSION AND CONCLUSION

My findings suggest that in addition to permanent geographic proximity, face-to-face interactions during temporary colocation is an effective mechanism that enhances the exchange of information necessary to establish collaborative relationships and diffuse knowledge. Even though I was careful in my empirical design to address endogeneity concerns – through attendee matching, DiD modeling, isolating collaborations between attendees who have not published together or cited each other before, and dividing the sample to obtain similar versus dissimilar attendees – there are still issues with the choice of attending the conference not being random. Nevertheless, my results add to the literature empirically by being the first to quantitatively show the effects of temporarily colocating at conferences on long-term subsequent collaboration and citation behavior.

Permanent colocation enables convenient interactions and more collaborative trials between limited numbers of neighbors, but is associated with high setup costs of moving permanently together if being proximate is even a possibility. Temporary colocation in contrast is less costly and more plausible. Although short colocation timeframes limit the number of collaborative trials, bringing many researchers together at conferences expose attendees to a more diverse set of ideas. Moreover, the limited timeframe force potential collaborators to be more concentrated on the work and collaborative pitches to be of better quality, as there is less time for trial and error. Thus, this work contributes theoretically to the growing literature exploring the phenomenon of collaboration and sheds new light on temporary colocation as a mechanism for the formation of sustained collaborative relationships and citation links. Since I track the collaborative and citation behavior of each conference attendee at the individual level, this work contributes to the literature on the micro-foundations of innovation.

This research also speaks to the agglomeration literature at the individual level of analysis. Rewards of agglomeration stem from individuals or organizations being permanently located close to one another to reap from knowledge spillovers (Head et al., 1995, Chung and Alcácer, 2002), economies of scale (Krugman, 1991), and labor pooling (Ellison et al., 2014). On the other hand, diseconomies of agglomeration point to disadvantages such as increased competition (Alcacer and Chung, 2007) and congestion. Temporary colocation may be a substitute to gain from benefits of being permanently colocated while avoiding the pitfalls of excessive competition.

Policymakers should consider supporting and fostering conferences as an alternative and less costly form of colocation and proximity that temporarily brings individuals together to exchange ideas and diffuse knowledge. Researchers can draw direct implications on the course of their research trajectory from my results. Similarly, this work informs whether managers of science and technology-intensive firms should commit substantial funds for employees to participate in professional or academic conferences. While overall productivity may not be dramatically improved, the subsequent direction of R&D activities will likely be. Thus, they should be careful in choosing conference topics appropriate for their organization's innovation strategy, whether looking to exploit or explore existing technology portfolios and know-how.

Future Research

It is important to stress that the findings herein shed light on the research trajectory following the attendance of a specific conference, rather than showing the marginal benefit of attending an additional conference. Without information and variation on how many conferences researchers attend in a given timespan, I can only speculate on how it may affect their overall productivity including collaborations and citations. For instance, if attendees in my sample participated in an equal or fewer numbers of conferences than their matched counterpart, the small effect sizes observed herein would be a conservative and underestimation of their ensuing overall productivity. On the other hand, if attendees in my sample were present in more conferences than the matched researchers, then the most substantial effect of attending conferences is the subsequent research trajectory of attendees rather than overall productivity. Future research could find this marginal effect by systematically obtaining the number of conferences a researcher attends through survey or content analysis of CVs.

The setting of this paper presents a homogeneous structure in that GRCs are small, weeklong meetings taking place in a geographically isolated area. Future research could compare how different conference structures affect outcomes to obtain greater generalizability. This would help conference organizers in designing a setup that would be the most effective for their goals. These factors include size – big conferences with thousands or more attendees versus medium conferences with a few hundred versus small conferences sometimes with less than one hundred attendees, length – one to two-day conferences versus weeklong ones, location – in isolated rural areas versus urban cities, setup – parallel versus sequential sessions, etc. Moreover, these settings can also be further extended to industrial meetings and trade shows beyond the scientific institution.

Another interesting extension is to investigate the same questions but in the context of virtual conferences. Virtual meetings have been proposed and increasingly organized as an alternative to decrease the cost of physical ones especially with the widespread penetration of the Internet. Although they bring participants together and attempt to create a virtual sense of spatial proximity, they lack the direct connections to other attendees afforded by physical conferences. Thus investigating this question is interesting empirically as the effect between attending physical and virtual conferences on subsequent productivity, collaborative behavior and diffusion can be compared, but also theoretically as it introduces another dimension of proximity – virtual proximity.

The empirical evidence presented in this work suggests that conferences not only spur collaboration between attendees where other potential collaborators are substituted by ones encountered at the conference, they also positively influence the quality and inventive direction of the collaborative output as well as within-attendee citations between the more similar attendees. Assuming path dependence of research and the cumulative nature of knowledge, these results imply that over time conferences can have a significant impact in steering the research path of attendees – from the works that they cite and build upon, to the colleagues who they choose to collaborate with, and the research direction they undertake while collaborating with fellow attendees.

The unique dataset of attendees participating in a small and weeklong conference ensures in-person interactions not captured in extant works on collaborations that mostly rely on existing relationships in the form of co-authorship on publications or co-inventorship on patents. Thus, this work sheds light on the emerging literature that investigates how collaborative relationships form by focusing on the effect of temporary colocation as a catalyst that facilitates matching between potential collaborators and their sustained long-term collaborative outputs. Conferences also enable researchers to showcase their research better and establish a community as evidenced by the increase in within-attendee citation between attendees who have never been cited by other attendees before. Moreover, all these results have stronger effects on junior researchers than senior researchers, and surprisingly non-presenters versus presenters. Thus, despite the high costs and other critiques of conferences, they may be still worthwhile to attend especially for junior researchers and even as a non-presenter, both in terms of finding suitable collaborators to work with and disseminating work amongst peers.

REFERENCES

ADAMS, J. D. 1990. Fundamental Stocks of Knowledge and Productivity Growth. *The Journal of Political Economy,* 98**,** 673-702.

AGRAWAL, A. & GOLDFARB, A. 2008. Restructuring Research: Communication Costs and the Democratization of University Innovation. *American Economic Review,* 98**,** 1578-1590.

ALCACER, J. & CHUNG, W. 2007. Location strategies and knowledge spillovers. *Management Science,* 53**,** 760-776.

ALMEIDA, P. & KOGUT, B. 1999. Localization of Knowledge and the Mobility of Engineers in Regional Networks. *Management Science,* 45**,** 905-917.

BOUDREAU, K., GANGULI, I., GAULE, P., GUINAN, E. & LAKHANI, K. 2012. Colocation and scientific collaboration: evidence from a field experiment. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*.

CATALINI, C. 2012. Microgeography and the Direction of Inventive Activity. *Working paper*.

CHUNG, W. & ALCÁCER, J. 2002. Knowledge seeking and location choice of foreign direct investment in the United States. *Management Science,* 48**,** 1534-1554.

ELLISON, G., GLAESER, E. L. & KERR, W. 2014. What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review,* forthcoming.

FAFCHAMPS, M., VAN DER LEIJ, M. J. & GOYAL, S. 2010. Matching and Network Effects. *Journal of the European Economic Association,* 8**,** 203-231.

FLEMING, L. 2001. Recombinant Uncertainty in Technological Search. *Management Science,* 47**,** 117-132.

FREEMAN, R. B., GANGULI, I. & MURCIANO-GOROFF, R. 2013. Why and wherefore of increased scientific collaboration. *The Changing Frontier: Rethinking Science and Innovation Policy.* University of Chicago Press.

GIROTRA, K., TERWIESCH, C. & ULRICH, K. T. 2010. Idea Generation and the Quality of the Best Idea. *Management Science,* 56**,** 591-605.

GREEN, M. 2008. Are international medical conferences an outdated luxury the planet can‚Äôt afford? Yes. *BMJ: British Medical Journal,* 336**,** 1466.

HEAD, K., RIES, J. & SWENSON, D. 1995. Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States. *Journal of international economics,* 38**,** 223-247.

HENDERSON, R. M. & CLARK, K. B. 1990. Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly,* 35**,** 9-30.

IOANNIDIS, J. A. 2012. ARe medical conferences useful? and for whom? *JAMA,* 307**,** 1257-1258.

JANIS, I. L. 1971. Groupthink. *Psychology Today,* 5**,** 43-46, 74-76.

JONES, B. F. 2009. The Burden of Knowledge and the “Death of the Renaissance Man”: Is Innovation Getting Harder? *Review of Economic Studies,* 76**,** 283-317.

KATZ, J. 1994. Geographical proximity and scientific collaboration. *Scientometrics,* 31**,** 31-43.

KRUGMAN, P. R. 1991. *Geography and trade*, MIT press.

MANSFIELD, E. 1972. Contribution of R&D to Economic Growth in the United States. *Science,* 175**,** 477-486.

MCFADYEN, M. A. & CANNELLA, A. A. J. 2004. Social Capital and Knowledge Creation: Diminishing Returns of the Number and Strength of Exchange Relationships *Academy of Management Journal,* 47**,** 735-746.
MCKIE, D. 1960. The origins and foundation of the Royal Society of London. *notes and Records of the Royal Society of London,* 15**,** 1-37.
ROSENBERG, N. 1974. Science, Invention and Economic Growth. *The Economic Journal,* 84**,** 90-108.
SAB 2013. Trends in Scientific Conferences. http://www.scienceboard.net/studies/studies.asp?studyId=142: Science Advisory Board.
SAXENIAN, A. 1994. Regional advantage: Culture and competition in Silicon Valley and Route 128. *Harvard University, Cambridge, MA*.
SINGH, J. & FLEMING, L. 2010. Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Management Science,* 56**,** 41-56.
TORVIK, V. I. & SMALHEISER, N. R. 2009. Author Name Disambiguation in MEDLINE. *ACM transactions on knowledge discovery from data,* 3**,** 1-29.
WEITZMAN, M. L. 1998. Recombinant Growth. *The Quarterly Journal of Economics,* 113**,** 331-360.
WUCHTY, S., JONES, B. F. & UZZI, B. 2007. The Increasing Dominance of Teams in Production of Knowledge. *Science,* 316**,** 1036-1039.
ZUCKER, L. G., DARBY, M. R. & BREWER, M. B. 1998. Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises. *The American Economic Review,* 88**,** 290-306.

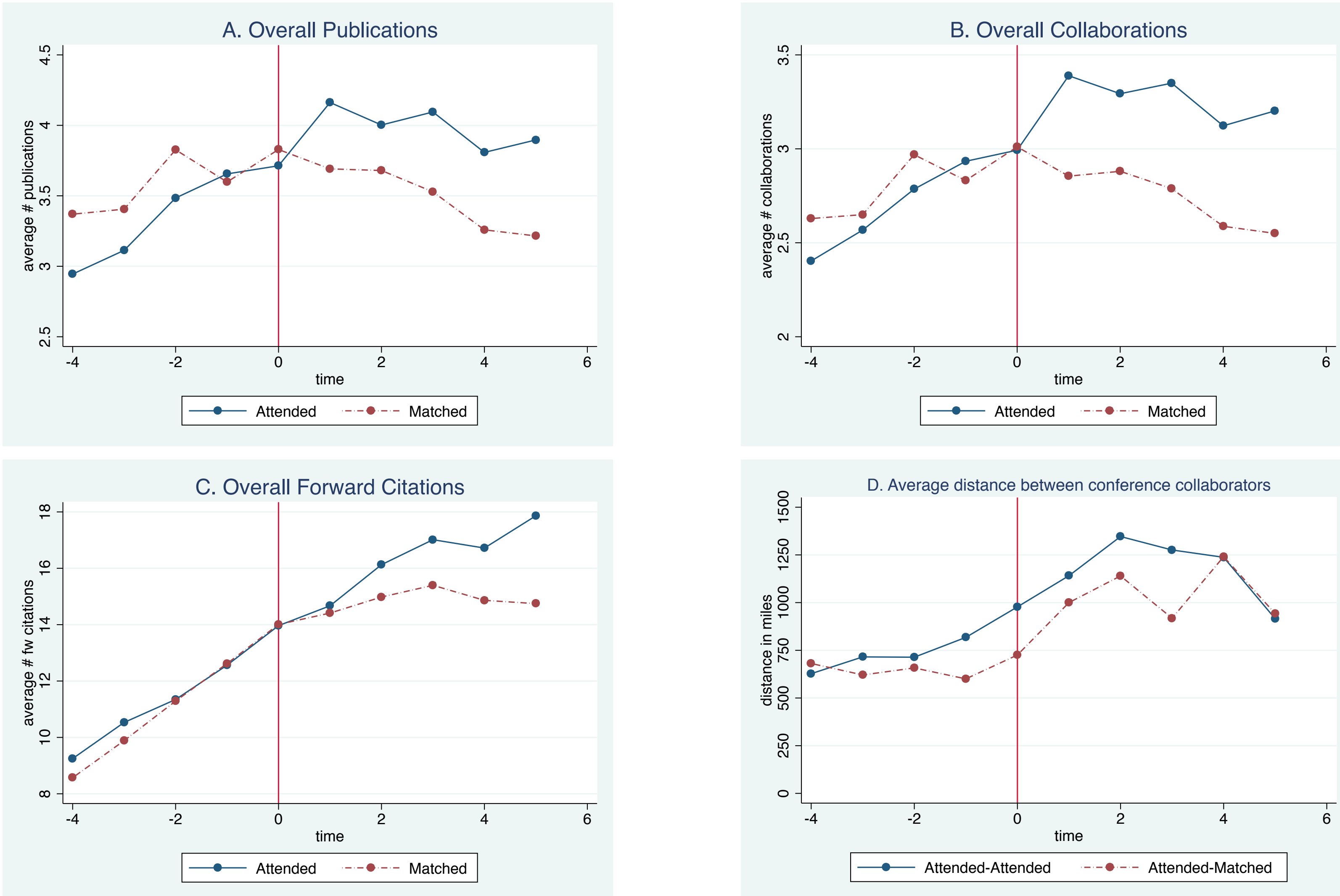

Figure 1 – These figures show yearly average trends before and after conference of overall publications, collaborations, forward citations for attendees and non-attendees for the full sample, as well as yearly trends before and after conference of the average collaborative distance for publications between attendees and non-attendees.

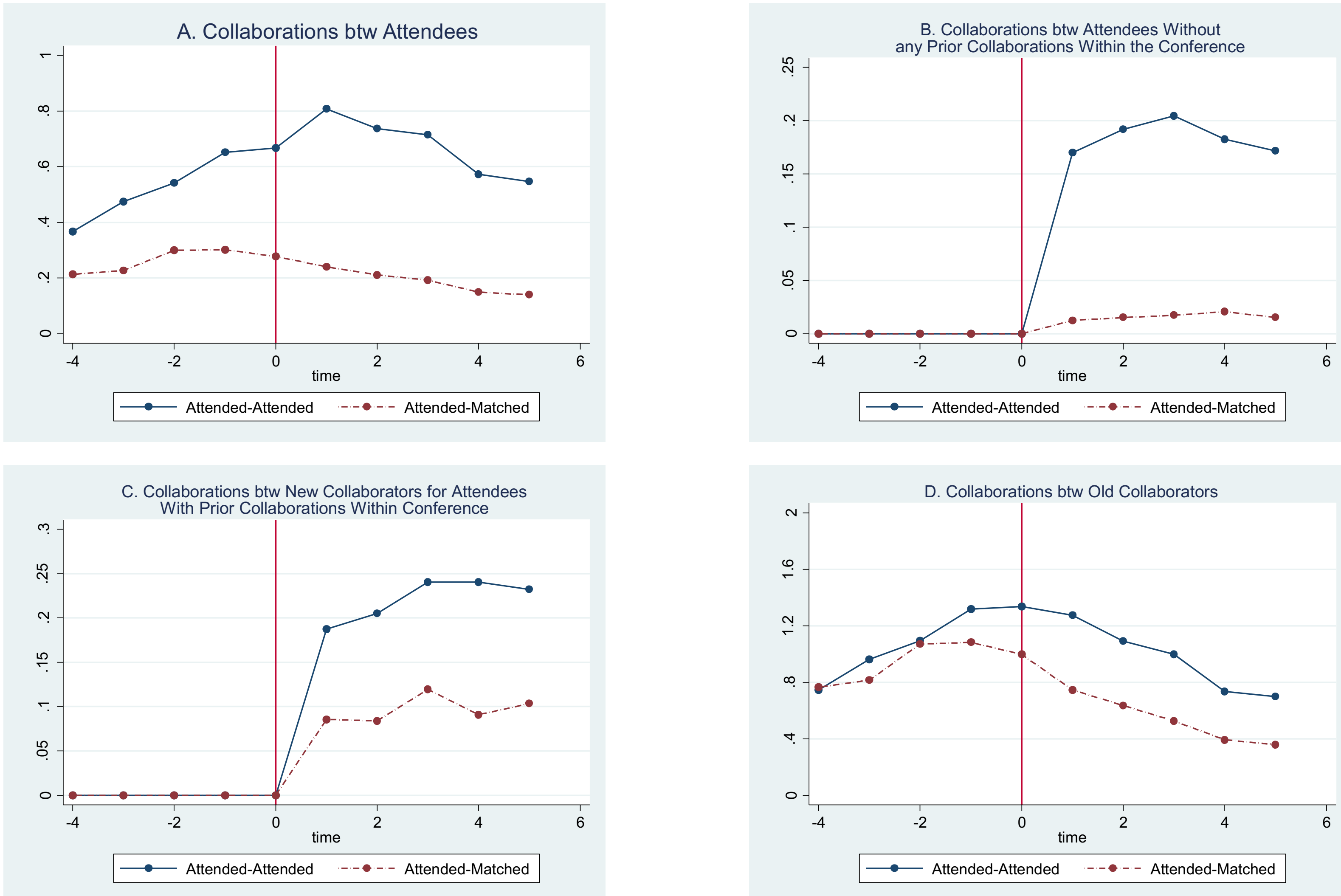

Figure 2 – These figures show yearly average trends before and after conference of collaborations amongst attendees and non-attendees for the full sample.

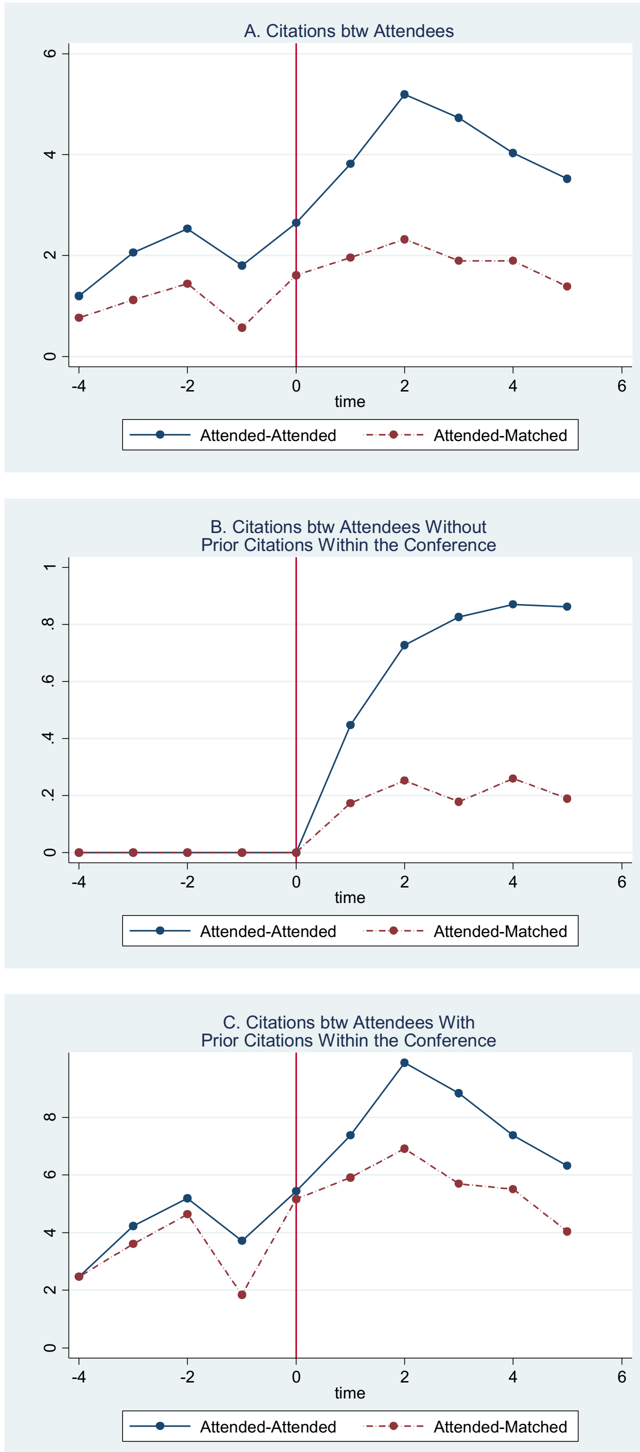

Figure 3 – These figures show yearly average trends before and after conference of forward citations five years after publication between attendees and non-attendees for the full sample.

Characteristic	Attended	Matched	Two tailed t-test
experience	13.20	12.57	0.09
# publications	3.38	3.61	0.30
# collaborations	2.74	2.82	0.59
# forward citations	11.53	11.28	0.78
n	1265	2016	

Table 1 – This table shows the comparison of attended and matched researcher observables based on 5-year average measures prior to conference.

Variable	# Obs	Mean	Std. Dev.	Min	Max
post	32810	0.500	0.500	0	1
attended	32810	0.386	0.487	0	1
post x attended	32810	0.193	0.394	0	1
junior	32810	0.477	0.499	0	1
presenter	32810	0.432	0.495	0	1
experience	32800	12.887	9.939	0	51
distance to conference	30180	2400.9	2096.4	25.5	11530.5
# publications	32810	3.597	6.384	0	131
# collaborations	32810	2.863	4.688	0	130
# collaborators	32810	9.370	16.300	0	325
# collaborations within attended	32810	0.372	1.119	0	21
# collaborations within attended_m	32810	0.353	1.105	0	21
# collaborators within attended	32810	0.292	0.795	0	10
# collaborators within attended_m	32810	0.261	0.724	0	10
# collaborations within attendees new	32810	0.051	0.323	0	11
# collaborations within attendees old	32810	0.321	1.058	0	21
# citations for within-attendee collaborations	2353	5.173	13.576	0	338
zero citation indicator	2353	0.414	0.493	0	1
top90th citation indicator	2353	0.103	0.304	0	1
# MeSH in common with conference	2353	7.726	3.077	0	17
MeSH fraction from both	5295	0.194	0.129	0	1
MeSH fraction from one or other	5295	0.211	0.159	0	1
average collaborative distance	4675	907.5	1703	0	11862.4
# citations	32810	13.435	27.589	0	661
# citers	32810	50.197	114.598	0	2103
# citations within attended	32810	2.134	8.202	0	247
# citations within attended_m	32810	1.753	7.000	0	247
# citers within attended	32810	0.815	2.159	0	39
# citers within attended_m	32810	0.673	1.853	0	39

Table 2 – This table shows summary statistics for all variables used in the analysis.

		Collaborations (within attendees)				**Citations** (within attendees)		
		All collaborations within conference	**No prior collaborations within conference**	**Prior collaborations within conference (new collaborators)**	**Prior collaborations within conference (old collaborators)**	**All citations within conference**	**No prior citations within conference**	**Prior citations within conference**
	Reason for data cut		Address endogeneity of existing collaborators attending the same conference				Address endogeneity of existing citers attending the same conference	
Full sample		+ 40.8%**	+ 11.8 times**	+ 2.5 times**	+ 32.5%**	+ 15.1%	+ 4.2 times**	+ 14.9%
Dissimilar	Understand effect of conference when attendee is dissimilar or similar to others and address endogeneity of going	+ 22.8%**	+ 14.5 times**	+ 84.0%**	+ 19.0%**	+ 16.8%	+ 7.5 times**	+ 18.3%
Similar		+ 58.4%**	+ 10.0 times**	+ 3.2 times**	+ 45.4%**	+ 22.0%*	+ 3.7 times**	+ 18.7%
Junior	Understand effect of conference depending on career stage of attendee		+ 17.2 times**	+ 2.9 times**	+ 62.1%**	+ 31.1%**	+ 5.1 times**	+ 24.5%
Senior			+ 8.9 times**	+ 2.3 times**	+ 17.6%*	+ 19.7%*	+ 3.9 times**	+ 20.9%+
Presenter	Understand effect of conference depending on role taken during conference		+ 10.8 times**	+ 2.3 times**	+ 20.9%*	+ 18.3%*	+ 4.5 times**	+ 22.0%*
Non-presenter			+ 12.0 times**	+ 2.7 times**	+ 49.8%**	+ 30.9%*	+ 4.4 times**	+ 20.3%

+ p<0.10, * p<0.05, ** p<0.01

Table 3 – This table summarizes the various data samples used in the analysis of within attendee collaborations and citations, and the respective results compared to collaborations and citations between attendees and matched non-attendees.

Table 4A	**Model 1**	**Model 2**	**Model 3**
Full sample	# publications	# collaborations	# collaborations within attendees
post	-0.0244**	-0.0235**	-0.335**
	-0.00872	-0.00638	-0.0386
attended	0.0152	0.0534**	0.774**
	-0.032	-0.011	-0.0728
post*attended	0.0307*	0.0132*	0.342**
	-0.0136	-0.00672	-0.0516
ln(experience)	0.181**	0.0427**	-0.327**
	-0.0283	-0.00853	-0.0538
ln(publications)		1.371**	1.525**
		-0.00815	-0.0223
ln(citations)	-0.00944	0.00138	-0.0168
	-0.0185	-0.0063	-0.0225
ln(collaborations)	1.368**		
	-0.0267		
ln(distance to conference)	-0.00969	0.00556	-0.116**
	-0.00702	-0.0041	-0.0256
_cons	-1.251**	-1.313**	-2.016**
	-0.0717	-0.0453	-0.301
lnalpha			
_cons	-1.863**	-2.284**	1.120**
	-0.119	-0.0526	-0.0407
conference fe	y	y	y
N	30170	30170	30170
Log lik.	-42592.5	-38306.1	-15010.9

+ p<0.10, * p<0.05, ** p<0.01

Table 4B	**Model 1**	**Model 2**	**Model 3**
Dissimilar sample	# publications	# collaborations	# collaborations within attendees
post	-0.0146	-0.0188*	-0.241**
	-0.0109	-0.00885	-0.0609
attended	0.112*	0.0273	0.869**
	-0.0538	-0.022	-0.125
post*attended	0.0475*	0.00549	0.205**
	-0.0232	-0.0099	-0.0767
ln(experience)	0.302**	0.0403*	-0.389**
	-0.057	-0.0175	-0.0904
ln(publications)		1.280**	1.391**
		-0.013	-0.0421
ln(citations)	-0.0437	0.0322**	0.0651
	-0.036	-0.0113	-0.0431
ln(collaborations)	1.228**		
	-0.0406		
ln(distance to	0.0146	-0.00168	-0.121*
conference)	-0.0114	-0.00701	-0.0511
_cons	-1.384**	-1.176**	-1.856**
	-0.11	-0.0614	-0.533
lnalpha			
_cons	-1.573**	-2.186**	1.268**
	-0.198	-0.087	-0.0613
conference fe	y	y	y
N	10830	10830	10830
Log lik.	-17091.7	-15306.1	-5834.6

+ p<0.10, * p<0.05, ** p<0.01

Table 4C	**Model 1**	**Model 2**	**Model 3**
Similar sample	# publications	# collaborations	# collaborations within attendees
post	-0.0573**	-0.0460**	-0.419**
	(0.00774)	(0.00714)	(0.0546)
attended	-0.0257	0.0725**	0.736**
	(0.0220)	(0.0167)	(0.0877)
post*attended	0.0114	0.0250**	0.460**
	(0.0203)	(0.00811)	(0.0655)
ln(experience)	0.0925**	0.0523**	-0.278**
	(0.0127)	(0.00880)	(0.0504)
ln(publications)		1.455**	1.623**
		(0.0103)	(0.0255)
ln(citations)	0.0367**	-0.0113**	-0.0677**
	(0.00524)	(0.00437)	(0.0193)
ln(collaborations)	1.505**		
	(0.0302)		
ln(distance to	-0.0156*	0.0157**	-0.100**
conference)	(0.00792)	(0.00547)	(0.0315)
_cons	-1.495**	-1.446**	-2.123**
	(0.0947)	(0.0662)	(0.557)
lnalpha			
_cons	-2.200**	-2.430**	1.004**
	(0.105)	(0.0833)	(0.0468)
conference fe	y	y	y
N	19340	19340	19340
Log lik.	-25141.2	-22862.0	-9135.6

+ p<0.10, * p<0.05, ** p<0.01

Table 4A, B & C – These tables show difference-in-differences QML Poisson count regression models with panel random effects for the full sample in A, dissimilar sample in B, and similar sample in C. The dependent variables are respectively: overall publications in model 1, overall collaborations in model 2, and collaborations between attendees in model 3. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.

Table 5A	No prior collaborations w/in conference	Prior collaborations w/in conference	
	Model 1	**Model 2**	**Model 3**
Full sample	# collaborations within attendees	# collaborations within attendees (between new collaborators)	# collaborations within attendees (between old collaborators)
post			-0.532**
			(0.0488)
attended	2.465**	0.901**	0.0442
	(0.192)	(0.118)	(0.0516)
post*attended			0.281**
			(0.0673)
ln(experience)	-0.453**	-0.0301	-0.273**
	(0.106)	(0.0805)	(0.0327)
ln(publications)	1.536**	1.298**	1.375**
	(0.0979)	(0.0636)	(0.0253)
ln(citations)	-0.0908	0.111*	-0.0647**
	(0.0752)	(0.0547)	(0.0207)
ln(distance to conference)	-0.126+	0.0278	-0.0246
	(0.0710)	(0.0485)	(0.0223)
_cons	-4.816**	-5.784**	-1.439**
	(0.780)	(0.421)	(0.277)
lnalpha			
_cons	1.005**	0.604**	-0.401**
	(0.164)	(0.124)	(0.0526)
conference fe	y	y	y
N	9330	5755	11510
Log lik.	-1407.7	-2032.4	-11373.8

+ p<0.10, * p<0.05, ** p<0.01

Table 5B	No prior collaborations w/in conference	Prior collaborations w/in conference	
	Model 1	**Model 2**	**Model 3**
Dissimilar sample	# collaborations within attendees	# collaborations within attendees (between new collaborators)	# collaborations within attendees (between old collaborators)
post			-0.449**
			(0.0677)
attended	2.674**	0.610**	-0.0712
	(0.303)	(0.209)	(0.0989)
post*attended			0.174*
			(0.0828)
ln(experience)	-0.471*	-0.258	-0.344**
	(0.210)	(0.194)	(0.0638)
ln(publications)	1.444**	1.248**	1.255**
	(0.183)	(0.101)	(0.0330)
ln(citations)	-0.108	0.0653	0.0327
	(0.113)	(0.0949)	(0.0426)
ln(distance to	-0.123	0.106	-0.0397
conference)	(0.0819)	(0.0847)	(0.0441)
_cons	-4.403**	-5.252**	-1.101**
	(0.770)	(0.764)	(0.388)
lnalpha			
_cons	1.014**	0.402*	-0.242**
	(0.339)	(0.157)	(0.0821)
conference fe	y	y	y
N	3435	1980	3960
Log lik.	-496.1	-888.4	-4504.4

+ p<0.10, * p<0.05, ** p<0.01

Table 5C	No prior collaborations w/in conference	Prior collaborations w/in conference	
	Model 1	**Model 2**	**Model 3**
Similar sample	# collaborations within attendees	# collaborations within attendees (between new collaborators)	# collaborations within attendees (between old collaborators)
post			-0.608**
			(0.0558)
attended	2.303**	1.155**	0.102+
	(0.277)	(0.192)	(0.0585)
post*attended			0.371**
			(0.0606)
ln(experience)	-0.380**	0.124	-0.210**
	(0.109)	(0.125)	(0.0403)
ln(publications)	1.665**	1.412**	1.458**
	(0.129)	(0.0745)	(0.0319)
ln(citations)	-0.00358	0.203*	-0.126**
	(0.0882)	(0.0869)	(0.0252)
ln(distance to	-0.138*	0.00252	-0.00438
conference)	(0.0702)	(0.0678)	(0.0289)
_cons	-6.569*	-6.443*	-1.632**
	(2.651)	(2.779)	(0.582)
lnalpha			
_cons	0.809**	0.588**	-0.578**
	(0.227)	(0.182)	(0.0584)
conference fe	y	y	y
N	5895	3775	7550
Log lik.	-897.8	-1125.9	-6819.8

+ p<0.10, * p<0.05, ** p<0.01

Table 5A, B & C – These tables show simplified and difference-in-differences QML Poisson count regression models with panel random effects for the full sample in A, dissimilar sample in B, and similar sample in C. The dependent variables are respectively: collaborations between attendees with no prior collaborations within the conference in model 1, collaborations between new collaborators for attendees with prior collaborations within the conference in model 2, and collaborations between old collaborators in model 3. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.

Table 6A	All collaborative outputs	No prior collaborations w/in conference	Prior collaborations w/in conference
	Model 1	**Model 2**	**Model 3**
	# citations for within-attendee collaborations	# citations for within-attendee collaborations	# citations for within-attendee collaborations
attended	0.345*	-0.149	0.446**
	(0.148)	(0.346)	(0.163)
ln(average experience)	0.0616	0.0630	0.0744
	(0.203)	(0.263)	(0.253)
ln(collaborators)	0.505**	0.635*	0.500*
	(0.194)	(0.273)	(0.210)
_cons	1.197**	0.950	1.165*
	(0.450)	(1.504)	(0.504)
conference fe	y	y	y
N	2353	274	2079
Log lik.	-13159.5	-1708.5	-11272.4

+ p<0.10, * p<0.05, ** p<0.01

Table 6A – This table shows QML Poisson count regression models for the sample of collaborative outputs conditional on collaborations between attendees forming. The dependent variables are respectively: 10-year forward citations of the collaborative output in model 1, 10-year forward citations of the collaborative output if the output is between collaborators with no prior collaborative link within the conference in model 2, and 10-year forward citations of the collaborative output if the output is between collaborators with prior collaborative links within the conference in model 3. Robust standard errors are in parentheses. Conference fixed effects are also included.

Table 6B	All collaborative outputs	No prior collaborations w/in conference	Prior collaborations w/in conference	All collaborative outputs	No prior collaborations w/in conference	Prior collaborations w/in conference
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
	zero citation indicator	zero citation indicator	zero citation indicator	top90th citation indicator	top90th citation indicator	top90th citation indicator
attended	-0.619**	-0.985*	-0.576**	0.508*	-0.0501	0.584*
	(0.141)	(0.397)	(0.152)	(0.217)	(0.521)	(0.247)
ln(average	0.191	0.620	0.0979	-0.332	0.377	-0.468+
experience)	(0.161)	(0.480)	(0.174)	(0.223)	(0.469)	(0.267)
ln(collaborators)	-0.752**	-0.428	-0.817**	0.974**	1.715**	0.917**
	(0.144)	(0.484)	(0.154)	(0.205)	(0.568)	(0.223)
_cons	-0.194	-0.476	-0.0288	-2.669**	-7.337**	-2.075*
	(0.623)	(1.894)	(0.680)	(0.860)	(2.500)	(0.964)
conference fe	y	y	y	y	y	y
N	2353	255	2079	1842	190	1629
Log lik.	-1179.4	-116.2	-1046.3	-614.3	-80.27	-521.3

+ p<0.10, * p<0.05, ** p<0.01

Table 6B – This table shows logistic regression models for the sample of collaborative outputs conditional on collaborations between attendees forming. The dependent variables are respectively: zero citation indicator for all collaborative outputs in model 1, zero citation indicator if the output is between collaborators with no prior collaborative link within the conference in model 2, zero citation indicator if the output is between collaborators with prior collaborative links within the conference in model 3, top 90th percentile citation indicator for all collaborative outputs in model 4, top 90th percentile citation if the output is between collaborators with no prior collaborative link within the conference in model 5, and top 90th percentile citation indicator if the output is between collaborators with prior collaborative links within the conference in model 6. Robust standard errors are in parentheses. Conference fixed effects are also included.

Table 7	**Model 1**	**Model 2**	**Model 3**
	# MeSH in common with conference	MeSH fraction from one or other	MeSH fraction from both
attended	0.106**	0.0256**	-0.00995*
	(0.0219)	(0.00553)	(0.00494)
ln(collaborators)	0.0640**		
	(0.0211)		
_cons	2.233**	0.179**	0.183**
	(0.0906)	(0.0117)	(0.00913)
conference fe	y	y	y
N	2353	5294	5294
Log lik.	-5961.8		
R2		0.0740	0.0372

+ p<0.10, * p<0.05, ** p<0.01

Table 7 – This table shows QML Poisson count and OLS regression models for the sample of collaborative outputs conditional on collaborations between attendees forming. The dependent variables are respectively: the number of common MeSH keywords between the collaborative output and the conference in model 1, the fraction of MeSH keywords of the collaborative output from one or the other attendee coauthors in model 2, and the fraction of MeSH keywords from both attendee coauthors in model 3. Robust standard errors are in parentheses. Conference fixed effects are also included.

Table 8A	**Model 1**	**Model 2**
Full sample	# citations	# citations within attendees
post	0.288**	0.560**
	(0.0231)	(0.0671)
attended	-0.0343	0.707**
	(0.0503)	(0.0914)
post*attended	0.0457	0.141
	(0.0379)	(0.0859)
ln(experience)	0.830**	0.519**
	(0.0300)	(0.0411)
ln(publications)	-0.0220	0.0497
	(0.0757)	(0.0971)
ln(collaborations)	0.216**	0.312**
	(0.0800)	(0.119)
ln(distance to conference)	-0.100**	-0.120**
	(0.0181)	(0.0326)
_cons	1.089**	-1.092**
	(0.191)	(0.271)
lnalpha		
_cons	0.433**	1.396**
	(0.0244)	(0.0251)
conference fe	y	y
N	30170	30170
Log lik.	-98717.3	-48327.4

+ p<0.10, * p<0.05, ** p<0.01

Table 8B	**Model 1**	**Model 2**
Dissimilar sample	# citations	# citations within attendees
post	0.234**	0.420**
	(0.0412)	(0.0848)
attended	0.0609	0.865**
	(0.0832)	(0.160)
post*attended	0.0264	0.155
	(0.0593)	(0.106)
ln(experience)	0.931**	0.664**
	(0.0583)	(0.0838)
ln(publications)	-0.0974	0.0839
	(0.148)	(0.180)
ln(collaborations)	0.323*	0.413*
	(0.163)	(0.203)
ln(distance to conference)	-0.121**	-0.108*
	(0.0319)	(0.0519)
_cons	0.782**	-2.042**
	(0.282)	(0.451)
lnalpha		
_cons	0.577**	1.357**
	(0.0465)	(0.0537)
conference fe	y	y
N	10830	10830
Log lik.	-40990.7	-20024.1

+ p<0.10, * p<0.05, ** p<0.01

Table 8C	**Model 1**	**Model 2**
Similar sample	# citations	# citations within attendees
post	0.354**	0.673**
	(0.0265)	(0.0718)
attended	-0.124*	0.565**
	(0.0547)	(0.114)
post*attended	0.0910	0.199*
	(0.0574)	(0.0828)
ln(experience)	0.657**	0.343**
	(0.0374)	(0.0692)
ln(publications)	0.0597	0.142+
	(0.0425)	(0.0790)
ln(collaborations)	0.0857+	0.0983
	(0.0452)	(0.0923)
ln(distance to conference)	-0.0836**	-0.105**
	(0.0214)	(0.0343)
_cons	1.823**	-0.0736
	(0.225)	(0.391)
lnalpha		
_cons	0.278**	1.383**
	(0.0348)	(0.0377)
conference fe	y	y
N	19340	19340
Log lik.	-57264.5	-28025.6

+ p<0.10, * p<0.05, ** p<0.01

Table 8A, B & C – These tables show difference-in-differences QML Poisson count regression models with panel random effects for the full sample in A, dissimilar sample in B, and similar sample in C. The dependent variables are respectively: overall forward citations in model 1, and between-attendee citations in model 2. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.

Table 9A	Citations btw att w/ no prior citation links w/in conference	Citations btw att w/ prior citation links w/in conference
	Model 1	**Model 2**
Full sample	# citations within attendees	# citations within attendees
post		0.482**
		(0.0606)
attended	1.435**	0.133
	(0.135)	(0.110)
post*attended		0.139
		(0.0845)
ln(experience)	-0.0332	0.196**
	(0.0929)	(0.0429)
ln(publications)	0.0524	0.0645
	(0.371)	(0.111)
ln(collaborations)	0.520	0.289*
	(0.414)	(0.139)
ln(distance to conference)	-0.183**	-0.0283
	(0.0577)	(0.0272)
_cons	-2.546**	0.135
	(0.599)	(0.276)
lnalpha		
_cons	1.786**	0.101**
	(0.0622)	(0.0321)
conference fe	y	y
N	17880	12290
Log lik.	-7556.8	-40543.4

+ p<0.10, * p<0.05, ** p<0.01

Table 9B	Citations btw att w/ no prior citation links w/in conference	Citations btw att w/ prior citation links w/in conference
	Model 1	**Model 2**
Dissimilar sample	# citations within attendees	# citations within attendees
post		0.363**
		(0.0988)
attended	2.015**	0.303*
	(0.256)	(0.148)
post*attended		0.168
		(0.123)
ln(experience)	-0.333*	0.288**
	(0.153)	(0.0895)
ln(publications)	-0.0124	0.112
	(0.745)	(0.224)
ln(collaborations)	0.701	0.371
	(0.783)	(0.246)
ln(distance to conference)	0.0722	-0.0211
	(0.105)	(0.0410)
_cons	-3.820**	-0.614
	(0.902)	(0.451)
lnalpha		
_cons	1.758**	0.0771
	(0.158)	(0.0538)
conference fe	y	y
N	6090	4740
Log lik.	-2227.9	-17573.7

+ p<0.10, * p<0.05, ** p<0.01

Table 9C	Citations btw att w/ no prior citation links w/in conference	Citations btw att w/ prior citation links w/in conference
	Model 1	**Model 2**
Similar sample	# citations within attendees	# citations within attendees
post		0.581**
		(0.0716)
attended	1.307**	-0.0343
	(0.127)	(0.0971)
post*attended		0.171
		(0.108)
ln(experience)	0.0396	0.0433
	(0.113)	(0.0697)
ln(publications)	0.157	0.146
	(0.317)	(0.0975)
ln(collaborations)	0.354	0.0889
	(0.362)	(0.112)
ln(distance to conference)	-0.220**	-0.00628
	(0.0626)	(0.0372)
_cons	-3.616	0.873*
	(9.294)	(0.370)
lnalpha		
_cons	1.699**	0.0459
	(0.0765)	(0.0504)
conference fe	y	y
N	11790	7550
Log lik.	-5302.5	-22751.3

+ p<0.10, * p<0.05, ** p<0.01

Table 9A, B & C – These tables show simplified and difference-in-differences QML Poisson count regression models with panel random effects for the full sample in A, dissimilar sample in B, and similar sample in C. The dependent variables are respectively: forward citations from attendees with no prior citation links within the conference in model 1, and forward citations from attendees with prior citation links within the conference in model 2. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.

Table 10A	No prior collaborations w/in conference	Prior collaborations w/in conference	
Junior attendees	**Model 1**	**Model 2**	**Model 3**
	# collaborations within attendees	# collaborations within attendees (between new collaborators)	# collaborations within attendees (between old collaborators)
post			-0.736**
			(0.0891)
attended	2.847**	1.058**	0.119+
	(0.333)	(0.285)	(0.0619)
post*attended			0.483**
			(0.0983)
ln(experience)	-0.581**	-0.0344	-0.406**
	(0.168)	(0.220)	(0.0526)
ln(publications)	1.947**	1.371**	1.576**
	(0.160)	(0.119)	(0.0374)
ln(citations)	-0.183*	0.122	-0.104**
	(0.0875)	(0.139)	(0.0256)
ln(distance to conference)	-0.105	0.268*	-0.0173
	(0.122)	(0.117)	(0.0328)
_cons	-4.315	-7.048	-1.666**
	(9.050)	(5.222)	(0.540)
lnalpha			
_cons	0.864**	0.828**	-0.742**
	(0.253)	(0.238)	(0.0728)
conference fe	y	y	y
N	4255	2515	5030
Log lik.	-518.5	-585.7	-4231.2

+ p<0.10, * p<0.05, ** p<0.01

Table 10B	No prior collaborations w/in conference	Prior collaborations w/in conference	
Senior attendees	**Model 1**	**Model 2**	**Model 3**
	# collaborations within attendees	# collaborations within attendees (between new collaborators)	# collaborations within attendees (between old collaborators)
post			-0.426**
			(0.0531)
attended	2.190**	0.834**	-0.00196
	(0.239)	(0.150)	(0.0843)
post*attended			0.162*
			(0.0670)
ln(experience)	-0.894**	-0.367	-0.117
	(0.285)	(0.255)	(0.103)
ln(publications)	1.263**	1.279**	1.237**
	(0.0975)	(0.0779)	(0.0294)
ln(citations)	-0.0104	0.103+	-0.0142
	(0.0937)	(0.0530)	(0.0408)
ln(distance to conference)	-0.131+	-0.0638	-0.0451
	(0.0722)	(0.0758)	(0.0406)
_cons	-3.032**	-4.224**	-1.605**
	(0.990)	(0.980)	(0.345)
lnalpha			
_cons	0.853**	0.473**	-0.273**
	(0.186)	(0.125)	(0.0562)
conference fe	y	y	y
N	5075	3240	6480
Log lik.	-867.7	-1432.0	-7063.2

+ p<0.10, * p<0.05, ** p<0.01

Table 10A & B – These tables show simplified and difference-in-differences QML Poisson count regression models with panel random effects for the full sample with junior attendees in A and senior attendees in B. The dependent variables are respectively: collaborations between attendees with no prior collaborations within the conference in model 1, collaborations with new collaborators for attendees with prior collaborations within the conference in model 2, and collaborations between old collaborators 3. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.

Table 11A	Citations btw attendees	Citations btw att w/ no prior citation links w/in conference	Citations btw att w/ prior citation links w/in conference
Junior attendees	**Model 1** # citations within attendees	**Model 2** # citations within attendees	**Model 3** # citations within attendees
post	0.891**		0.788**
	(0.0819)		(0.0890)
attended	0.746**	1.621**	0.0260
	(0.151)	(0.222)	(0.151)
post*attended	0.271**		0.219
	(0.105)		(0.134)
ln(experience)	0.879**	0.0411	0.366**
	(0.0880)	(0.123)	(0.0984)
ln(publications)	0.315**	0.00921	0.327**
	(0.102)	(0.329)	(0.110)
ln(collaborations)	-0.0950	0.200	-0.114
	(0.112)	(0.383)	(0.120)
ln(distance to conference)	-0.131*	-0.208*	-0.0150
	(0.0592)	(0.100)	(0.0438)
_cons	-1.946**	-1.212	-0.525
	(0.554)	(4.125)	(0.847)
lnalpha			
_cons	1.558**	1.832**	-0.0671
	(0.0462)	(0.0881)	(0.0741)
conference fe	y	y	y
N	13540	4730	4080
Log lik.	-16350.6	-2681.4	-12606.0

+ p<0.10, * p<0.05, ** p<0.01

Table 11B	Citations btw attendees	Citations btw att w/ no prior citation links w/in conference	Citations btw att w/ prior citation links w/in conference
Senior attendees	**Model 1** # citations within attendees	**Model 2** # citations within attendees	**Model 3** # citations within attendees
post	0.381**		0.322**
	(0.0574)		(0.0605)
attended	0.720**	1.349**	0.196+
	(0.104)	(0.207)	(0.112)
post*attended	0.180*		0.190+
	(0.0763)		(0.101)
ln(experience)	0.624**	0.470	0.442**
	(0.174)	(0.385)	(0.126)
ln(publications)	-0.0152	-0.384	-0.00153
	(0.132)	(0.294)	(0.139)
ln(collaborations)	0.426**	0.620+	0.400*
	(0.146)	(0.335)	(0.158)
ln(distance to conference)	-0.0918**	-0.187*	-0.0149
	(0.0348)	(0.0922)	(0.0313)
_cons	-1.539*	-2.662*	-0.695+
	(0.645)	(1.211)	(0.388)
lnalpha			
_cons	1.265**	1.724**	0.146**
	(0.0388)	(0.103)	(0.0360)
conference fe	y	y	y
N	16630	4210	8210
Log lik.	-31411.8	-2517.4	-27527.1

+ p<0.10, * p<0.05, ** p<0.01

Table 11A & B – These tables show simplified and difference-in-differences QML Poisson count regression models with panel random effects for the full sample with junior attendees in A and senior attendees in B. The dependent variables are respectively: forward citations from attendees in model 1, forward citations from attendees with no prior citation links within the conference in model 2, and forward citations from attendees with prior citation links within the conference in model 3. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.

Table 12A	No prior collaborations w/in conference	Prior collaborations w/in conference	
Presenters	**Model 1**	**Model 2**	**Model 3**
	# collaborations within attendees	# collaborations within attendees (between new collaborators)	# collaborations within attendees (between old collaborators)
post			-0.499**
			(0.0603)
attended	2.375**	0.820**	-0.170*
	(0.231)	(0.147)	(0.0798)
post*attended			0.190*
			(0.0908)
ln(experience)	-0.289	0.0719	-0.243**
	(0.186)	(0.131)	(0.0765)
ln(publications)	1.258**	1.171**	1.230**
	(0.127)	(0.0889)	(0.0340)
ln(citations)	-0.0858	0.0516	-0.0249
	(0.0936)	(0.0620)	(0.0317)
ln(distance to	-0.110	-0.0878	-0.0679
conference)	(0.0785)	(0.0594)	(0.0451)
_cons	-4.699**	-4.865**	-1.005**
	(0.785)	(0.631)	(0.380)
lnalpha			
_cons	0.557	0.274+	-0.271**
	(0.351)	(0.158)	(0.0571)
conference fe	y	y	y
N	4060	2695	5390
Log lik.	-656.4	-1187.6	-5838.9

+ p<0.10, * p<0.05, ** p<0.01

Table 12B	No prior collaborations w/in conference	Prior collaborations w/in conference	
Non-presenters	**Model 1**	**Model 2**	**Model 3**
	# collaborations within attendees	# collaborations within attendees (between new collaborators)	# collaborations within attendees (between old collaborators)
post			-0.587**
			(0.0692)
attended	2.484**	0.973**	0.227**
	(0.233)	(0.237)	(0.0653)
post*attended			0.404**
			(0.0843)
ln(experience)	-0.574**	-0.222	-0.249**
	(0.127)	(0.165)	(0.0404)
ln(publications)	1.764**	1.484**	1.525**
	(0.124)	(0.0882)	(0.0355)
ln(citations)	-0.0607	0.196*	-0.0914**
	(0.0880)	(0.0835)	(0.0259)
ln(distance to	-0.155	0.182*	0.00544
conference)	(0.0960)	(0.0893)	(0.0300)
_cons	-5.124	-7.620	-1.642**
	(6.228)	(5.024)	(0.304)
lnalpha			
_cons	1.148**	0.972**	-0.609**
	(0.218)	(0.186)	(0.0655)
conference fe	y	y	y
N	5270	3060	6120
Log lik.	-733.7	-828.5	-5479.9

+ p<0.10, * p<0.05, ** p<0.01

Table 12A & B – These tables show simplified and difference-in-differences QML Poisson count regression models with panel random effects for the full sample with presenters in A and non-presenters in B. The dependent variables are respectively: collaborations between attendees with no prior collaborations within the conference in model 1, collaborations with new collaborators for attendees with prior collaborations within the conference in model 2, and collaborations between old collaborators 3. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.

Table 13A	Citations btw attendees	Citations btw att w/ no prior citation links w/in conference	Citations btw att w/ prior citation links w/in conference
Presenters	**Model 1** # citations within attendees	**Model 2** # citations within attendees	**Model 3** # citations within attendees
post	0.416**		0.350**
	(0.0776)		(0.0747)
attended	0.801**	1.496**	0.230+
	(0.127)	(0.265)	(0.125)
post*attended	0.168*		0.199*
	(0.0816)		(0.0985)
ln(experience)	0.429**	0.0112	0.175*
	(0.0889)	(0.129)	(0.0696)
ln(publications)	0.0351	-0.289	0.0563
	(0.148)	(0.417)	(0.125)
ln(collaborations)	0.400*	0.411	0.376**
	(0.170)	(0.458)	(0.145)
ln(distance to conference)	-0.143**	-0.163+	-0.0311
	(0.0364)	(0.0884)	(0.0309)
_cons	-0.766*	-1.327+	0.0697
	(0.368)	(0.769)	(0.306)
lnalpha			
_cons	1.184**	1.714**	0.0495
	(0.0555)	(0.144)	(0.0508)
conference fe	y	y	y
N	13510	3495	6520
Log lik.	-25888.1	-1916.7	-22899.4

+ p<0.10, * p<0.05, ** p<0.01

Table 13B	Citations btw attendees	Citations btw att w/ no prior citation links w/in conference	Citations btw att w/ prior citation links w/in conference
Non-Presenters	**Model 1** # citations within attendees	**Model 2** # citations within attendees	**Model 3** # citations within attendees
post	0.720**		0.632**
	(0.0894)		(0.0891)
attended	0.425**	1.486**	-0.159
	(0.156)	(0.177)	(0.111)
post*attended	0.269*		0.185
	(0.129)		(0.124)
ln(experience)	0.419**	0.126	0.0194
	(0.0675)	(0.101)	(0.0543)
ln(publications)	0.151	-0.146	0.168
	(0.103)	(0.289)	(0.109)
ln(collaborations)	0.0786	0.409	0.0435
	(0.117)	(0.302)	(0.124)
ln(distance to conference)	-0.0641	-0.189*	0.0178
	(0.0439)	(0.0897)	(0.0376)
_cons	-4.016**	-3.439**	-1.321**
	(0.564)	(1.001)	(0.422)
lnalpha			
_cons	1.521**	1.848**	0.0367
	(0.0445)	(0.0991)	(0.0515)
conference fe	y	y	y
N	16660	5445	5770
Log lik.	-22097.9	-3290.3	-17398.9

+ p<0.10, * p<0.05, ** p<0.01

Table 13A & B – These tables show simplified and difference-in-differences QML Poisson count regression models with panel random effects for the full sample with presenters in A and non-presenters in B. The dependent variables are respectively: forward citations from attendees in model 1, forward citations from attendees with no prior citation links within the conference in model 2, and forward citations from attendees with prior citation links within the conference in model 3. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.

APPENDIX A

Table A1	collaborations	collaborators	collaborations btw attendees	collaborators btw attendees
collaborations	1			
collaborators	0.8787	1		
collaborations btw attendees	0.2652	0.244	1	
collaborators btw attendees	0.1839	0.1935	0.7672	1

	forward cites	forward citors	forward cites btw attendees	forward citors btw attendees
forward cites	1			
forward citers	0.9332	1		
forward cites btw attendees	0.3411	0.3177	1	
forward citers btw attendees	0.3723	0.3608	0.8435	1

Table A1 – This table shows the correlation matrix between collaborations and collaborators, as well as collaborations and collaborators between attendees. Similarly for citations, it shows the correlation matrix between forward citations and citers, as well as forward citations and citers between attendees.

Table A2			Collaborations between attendees	
	Model 1	**Model 2**	**Model 3**	**Model 4**
	# collaborations	# collaborators	# collaborations within attendees	# collaborators within attendees
post	-0.0235**	0.147**	-0.335**	-0.275**
	(0.00589)	(0.0121)	(0.0413)	(0.0417)
attended	0.0534**	-0.00799	0.774**	0.820**
	(0.0116)	(0.0194)	(0.0677)	(0.0730)
post*attended	0.0132	0.0461*	0.342**	0.386**
	(0.00885)	(0.0185)	(0.0494)	(0.0546)
ln(experience)	0.0427**	0.0344*	-0.327**	-0.266**
	(0.00931)	(0.0140)	(0.0454)	(0.0440)
ln(publications)	1.371**	1.254**	1.525**	1.212**
	(0.00876)	(0.0108)	(0.0179)	(0.0253)
ln(citations)	0.00138	0.0549**	-0.0168	-0.0459*
	(0.00640)	(0.00916)	(0.0228)	(0.0210)
ln(distance to	0.00556	0.0424**	-0.116**	-0.114**
conference)	(0.00509)	(0.00638)	(0.0288)	(0.0258)
_cons	-1.313**	-0.223**	-2.016**	-2.098**
	(0.0558)	(0.0838)	(0.301)	(0.276)
lnalpha				
_cons	-2.284**	-1.523**	1.120**	0.951**
	(0.0647)	(0.0386)	(0.0404)	(0.0401)
conference fe	y	y	y	y
N	30170	30170	30170	30170
Log lik.	-38306.1	-76066.3	-15010.9	-14493.9

+ p<0.10, * p<0.05, ** p<0.01

Table A2 – This table shows difference-in-differences QML Poisson count regression models with panel random effects for the full sample. The dependent variables are respectively: overall collaborations and collaborators in models 1 and 2, and collaborations and collaborators between attendees in models 3 and 4. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.

Table A3			Citations between attendees	
	Model 1	**Model 2**	**Model 3**	**Model 4**
	# citations	#citers	# citations within attendees	#citers within attendees
post	0.288**	0.624**	0.560**	0.587**
	(0.0255)	(0.0266)	(0.0594)	(0.0290)
attended	-0.0343	-0.0745	0.707**	0.703**
	(0.0590)	(0.0636)	(0.116)	(0.0673)
post*attended	0.0457	0.0570	0.141	0.134**
	(0.0401)	(0.0400)	(0.0865)	(0.0424)
ln(experience)	0.830**	0.758**	0.519**	0.382**
	(0.0378)	(0.0355)	(0.0459)	(0.0358)
ln(publications)	-0.0220	-0.0854	0.0497	0.170*
	(0.0834)	(0.105)	(0.0974)	(0.0806)
ln(collaborations)	0.216*	0.321**	0.312**	0.0782
	(0.0907)	(0.111)	(0.120)	(0.0920)
ln(distance to conference)	-0.100**	-0.0985**	-0.120**	-0.102**
	(0.0189)	(0.0198)	(0.0304)	(0.0241)
_cons	1.089**	2.519**	-1.092**	-1.504**
	(0.220)	(0.190)	(0.316)	(0.262)
lnalpha				
_cons	0.433**	0.569**	1.396**	1.096**
	(0.0282)	(0.0327)	(0.0332)	(0.0343)
conference fe	y	y	y	y
N	30170	30170	30170	30170
Log lik.	-98717.3	-293033.6	-48327.4	-26523.6

+ p<0.10, * p<0.05, ** p<0.01

Table A3 – This table shows difference-in-differences and simplified QML Poisson count regression models with panel random effects for the full sample. The dependent variables are respectively: overall forward citations and citers in models 1 and 2, and forward citations and citers between attendees in models 3 and 4. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.

APPENDIX B

Table B1	Collaborations within attendees			
	Model 1	**Model 2**	**Model 3**	**Model 4**
	# collaborations within attendees (w att-mat controls)	# collaborations within attendees (w mat-mat controls)	# collaborators within attendees (w att-mat controls)	# collaborators within attendees (w mat-mat controls)
post	-0.335**	-0.682**	-0.275**	-0.731**
	(0.0408)	(0.0584)	(0.0404)	(0.0431)
attended	0.774**	0.842**	0.820**	0.930**
	(0.0616)	(0.0644)	(0.0610)	(0.0543)
post*attended	0.342**	0.692**	0.386**	0.846**
	(0.0459)	(0.0640)	(0.0449)	(0.0564)
ln(experience)	-0.327**	-0.268**	-0.266**	-0.218**
	(0.0483)	(0.0417)	(0.0471)	(0.0369)
ln(publications)	1.525**	1.495**	1.212**	1.201**
	(0.0259)	(0.0248)	(0.0239)	(0.0267)
ln(citations)	-0.0168	-0.0162	-0.0459*	-0.0519**
	(0.0252)	(0.0269)	(0.0194)	(0.0192)
ln(distance to	-0.116**	0.00439	-0.114**	-0.0379
conference)	(0.0266)	(0.0284)	(0.0258)	(0.0238)
_cons	-2.016**	-3.064**	-2.098**	-3.049**
	(0.358)	(0.325)	(0.343)	(0.249)
lnalpha				
_cons	1.120**	1.022**	0.951**	0.741**
	(0.0379)	(0.0388)	(0.0466)	(0.0426)
conference fe	y	y	y	y
N	30170	30170	30170	30170
Log lik.	-15010.9	-14648.1	-14493.9	-13665.9

+ p<0.10, * p<0.05, ** p<0.01

Table B1 – This table shows difference-in-differences and simplified QML Poisson count regression models with panel random effects for the full sample. The dependent variables are respectively: collaborations between attendees with attended-matched controls in models 1 and matched-matched controls in model 2, and collaborators between attendees with attended-matched controls in model 3 and matched-matched controls in model 4. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.

Table B2	Citations between attendees			
	Model 1	**Model 2**	**Model 3**	**Model 4**
	# citations within attendees (w att-mat controls)	# citations within attendees (w mat-mat controls)	# citers within attendees (w att-mat controls)	# citers within attendees (w mat-mat controls)
post	0.560**	0.112+	0.587**	0.137**
	(0.0642)	(0.0579)	(0.0309)	(0.0368)
attended	0.707**	0.871**	0.703**	0.959**
	(0.0820)	(0.0855)	(0.0688)	(0.0576)
post*attended	0.141+	0.587**	0.134**	0.580**
	(0.0818)	(0.0874)	(0.0425)	(0.0475)
ln(experience)	0.519**	0.528**	0.382**	0.393**
	(0.0464)	(0.0399)	(0.0490)	(0.0420)
ln(publications)	0.0497	0.0899	0.170*	0.216*
	(0.0992)	(0.107)	(0.0729)	(0.0968)
ln(collaborations)	0.312**	0.295*	0.0782	0.0808
	(0.121)	(0.122)	(0.0818)	(0.112)
ln(distance to conference)	-0.120**	-0.0882**	-0.102**	-0.0639**
	(0.0325)	(0.0334)	(0.0272)	(0.0203)
_cons	-1.092**	-1.606**	-1.504**	-2.098**
	(0.284)	(0.328)	(0.251)	(0.278)
lnalpha				
_cons	1.396**	1.288**	1.096**	0.841**
	(0.0303)	(0.0335)	(0.0338)	(0.0343)
conference fe	y	y	y	y
N	30170	30170	30170	30170
Log lik.	-48327.4	-44779.2	-26523.6	-24274.9

+ p<0.10, * p<0.05, ** p<0.01

Table B2 – This table shows difference-in-differences and simplified QML Poisson count regression models with panel random effects for the full sample. The dependent variables are respectively: forward citation from attendees with attended-matched controls in models 1 and matched-matched controls in model 2, and forward citers from attendees with attended-matched controls in model 3 and matched-matched controls in model 4. Robust standard errors are in parentheses with clustering on the individual. Conference fixed effects are also included.