Scientometrics
https://doi.org/10.1007/s11192-025-05291-5

Air travel and research collaboration: a quasi-experimental insight

Adam Ploszaj[1]

Received: 31 May 2024 / Accepted: 14 March 2025

Abstract
This article analyzes the impact of the availability of long-haul flights on international scientific collaboration. The background assumption is that a wider availability of long-distance flights enables greater mobility for scientists, which in turn increases the likelihood of long-distance research collaboration. The analytical framework of this article is based on a quasi-experimental approach. Specifically, a discontinuity in the global network of air routes due to regulatory requirements is used as a source of random selection—i.e., a natural experiment. Two complementary methods are used to model the effect: instrumental variable design and regression discontinuity. Combining the bibliometric data from Microsoft Academic Graph and flight data from the International Civil Aviation Organization, this paper provides evidence of a causal relation between air transport availability and research collaboration. The results show that direct long-haul flights positively impact the number of papers co-authored by scholars based in both the connected cities. The implications of these results for science policy are discussed in the context of the virtualization of scientific communication resulting from the COVID-19 pandemic and air transport's negative climate impacts.

Keywords Research collaboration · Academic mobility · Air travel · Natural experiment · Quasi experiment

Introduction

As a social practice, contemporary science can be characterized by widespread collaboration and internationalization. Numerous studies have documented the development of international research collaboration, academic mobility, and knowledge flows, as well as their patterns, causes, and consequences (Olechnicka et al., 2019). The dominant view is that scientific mobility is beneficial for individual scientists and their careers, as well as for scientific institutions and general advancement (Hall et al., 2018; Jacob & Meek, 2013; Leahey, 2016; Netz et al., 2020; Petersen, 2018). This topic is important not only from

✉ Adam Ploszaj
a.ploszaj@uw.edu.pl

[1] Science Studies Lab & Centre for European Regional and Local Studies EUROREG, University of Warsaw, Warsaw, Poland

Published online: 03 April 2025

a theoretical point of view but also from the perspective of scientific and, more broadly, development policy. Many institutions and countries have specific policies supporting scientific mobility based on the assumption that it can yield positive results, including such iconic initiatives as the Fulbright in the United States (Arndt, 1996) and the Marie Skłodowska-Curie Actions in the European Union (Souto-Otero, M et al., 2017).

At the same time—as in the case of many, if not all, socioeconomic issues—conducting a strict causal analysis of the effects of scientific mobility is highly complicated, or even impossible. Indeed, it is difficult to imagine an experimental study in which fate alone decides which scientist obtains a scholarship to go abroad and which stays in their own country. Cost would make such an experiment impossible, not to mention ethical issues. This leaves two possibilities. The first is observational research, which by definition prohibits the confirmation or rejection of causal hypotheses. The second option is a quasi-experimental approach, which does not involve an experiment in the strict sense. Instead, appropriate techniques are applied to suitable data and specific assumptions are fulfilled, allowing conclusions to be drawn about cause and effect from a given analysis (Campbell & Stanley, 2011). The quasi-experimental approach is often based on the occurrence of certain unique circumstances that introduce a random element and, consequently, divide the examined sample into an experimental and a control group. In such a situation, we speak of natural experiments. A natural experiment differs from a proper experiment: in the latter, the random element is introduced as part of the research procedure, while in the former the random factor is outside the researchers' control (which also cleverly removes all problematic legal and ethical issues). However, in practice, the use of natural experiments is limited because finding an appropriate quasi-experimental situation is more difficult than one might expect (Dunning, 2012; Sekhon & Titiunik, 2012).

This article uses a highly specific quasi-experimental situation to analyze the impact of long-distance mobility on research collaboration. In particular, the article analyzes how the availability of long-distance (generally speaking: intercontinental) air connections translates into scientific collaboration. An important assumption here is that the physical possibility of travel materializes in the form of traveling scholars (scientific mobility). Accordingly, scientists can meet and collaborate, leading to measurable effects in the form of co-authored publications. The key element that translates into materialized collaboration is thus not the availability of transport infrastructure but rather the face-to-face meetings it enables.

Theoretically, this study is embedded in two key concepts explaining the spatial aspects of research collaboration. The first is the importance of spatial distance for establishing, maintaining, and conducting scientific collaboration. The basic rule in this case is that the likelihood of collaboration declines as the distance between prospective collaborators increases. This thesis finds support in a large body of published research at the micro (e.g., collaboration likelihood within buildings or campuses), meso (e.g., cities, regions, countries), and macro (e.g., global) scales (Allen, 2007; Fernández et al., 2016; Kabo et al., 2014; Katz & Martin, 1997). The second concept embeds research collaboration in the gravity model borrowed from physics, adding a mass variable to the distance variable. Here, mass is typically interpreted as research capacities, measured, for instance, using the volume of published papers, submitted patents, received grants, or research workforce. In effect, the basic gravity model theoretical framework can be written as follows: The probability and intensity of research collaboration are negatively associated with geographical distance, which separates the units in question, and positively affected by their accumulated research potential. Again, numerous studies have used this approach (See e.g.: Andersson & Persson, 1993; Hoekman et al., 2009; Picci, 2010; Plotnikova & Rake, 2014; Sebestyén

& Varga, 2013). This article uses the gravity framework, while deepening and improving our understanding of the phenomenon under study by employing a quasi-experimental design and including variables describing actual long-distance transport accessibility as a factor enabling scientific mobility, which in turn impacts research collaboration.

The analyses presented in this article seek to test the hypothesis that greater availability of long-distance air connections positively influences long-distance research collaboration. A key aspect of the hypothesis is the centrality of a variable that is not directly observed, namely the mobility of scientists. The existence of air connections should not (based on existing theories) directly translate into research collaboration. However, the availability of air connections facilitates this mobility in the short term, such as for conference trips or project meetings, as well as in the medium and long term, such as internships, study visits, and international fellowships. The mobility of scientists enables face-to-face meetings, which significantly increases the chance of establishing collaboration (Boudreau et al., 2017; Chai & Freeman, 2019; Duede et al., 2024; Jin et al., 2024). As such, the study's conceptual framework can be presented as follows: (1) Better availability of long-distance flights (observed variable) influences (2) greater mobility of scientists (unobserved variable), which in turn increases the likelihood of (3) long-distance research collaboration (observed variable). Scientific collaboration is operationalized as the co-authorship of scientific publications. Of course, this is not the only form or indication of scientific collaboration, but it is a commonly used measure due to both the availability and large volume of data, which facilitates statistical analysis by providing large samples.

Prior work

This study fits directly into the field of transport accessibility and its effects on academic mobility, research collaboration, and scientists' productivity. It is worth noting that few analyses of mobility and scientific collaboration consider transport accessibility, especially given the hundreds of publications devoted to the topic of collaboration in science. Even when included, spatial separation is often simply measured as geographical distance along the surface of the earth (or colloquially: "as the crow flies") (Frenken et al., 2009a, 2009b). Furthermore, while a handful of studies have accounted for transport accessibility, they have typically been descriptive (i.e., non-causal) and focused on specific cases, such as countries, scientific fields, or groups of institutions. Road travel time has been used to explain co-patenting in Sweden (Andersson & Ejermo, 2005), with a subsequent study using both road travel time and air travel time (Ejermo & Karlsson, 2006). Road travel time has also been used when analyzing scholarly co-authorship networks in the Netherlands (Frenken, Hoekman, et al., 2009). Chinese authors have hypothesized that high-speed railway accessibility may be one factor explaining the intensity of research collaboration between cities in China (Ma et al., 2014). European data suggest that regions with major international airports are likely to develop intensive international scientific collaboration (Hoekman et al., 2010). A cross-sectional study of four large public US universities found that more flight connections (connectivity) and closer airport proximity (accessibility) increase the expected number of co-authored scientific papers (Ploszaj et al., 2020). Global flight network data have been used to analyze variation in citations of collaborative research (Naik et al., 2023). A sample of scholars from the University of British Columbia was used to investigate the influence of career stage, research productivity, field of expertise, and other variables on academic air travel and the associated emissions (Wynes et al., 2019). Moreover, it has been found that universities with better connectivity via air transport networks tend to be ranked higher (Guo et al., 2017). Among this research, several papers stand out due to their use of a quasi-experimental approach. Catalini et al. (2020) used a quasi-experimental design to examine the impact of a new, low-fare air route on the probability of research collaboration. However, their research was limited in terms of space and field, as they focused on United States-based chemistry scholars. Bahar et al. (2023) analyzed relationships between direct long-haul flights and collaborative patent applications. Dong et al. (2020) employed an instrumental variable approach to study the impact of high-speed railways on scientific collaboration. Yao and Li (2022) and Kand et al. (2023) analyzed to impact of Chinese high-speed rail on co-patenting. Similarly, Koh et al. (2024) showed that the development of the Begin subway impacted collaborative patents on an intra-urban scale. Hu et al. (2022), using a difference-in-differences framework, provide evidence that nonstop flights between China and the U.S. significantly enhance the production of highly cited joint papers. However, most of these findings are limited to single country, specifically US or China, restricting their broader applicability. Among the above-mentioned quasi-experimental studies, only Bahar et al. (2023) examine the global scale; however, their focus is on the impact of air availability on collaborative patents. In contrast, the present study addresses a clear research gap by investigating the causal relationship between long-haul air accessibility and research collaboration, operationalized through co-authored scientific publications.

Data and methods

This study followed the natural experiment identified by Campante and Yanagizawa-Drott (2018).They noted that cities that are just under 6000 miles apart are distinctly more likely to have direct air links, especially compared with those slightly over the 6000-mile threshold. They then determined that this effect did not correlate with any objective factors related to geography, geopolitics, economy, or aviation technology, but was rather due to an arbitrary administrative decision. In accordance with aviation regulations in force in many countries from the 1980s to the second decade of the twenty-first century, airlines were required to provide a double crew for flights scheduled to last longer than 12 h, including pilots and all other crew members. This 12-h limit was set arbitrarily. In practice, this meant that, for flights scheduled to last even 1 min longer than 12 h, the airline had to pay for a double crew and provide a space where additional crew members could rest. Therefore, exceeding the 12-h flight time limit resulted in a sudden increase in crew costs. This cost factor presumably deterred airlines from launching such flights, an assumption that was confirmed by Campante and Yanagizawa-Drott (2018). Using historical data, they identified a clear discontinuity in the global air travel network exactly when expected: at flight lengths of 12 h. Furthermore, the authors noted that a 12-h flight translates to a distance of roughly 6000 miles. This is an important observation because while data on flight distance are readily available, data on flight duration are more difficult to obtain. Substituting distance data for flight length data again shows the expected regularity. There are significantly more connections between airport pairs just below 6000 miles than just above this threshold. This jump is difficult to explain other than by the above-described regulations. Since 12 h to 6000 miles was arbitrarily determined, this situation can be used as a natural experiment. Campante and Yanagizawa-Drott (2018) used this natural experiment to investigate the causal relationship between long-haul air flights and economic development (using night lights satellite data as a proxy for economic development) and business links (using international firm ownership data). In this paper, I applied Campante and Yanagizawa-Drott's (2018) natural experiment framework to other data, thereby not only examining the phenomenon of scientific mobility and research collaboration but also testing the natural experiment from a different angle.

An important aspect of this study's methodology is its temporal dimension. The main analyses were conducted for 2014. Far from being arbitrary, this choice related directly to the type of natural experiment used. Air transport regulations that were the basis for the discussed discontinuity have changed since 2014. New regulations have been introduced, particularly in the United States (known as FAR 117) and the European Union (known as EU FTL). As such, the 12-h (or 6000-mile) threshold became significantly less pronounced after 2014. Consequently, 2014 is the last year for which the discussed natural experiment can be realistically used for quasi-experimental analyses. On the other hand, before 2014, the discontinuity had only occurred for 20-odd years. As Campante and Yanagizawa-Drott (2018) showed for 1989, despite the same aviation regulations being in force, the 12-h limit had no significant impact on the air transport network. This was because, prior to the early 1990s, there were a limited number of aircraft available that could cost-effectively fly such long routes.

This paper used two main data sources. First, the International Civil Aviation Organization's proprietary datasets, "Traffic by Flight Stage" and "Airport Traffic," which contain data on individual air operations in international air traffic (not only on the completion of the flight but also on the filling of seats available on a given flight, cargo volume, etc.), as well as data enabling the characterization of international airports (number of operations, number of passengers, etc.). The second source was Microsoft Academic Graph (MAG), version 2020-09-01, as hosted by the Collaborative Archive & Data Research Environment project of the Indiana University Network Science Institute. Based on MAG data, two dependent variables were calculated for this study (see below). The use of MAG as a source of bibliometric data was motivated, firstly, by its open-access availability, and secondly, by the fact that the affiliation data in MAG were geolocated, i.e., provided with information on the longitude and latitude of a given institution (Sinha et al., 2015; Wang et al., 2020). These two features clearly distinguished MAG from two other databases often used in bibliometric analyses: the Web of Science and Scopus. Both these databases are distributed on a commercial basis, and neither contain data on latitude and longitude.

Two dependent variables were used in this study. The first was defined as the number of co-authored publications written by authors affiliated with institutions assigned to the nearest international airport offering regular international flights (if there was more than one airport in a given metropolitan area, their data were aggregated). In other words, the variable indicates how many joint articles can be assigned to each pair of international airports. For papers published in 2014, this variable had a mean of 15.5 and assumed a value between 0 and 17,360. The latter value is the number of co-authored papers between institutions assigned to the Boston airport and institutions assigned to airports serving New York. Note that this variable was not normally distributed. The value of 0 (no co-authored articles) applied to as many as 80% of cases, while 1–3 co-authored papers constituted 8.1% of the sample, 4–6 constituted 2.3%, 7–9 constituted 1.2%, and 10 or more constituted the remaining 8.4%.

The second dependent variable was defined as the percentage of international publications—publications with at least one author affiliated to an institution located in a different country divided by the total number of publications at a given affiliation. This variable was calculated for institutions spatially aggregated to spatial grid cells with a size of

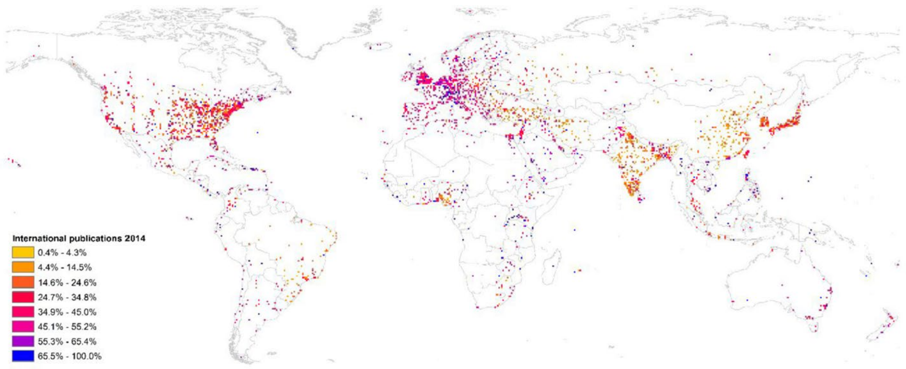

Fig. 1 Share of internationally co-authored publications in grid cells of 0.25×0.25° for papers published in 2014

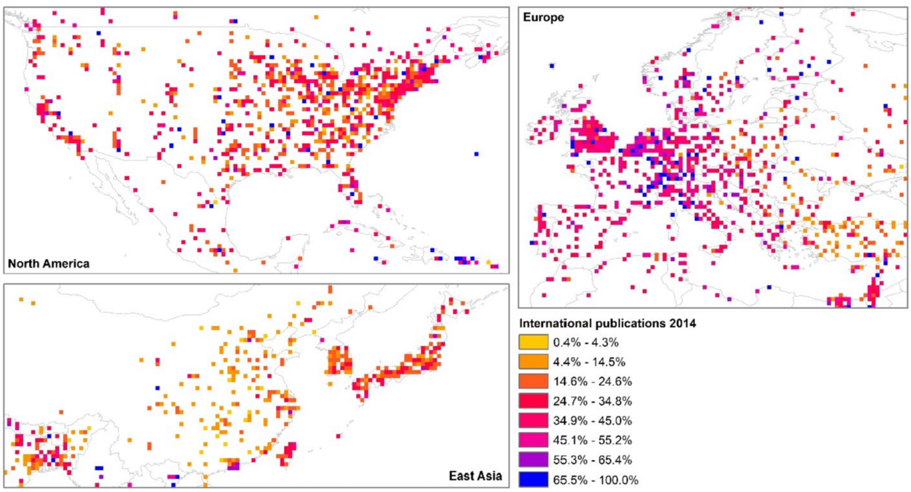

Fig. 2 Share of internationally co-authored publication in grid cells of 0.25×0.25° for papers published in 2014: close-up of selected areas

0.25×0.25°, which translates to a geographical space measuring approximately 27.5 km by 27.5 km. Data aggregated at the grid-cell level allows for clear presentation on a map (see Figs. 1 and 2) and facilitates the modeling procedure using the distance variable of the grid-cell centroid to the nearest airport. Since research activity is concentrated in a few places on Earth, it is unsurprising that, for 98.1% of grid cells with a size of 0.25×0.25°, the value of the variable is equal to 0. This is primarily due not to the absence of international articles assigned to them but rather to the absence of affiliated articles in these locations. For grid cells with at least one affiliated article published in 2014, the variable ranges from 0 to 100%, with an average of around 30%.

The study used two complementary quasi-experimental methods: regression discontinuity (RD) design and instrumental variable (IV) design. Both methods build upon regression analysis, leveraging specific characteristics of the studied phenomenon to approximate causal inference (Angrist & Pischke, 2009). A key aspect of this approach is the presence of a naturally occurring condition that introduces an element of randomness, effectively dividing the sample into a group exposed to a given influence (treatment group) and a group that is not (control group). This structure mirrors a traditional experiment in which random assignment underpins causal inference. In this analysis, the critical characteristic is the division of airport pairs into those just below and just above the 6000-mile threshold. As demonstrated above, this arbitrarily defined threshold significantly affects the probability of direct air connections between airports. By exploiting this exogenous variation, this analytical framework enables a robust examination of how long-haul air accessibility influences research collaboration across distant locations. The use of both RD and IV in this study allowed the use of two dependent variables, complementary, but with very different characteristics: network-pair data in the case of RD and grid-cell data in the case of IV. RD design seems more straightforward in its interpretation—especially since the result can be presented using striking and easy-to-read charts showing how values change in relation to the threshold. On the other hand, the use of instrumental variable design facilitated complementary analyses showing the spatial extent of the airport's impact.

The specification of the models largely followed Campate and Yanagizawa-Drott's (2018) proposed framework. The RD model was formulated as follows:

$$Y_{ij} = \alpha + \beta * Below6K_{ij} + g\left(d_{ij}\right) * \gamma + \varepsilon_{ij},$$

where Y_{ij} denotes the dependent variable, defined as the number of publications by co-authors affiliated with institutions located in places assigned to individual "airport pairs." $Below6K_{ij}$ is a discontinuity dummy equal to 1 if d_{ij} is less than 6000 miles. The proposed specification allowed the inclusion of control variables and various slopes above and below the 6000-mile threshold, as well as to model nonlinearities (by inclusion of polynomials).

In turn, the instrumental variable model was formulated as follows:

$$Y_{ic} = \alpha + \beta * ShareEIBelow6K_{ic} + X_{ic}\gamma + \varepsilon_{ic},$$

where Y_{ic} denotes the dependent variable of the percentage of publications written in international collaboration, defined at the grid-cell level (see Figs. 1 and 2). $ShareEIBelow6K_{ic}$ is a composite instrumental variable invented and calculated by Campante and Yanagizawa-Drott (2018), which they defined as the sum of the eigenvector centrality of airports 5,500 to 6000 miles away from airport *i*, divided by the sum of the eigenvector centrality of airports 5,500 to 6,500 miles away. The use of eigenvector centrality in the calculation of the variable made it possible to take into account not only the direct availability of long-haul flights but also the availability of connecting flights.

Finally, to account for the heterogeneity of the studied places and airports in the estimated models, a number of control variables were used, again in accordance with Campante and Yanagizawa-Drott (2018). The following control variables were included: total number of flights; number of daily, twice weekly, and weekly flights; total number of passengers and seats; number of connected cities and countries; time zone; distance to the equator; and GDP per capita at the country level. In the below model specifications, these variables are referred to as "airport controls." In addition, this study introduced a control variable specific to its context: "scientific capacity". In IV models, it is defined as the number of publications assigned to a given location, while in RD models, it is measured as the product of the number of publications assigned to airport areas i and j, Adding control variables to the RD model requires that they satisfy the covariate continuity assumption, meaning there should be no discontinuities in the covariates at the threshold of the assignment variable. The fulfillment of this assumption for airport controls was verified by Campante and Yanagizawa-Drott (2018). For the scientific capacity variable, the absence of discontinuities at the key cutoff point is illustrated in a graph included in the online appendix.

Results

Figure 3 illustrates the main results of this analysis and shows the number of co-authored publications by distance between airports. For the sake of readability, the number of articles has been aggregated into bins covering 100 miles each. The vertical line in the middle of the chart denotes the presumed cutoff threshold of 6000 miles. The results clearly demonstrate the significance of this threshold. Not only does exceeding the 6000-mile limit result in a significant drop in the number of co-authored publications, but the trend lines on both sides of the cutoff threshold are completely different. As such, the figure provides a strong basis for expecting that the anticipated effect (hypothesis of the paper) is supported by the data. The modeling results further confirm this expectation.

All six specifications of the RD model produced statistically significant results (Table 1). Depending on the specification, exceeding the 6000-mile limit resulted in reductions of between 2.6 and eight in the expected number of co-authored publications. Models (1) and (2) used an arbitrarily selected bandwidth (i.e., the distance to the right and left of the cutoff threshold), which determined the data included in the model. This was 500 miles in Model (1) and 1000 miles in Model (2). Subsequent models used the so-called optimal bandwidth (i.e., the bandwidth determined algorithmically by the model implemented in STATA based on given data characteristics). The use of alternative bandwidths is a good practice in RD modeling and a method of ensuring the robustness of the results (Imbens & Lemieux, 2008). In this case, the first three models differing only in bandwidth width generated similar RD estimate values and standard errors, which significantly increased the reliability of the obtained results. The remaining specifications (4–6) added control variables characterizing the studied airports to the model. Specifications (5) and (6) added a

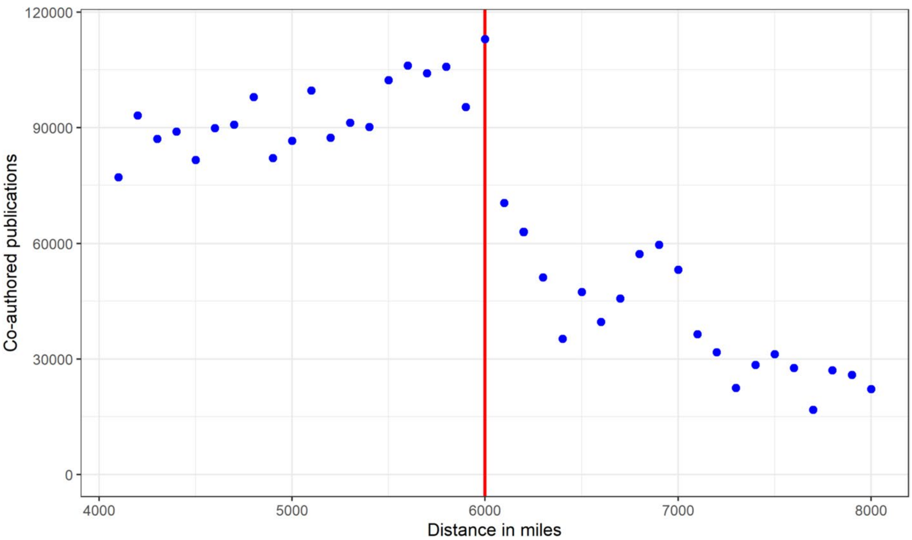

Fig. 3 Number of co-authored publications by distance between closest airports in 2014

Table 1 Effect on the number of co-authored publications, airport-pair level, in 2014

Dependent variable: number of co-authored publications						
Specification	(1)	(2)	(3)	(4)	(5)	(6)
Bandwidth in miles	500 miles	1000 miles	Optimal	Optimal	Optimal	Optimal
RD estimate	− 6.13	− 8.03	− 7.98	− 6.47	− 2.57	− 3.24
Robust standard errors	(2.58)**	(1.75)***	(1.77)***	(1.51)***	(1.10)**	(1.33)**
Covariates:						
Airport controls				✓	✓	✓
Scientific capacity					✓	✓
Polynomial order	1st	1st	1st	1st	1st	2nd
Observations: total	334,954	334,954	334,954	334,954	334,954	334,954
Observations: effective	41,477	81,842	80,540	87,562	86,027	124,884

Significance levels: $^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.1$

scientific capacity variable defined as the product of the number of publications assigned to airport areas *i* and *j*. The addition of the scientific capacity variable is consistent with the gravity model theoretical approach. The last specification (6) introduced a second-order polynomial to enable the modeling of curvilinear trends on both sides of the cutoff threshold. It should be emphasized that adding the scientific capacity variable clearly reduced the RD estimate value to − 2.6 in the case of specification (5) and − 3.2 in the case of specification (6), while still remaining statistically significant.

As explained above, the 6000-mile discontinuity became visible in the 1990s. In 1998, this discontinuity appeared practically nonexistent. Therefore, it could be expected that the effect observed in the 2014 data would be absent in 1989. The collected data confirmed this intuition to a certain extent. Figure 4 shows that, while there is a decline in the number of co-authored papers after crossing the 6000-mile threshold, it is also clear that this is part of a broader trend, wherein the number of co-authored publications declines with increasing distance. If there is a discontinuity here, it is certainly less pronounced compared to the case of the 2014 data.

The modeling results confirm the conclusions drawn from the observations in Fig. 2. Specifications (1–4), those that do not account for the scientific capacity variable, provide a statistically significant result, indicating the influence of the 6000-mile discontinuity. However, the full specifications (5) and (6), which do account for the scientific capacity (according to the gravity model) deliver a statistically insignificant result (see Table 2). This result aligns with expectations and ultimately strengthens the conclusion of the overall analysis. Since there was no discontinuity in the air traffic network at the 6000-mile point in 1989, it should not affect other changes. However, in 2014 the discontinuity at the same point was fully developed and could thus affect other variables—in our case, co-authored publications.

The second part of the analysis using the instrumental variable framework largely confirmed the results of the RD analysis. A positive estimate of the instrumental variable would indicate that, the better the developed network of long-distance connections, the greater the percentage of international co-authored publications (see Table 3). According to the assumptions of the gravity model, scientific capacity, calculated as the total number of scientific publications assigned to a given grid cell, was ultimately

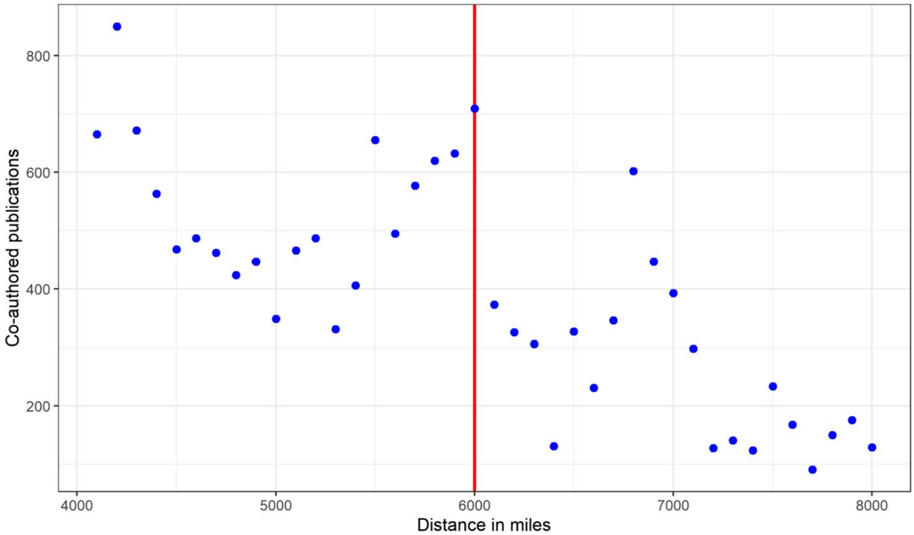

Fig. 4 Number of co-authored publications by distance between closest airports in 1989

Table 2 Effect on the number of co-authored publications, airport-pair level, in 1989

Dependent variable: number of co-authored publications						
Specification	(1)	(2)	(3)	(4)	(5)	(6)
Bandwidth	500 miles	1000 miles	Optimal	Optimal	Optimal	Optimal
RD estimate	− 0.065	− 0.078	− 0.074	− 0.060	− 0.014	− 0.018
Robust standard errors	(0.027)**	(0.020)***	(0.021)***	(0.019)***	(0.012)	(0.016)
Covariates:						
Airport controls				✓	✓	✓
Scientific capacity					✓	✓
Polynomial order	1st	1st	1st	1st	1st	2nd
Observations: total	334,954	334,954	334,954	334,954	334,954	334,954
Observations: effective	41,477	81,842	70,745	76,155	109,765	109,765

Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

a statistically significant variable. It is worth noting that adding baseline data on the percentage of international publications in 1989 (i.e., before the discontinuity occurred) to the model significantly increases its ability to explain the variance of the dependent variable. The R-squared value in specification (2) in Table 3 is 0.27, which is significantly higher than in the case of specification (1), which does not include this baseline variable.

The specifications presented in Table 4 show the extent of international airports' influence on international research collaboration. Specification (1) in Table 4 uses the interaction between the instrumental and distance variables of a given grid cell to the nearest international airport within the same country. A negative estimate means that, as

Table 3 Effect on the share of internationally co-authored publications, grid-cell level, in 2014

Dependent variable: share of internationally co-authored publications (%)		
Specification	(1)	(2)
ShareEI_below6000_w500	0.15***	0.12***
	(0.05)	(0.04)
Scientific capacity	0.19***	0.13***
	(0.04)	(0.03)
Co-authored publications in 1989		0.36***
		(0.03)
Airport controls	✓	✓
Observations	37,812	37,812
R-squared	0.15	0.27

Significance levels: ***p < 0.01, **p < 0.05, *p < 0.1

the distance increases, the airport's positive impact gradually decreases. In turn, specifications (2–7) were calculated for subsamples determined according to the distance from the airport: grid cells located up to 50 miles from the airport, then located at a distance from 50 to 100 miles, then located at a distance of 100 miles to 150 miles, etc., with an increment of 50 miles up to a value of 300 miles. The obtained results clearly show that the influence of the airport is highest in its vicinity. The estimate of the instrumental variable for the subsample below 50 miles was three times higher than for the 50–100 and 100–150-mile subsamples. Moreover, the result for the subsample below 50 miles was more statistically significant than those for the next two intervals. After crossing the 150-mile limit, the instrumental variable was no longer statistically significant, meaning that the analyzed impact reached no farther than 150 miles from the airport.

Discussion and conclusions

The analyses presented in this article support the hypothesis that greater availability of long-distance air connections positively affects long-distance research collaboration. The effect was observed and statistically significant for both tested dependent variables and both quasi-experimental methods. The research results are consistent with the findings of previous works examining the relationship between air accessibility and scientific collaboration (Andersson & Ejermo, 2005; Catalini et al., 2020; Hoekman et al., 2010; Ploszaj et al., 2020) or, more broadly, between transport accessibility—also accounting for additional means of transport—and scientific collaboration (Dong et al., 2020; Ejermo & Karlsson, 2006; Ma et al., 2014). Against this background, the main contribution of this paper is the application of a quasi-experimental method to global data. This is a significant extension of the only paper to date to have employed a quasi-experimental approach to study the relationship between air accessibility and scientific collaboration—a paper examining the impact of the introduction of low-cost airline flights on collaboration among chemistry scholars in the United States (Catalini et al., 2020).

The availability of long-distance air connections is neither the main nor the decisive factor in initiating and conducting fruitful research collaboration (Hall et al., 2018; Katz & Martin, 1997; Leahey, 2016; Sonnenwald, 2007). In the broadest sense, an appropriate

Table 4 Effect on the share of internationally co-authored publications, spatial patterns, grid-cell level, in 2014

Dependent variable: share of internationally co-authored publications (%)							
Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Subsample: miles from closest airport	< 500	< 50	50–100	100–150	150–200	200–250	250–300
ShareEI_below6000_w500	1.28***	0.45***	0.15**	0.16**	0.03	0.01	0.01
	(0.46)	(0.15)	(0.07)	(0.07)	(0.04)	(0.03)	(0.02)
ShareEI_below6000 * Distance to airport (interaction)	− 0.23*** (0.09)						
Covariates: airport controls, scientific capacity, co-authored publications in 1989	✓	✓	✓	✓	✓	✓	✓
Observations	157,341	14,480	23,332	23,206	20,717	17,615	15,647
R-squared	0.23	0.31	0.20	0.17	0.11	0.15	0.27

Significance levels: ***p < 0.01, **p < 0.05, *p < 0.1

capacity for research collaboration is needed. The results presented in this paper confirm this assumption. Model specifications that account for the scientific potential of collaborators (conceptualized here simply as the number of publications) explain a greater part of the variability of the outcome variables than models that do not take this variable into account. This is also consistent with, and expected on the basis of, the gravity model theory. Moreover, in models that account for scientific capacity, the estimates of the impact of air accessibility are lower and/or less statistically significant than in the case of models that neglect this variable. However, it should be emphasized that air accessibility variables remain statistically significant.

The results clearly show that the positive impact of an international airport offering long-distance connections is spatially limited: the impact is greatest near the airport, gradually decreases with increasing distance, and becomes negligible once the distance exceeds 150 miles. This result is consistent with the results of previous (non-quasi-experimental) studies (Ploszaj et al., 2020).

In a broad sense, this study fits into the discussion on the relationship between transport infrastructure and spatial connections, or socioeconomic development in general (Crescenzi & Rodríguez-Pose, 2012). Views on this issue range between two extremes. On the one hand, it is said that transport infrastructure influences the emergence of flows and connections. On the other, it has been argued that infrastructure should be created where these connections already exist. The results presented in this paper provide evidence that transport infrastructure can translate into increased collaboration, even over significant distances. At the same time, it should be emphasized that these results do not mean that there is no reverse relationship, that is, that existing relationships (including scientific collaboration) stimulate the launch of new air connections. Demand is, of course, an important factor in the supply of air services (Wang & Gao, 2021). However, this issue goes beyond the scope of this study and is impossible to investigate using the natural experiment implemented here.

An important limitation of this study is its inherently historical nature. Two aspects are worth mentioning here. Firstly, due to the abovementioned changes in aviation regulations, discontinuity (i.e., the basis for the natural experiment) declined after 2014. Accordingly, it would be impossible to repeat these analyses to obtain more current data. Secondly, the circumstances of scientific collaboration have changed drastically as a result of the COVID-19 pandemic. Remote collaboration tools have improved significantly and become common, causing, among other things, the (perhaps temporary) spread of virtual scientific conferences (Falk & Hagsten, 2021; Olechnicka et al., 2024, 2025). It is difficult to predict the extent to which virtual contacts will replace the need for scientific mobility. However, it should certainly be stated that new research on the relationship between air accessibility and scientific collaboration using data from the new, post-pandemic reality, is necessary. Moreover, such research should also endeavor to include virtual communication as an important covariate.

Future analyses of the impact of transport accessibility on research collaboration should consider the potential crowding-out effect, where the development of one type of collaboration may come at the expense of another. A study by Hu et al. (2022) examines whether increased international nonstop flights affect domestic and non-U.S. collaborations in China. Their findings suggest that the introduction of U.S.-China nonstop flights led to a shift in collaboration patterns, although the evidence for a strong crowding-out effect was inconclusive. Including this perspective in future studies could provide a more comprehensive understanding of whether improved long-haul air accessibility reallocates rather than expands collaboration networks, potentially affecting local or regional research dynamics.

Also, the issue of science policy implications ought to be considered. If direct long-distance air connections are beneficial for the development of research collaboration, should promoting the creation of new air connections be an element of scientific policy? This conclusion seems too far-reaching. Let us underline that better transport accessibility can only facilitate academic mobility, but it is not a sufficient condition for the development of research collaboration. Furthermore, while the role of physical mobility in research collaboration is well-established, recent geopolitical developments highlight additional constraints beyond infrastructure limitations. For example, Wang et al. (2023) demonstrate that political tensions between the U.S. and China have led to a measurable decline in air passenger flows, particularly affecting travel to university hubs, thereby potentially influencing international academic exchange.

Another important aspect emerging from the recommendations of this study is the harmful impact of air transport on the climate. Air transport emits relatively large amounts of greenhouse gases and scientists are one of the most internationally mobile professional groups (Arsenault et al., 2019; Hölbling et al., 2023; Schmidt, 2022). The challenge for science policy, at both the national and institutional levels, is therefore to find a balance between supporting the mobility of scientists and promoting environmental responsibility. On this basis, it could be argued that future research on the mobility of scientists, and its causes and effects, should focus on assessing the effectiveness of different types of mobility (e.g., short vs. long term) and the differences in effectiveness related to the characteristics of scientists (e.g., early vs. late career stages). Such efficiency assessments should consider environmental costs. Moreover, future research could also assess which forms of academic mobility could be replaced by remote communication, to what extent, and under what conditions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11192-025-05291-5.

Acknowledgements This work was supported by the National Science Centre Poland under Grant 2018/31/B/HS4/03997. Data for the project were made available by the Indiana University Network Science Institute and the Science Studies Laboratory of the University of Warsaw. I would like to thank Katy Börner for her warm hospitality and kind support during my stay at the Cyberinfrastructure for Network Science (CNS) Center at Indiana University Bloomington. I thank David Yanagizawa-Drott for kindly providing source data and code. I thank Xiaoran Yan for his generous assistance in processing bibliometric data. Finally, I would like to thank two anonymous reviewers whose insightful comments were very helpful in developing the final version of the paper.

Declarations

Conflict of interest The author has no competing interests to declare that are relevant to the content of this article.

References

Allen, T. J. (2007). *Managing the flow of technology: Technology transfer and the dissemination of technological information within the R&D organization*. MIT Press.

Andersson, Å. E., & Persson, O. (1993). Networking scientists. *The Annals of Regional Science, 27*(1), 11–21. https://doi.org/10.1007/BF01581830

Andersson, M., & Ejermo, O. (2005). How does accessibility to knowledge sources affect the innovativeness of corporations?—Evidence from Sweden. *The Annals of Regional Science, 39*(4), 741–765. https://doi.org/10.1007/s00168-005-0025-7

Angrist, J., & Pischke, J.-S. (2009). *Mostly harmless econometrics*. Princeton University Press.

Arndt, R. T. (1996). *The fulbright difference*. Transaction Publ.

Arsenault, J., Talbot, J., Boustani, L., Gonzalès, R., & Manaugh, K. (2019). The environmental footprint of academic and student mobility in a large research-oriented university. *Environmental Research Letters, 14*(9), 095001. https://doi.org/10.1088/1748-9326/ab33e6

Bahar, D., Choudhury, P., Kim, D. Y., & Koo, W. W. (2023). Innovation on wings: Nonstop flights and firm innovation in the global context. *Management Science, 69*(10), 6202–6223. https://doi.org/10.1287/mnsc.2023.4682

Boudreau, K. J., Brady, T., Ganguli, I., Gaule, P., Guinan, E., Hollenberg, A., & Lakhani, K. R. (2017). A field experiment on search costs and the formation of scientific collaborations. *The Review of Economics and Statistics, 99*(4), 565–576. https://doi.org/10.1162/REST_a_00676

Campante, F., & Yanagizawa-Drott, D. (2018). Long-range growth: Economic development in the global network of air links*. *The Quarterly Journal of Economics, 133*(3), 1395–1458. https://doi.org/10.1093/qje/qjx050

Campbell, D. T., & Stanley, J. C. (2011). *Experimental and quasi-experimental designs for research*. Wadsworth.

Catalini, C., Fons-Rosen, C., & Gaulé, P. (2020). How do travel costs shape collaboration? *Management Science, 66*(8), 3340–3360. https://doi.org/10.1287/mnsc.2019.3381

Chai, S., & Freeman, R. B. (2019). Temporary colocation and collaborative discovery: Who confers at conferences. *Strategic Management Journal, 40*(13), 2138–2164. https://doi.org/10.1002/smj.3062

Crescenzi, R., & Rodríguez-Pose, A. (2012). Infrastructure and regional growth in the European Union. *Papers in Regional Science, 91*(3), 487–513. https://doi.org/10.1111/j.1435-5957.2012.00439.x

Dong, X., Zheng, S., & Kahn, M. E. (2020). The role of transportation speed in facilitating high skilled teamwork across cities. *Journal of Urban Economics, 115*, 103212. https://doi.org/10.1016/j.jue.2019.103212

Duede, E., Teplitskiy, M., Lakhani, K., & Evans, J. (2024). Being together in place as a catalyst for scientific advance. *Research Policy, 53*(2), 104911. https://doi.org/10.1016/j.respol.2023.104911

Dunning, T. (2012). *Natural experiments in the social sciences: A design-based approach*. Cambridge University Press.

Ejermo, O., & Karlsson, C. (2006). Interregional inventor networks as studied by patent coinventorships. *Research Policy, 35*(3), 412–430. https://doi.org/10.1016/j.respol.2006.01.001

Falk, M. T., & Hagsten, E. (2021). When international academic conferences go virtual. *Scientometrics, 126*(1), 707–724. https://doi.org/10.1007/s11192-020-03754-5

Fernández, A., Ferrándiz, E., & León, M. D. (2016). Proximity dimensions and scientific collaboration among academic institutions in Europe: The closer, the better? *Scientometrics, 106*(3), 1073–1092. https://doi.org/10.1007/s11192-015-1819-8

Frenken, K., Hardeman, S., & Hoekman, J. (2009a). Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics, 3*(3), 222–232. https://doi.org/10.1016/j.joi.2009.03.005

Frenken, K., Hoekman, J., Kok, S., Ponds, R., van Oort, F., & van Vliet, J. (2009b). Death of distance in science? A gravity approach to research collaboration. In A. Pyka & A. Scharnhorst (Eds.), *Innovation Networks* (pp. 43–57). Springer.

Guo, W., Del Vecchio, M., & Pogrebna, G. (2017). Global network centrality of university rankings. *Royal Society Open Science, 4*(10), 171172. https://doi.org/10.1098/rsos.171172

Hall, K. L., Vogel, A. L., Huang, G. C., Serrano, K. J., Rice, E. L., Tsakraklides, S. P., & Fiore, S. M. (2018). The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *American Psychologist, 73*(4), 532–548. https://doi.org/10.1037/amp0000319

Hoekman, J., Frenken, K., & Tijssen, R. J. W. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy, 39*(5), 662–673. https://doi.org/10.1016/j.respol.2010.01.012

Hoekman, J., Frenken, K., & van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science, 43*(3), 721–738. https://doi.org/10.1007/s00168-008-0252-9

Hölbling, S., Kirchengast, G., & Danzer, J. (2023). Unmasking mobility patterns: International travel behavior and emissions of scientists in a higher research institution. *International Journal of Sustainability in Higher Education, 24*(9), 355–371. https://doi.org/10.1108/IJSHE-03-2023-0081

Hu, G., Ni, R., & Tang, L. (2022). Do international nonstop flights foster influential research? Evidence from Sino-US scientific collaboration. *Journal of Informetrics, 16*(4), 101348. https://doi.org/10.1016/j.joi.2022.101348

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*(2), 615–635. https://doi.org/10.1016/j.jeconom.2007.05.001

Jacob, M., & Meek, V. L. (2013). Scientific mobility and international research networks: Trends and policy tools for promoting research excellence and capacity building. *Studies in Higher Education, 38*(3), 331–344. https://doi.org/10.1080/03075079.2013.773789

Jin, Z., Cui, Z., Zhu, S., & Xu, L. (2024). Face-to-face contact and university–industry collaboration: Evidence from mobile signaling data in Beijing. *The Annals of Regional Science, 72*(4), 1255–1276. https://doi.org/10.1007/s00168-023-01242-1

Kabo, F. W., Cotton-Nessler, N., Hwang, Y., Levenstein, M. C., & Owen-Smith, J. (2014). Proximity effects on the dynamics and outcomes of scientific collaborations. *Research Policy, 43*(9), 1469–1485. https://doi.org/10.1016/j.respol.2014.04.007

Kang, M., Li, Y., Zhao, Z., Song, M., & Yi, J. (2023). Travel costs and inter-city collaborative innovation: Evidence of high-speed railway in China. *Structural Change and Economic Dynamics, 65*, 286–302. https://doi.org/10.1016/j.strueco.2023.02.002

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1–18. https://doi.org/10.1016/S0048-7333(96)00917-1

Koh, Y., Li, J., & Xu, J. (2024). Subway, collaborative matching, and innovation. *Review of Economics and Statistics*. https://doi.org/10.1162/rest_a_01279

Leahey, E. (2016). From sole investigator to team scientist: Trends in the practice and study of research collaboration. *Annual Review of Sociology, 42*(1), 81–100. https://doi.org/10.1146/annurev-soc-081715-074219

Ma, H., Fang, C., Pang, B., & Li, G. (2014). The effect of geographical proximity on scientific cooperation among Chinese cities from 1990 to 2010. *PLoS ONE, 9*(11), e111705. https://doi.org/10.1371/journal.pone.0111705

Naik, C., Sugimoto, C. R., Larivière, V., Leng, C., & Guo, W. (2023). Impact of geographic diversity on citation of collaborative research. *Quantitative Science Studies, 4*(2), 442–465. https://doi.org/10.1162/qss_a_00248

Netz, N., Hampel, S., & Aman, V. (2020). What effects does international mobility have on scientists' careers? *A Systematic Review. Research Evaluation, 29*(3), 327–351. https://doi.org/10.1093/reseval/rvaa007

Olechnicka, A., Ploszaj, A., & Celinska-Janowicz, D. (2019). *The geography of scientific collaboration*. Routledge.

Olechnicka, A., Ploszaj, A., & Zegler-Poleska, E. (2024). Virtual academic conferencing: A scoping review of 1984–2021 literature. Novel modalities vs. long standing challenges in scholarly communication. *Iberoamerican Journal of Science Measurement and Communication, 4*(1), 1–31. https://doi.org/10.47909/ijsmc.93

Olechnicka, A., Ploszaj, A., & Zegler-Poleska, E. (2025). The impact of the virtualization of scholarly conferences on the gender structure of conference contributors. *Scientometrics, 130*(1), 423–445. https://doi.org/10.1007/s11192-024-05212-y

Petersen, A. M. (2018). Multiscale impact of researcher mobility. *Journal of the Royal Society Interface, 15*(146), 20180580. https://doi.org/10.1098/rsif.2018.0580

Picci, L. (2010). The internationalization of inventive activity: A gravity model using patent data. *Research Policy, 39*(8), 1070–1081. https://doi.org/10.1016/j.respol.2010.05.007

Ploszaj, A., Yan, X., & Börner, K. (2020). The impact of air transport availability on research collaboration: A case study of four universities. *PLoS ONE, 15*(9), e0238360. https://doi.org/10.1371/journal.pone.0238360

Plotnikova, T., & Rake, B. (2014). Collaboration in pharmaceutical research: Exploration of country-level determinants. *Scientometrics, 98*(2), 1173–1202. https://doi.org/10.1007/s11192-013-1182-6

Schmidt, A. (2022). University air travel and greenhouse gas mitigation: An analysis of higher education climate policies. *International Journal of Sustainability in Higher Education, 23*(6), 1426–1442. https://doi.org/10.1108/IJSHE-07-2021-0318

Sebestyén, T., & Varga, A. (2013). Research productivity and the quality of interregional knowledge networks. *The Annals of Regional Science, 51*(1), 155–189. https://doi.org/10.1007/s00168-012-0545-x

Sekhon, J. S., & Titiunik, R. (2012). When natural experiments are neither natural nor experiments. *American Political Science Review, 106*(1), 35–57. https://doi.org/10.1017/S0003055411000542

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J. (Paul), & Wang, K. (2015). An overview of microsoft academic service (MAS) and applications. In: Proceedings of the 24th international conference on world wide web, 243–246. https://doi.org/10.1145/2740908.2742839

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology, 41*(1), 643–681. https://doi.org/10.1002/aris.2007.1440410121

Souto-Otero, M, Humburg, M, Franke, J, & European Commission, Directorate-General for Education, Youth, Sport and Culture. (2017). FP7 ex post and H2020 interim evaluation of Marie Skłodowska-Curie actions (MSCA): Final report. Publications Office. https://doi.org/10.2766/103174

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies, 1*(1), 396–413. https://doi.org/10.1162/qss_a_00021

Wang, S., & Gao, Y. (2021). A literature review and citation analyses of air travel demand studies published between 2010 and 2020. *Journal of Air Transport Management, 97*, 102135. https://doi.org/10.1016/j.jairtraman.2021.102135

Wang, Z., Tang, L., Cao, C., & Zhou, Z. (2023). The impact of U.S.-China tensions on people mobility. *China Review, 23*(4), 159–195.

Wynes, S., Donner, S. D., Tannason, S., & Nabors, N. (2019). Academic air travel has a limited influence on professional success. *Journal of Cleaner Production, 226*, 959–967. https://doi.org/10.1016/j.jclepro.2019.04.109

Yao, L., & Li, J. (2022). Intercity innovation collaboration and the role of high-speed rail connections: Evidence from Chinese co-patent data. *Regional Studies, 56*(11), 1845–1857. https://doi.org/10.1080/00343404.2021.2008340