

# Collaboration in pharmaceutical research: exploration of country-level determinants

Tatiana Plotnikova · Bastian Rake

Received: 6 April 2013 / Published online: 22 November 2013  
© Akadémiai Kiadó, Budapest, Hungary 2013

**Abstract** In this paper we focus on proximity as one of the main determinants of international collaboration in pharmaceutical research. We use various count data specifications of the gravity model to estimate the intensity of collaboration between pairs of countries as explained by the geographical, cognitive, institutional, social, and cultural dimensions of proximity. Our results suggest that geographical distance has a significant negative relation to the collaboration intensity between countries. The amount of previous collaborations, as a proxy for social proximity, is positively related to the number of cross-country collaborations. We do not find robust significant associations between cognitive proximity or institutional proximity with the intensity of international research collaboration. Our findings for cultural proximity do not allow of unambiguous conclusions concerning their influence on the collaboration intensity between countries.

**Keywords** International cooperation · Pharmaceuticals · Proximity

## Introduction

Collaboration has been found to be an increasingly common mode of knowledge creation. Early evidence suggests that there has been a steadily increasing trend towards collaborative research in the field of chemistry for the period from 1910 to 1960 (de Solla Price 1963). More recently, Adams et al. (2005) showed for a broad set of disciplines that scientific research originating from the 110 top U.S. research universities is increasingly

---

T. Plotnikova  
National Institute for Statistics and Economic Studies of the Grand Duchy of Luxembourg (STATEC),  
13, rue Erasme, 2013 Luxembourg, Luxembourg  
e-mail: tatiana.plotnikova@statec.etat.lu

B. Rake (✉)  
University of Passau, Innstrasse 27, 94032 Passau, Germany  
e-mail: bm.rake@gmail.com; bastian.rake@uni-passau.de

conducted in collaboration. Using an extended dataset of 19.9 million papers and 2.1 million patents, Wuchty et al. (2007) show that the trend towards collaborative knowledge production is particularly prevalent in publication data in such fields as science and engineering, social sciences, as well as in patent data. It is much less evident in arts and humanities. The general trend of increasing collaboration in the production of knowledge is accompanied by growing volume of international research. During the 1980s, the annual growth of the share of international research collaboration was slightly more than five percent, increasing to more than seven percent in the subsequent decade (Adams et al. 2005). The impact of knowledge produced in international collaboration is particularly pronounced since it is more frequently cited than research produced in different institutions within the same country, as well as in a single institution (Narin et al. 1991).

The increasing amount and the importance of international collaboration in scientific research motivated us to explore its determinants. Earlier contributions to the subject matter considered a variety of country-level factors, such as cognitive, geopolitical, social, historical, and economic aspects, as potential determinants of international scientific collaboration (Luukkonen et al. 1992). These factors include the relevant institutional settings such as research funding schemes, research infrastructure as well as the priorities set by scientists and policy makers to conduct research in particular fields and to collaborate with different partners (Miquel and Okubo 1994; Lundvall 1988, 1992). These aspects can be related to the national innovative capacity, i.e., the long-term ability of a country to generate and commercialize innovative technologies (Furman et al. 2002). Differences in countries' national innovative capacities may induce the exchange of knowledge and approaches for problem solving among organizations and individuals embedded in national contexts with country-specific scientific and technological advantages (Bartholomew 1997; Shan and Hamilton 1991; Dosi et al. 1990). A self-organizing network of international scientific collaborations emerges in which the collaboration patterns of lower order subsystems, such as pharmaceuticals, are affected by macro-level factors which are at the same time influenced by collaborations in lower order subsystems (cf., Wagner and Leydesdorff 2005).

In contrast to many earlier contributions providing descriptive evidence of international scientific collaboration, we aim to deliver the empirical evidence of the impact of various factors on the intensity of international collaboration through regression analysis. More precisely, our objective is to explain the intensity of international collaborations in pharmaceutical research by similarities and differences in the scientific environment of collaborating countries, such as support of science, the institutional settings, overlaps in the knowledge bases, and cultural linkages.

We rely on the literature in economic geography highlighting multiple dimensions of proximity (cf., Boschma 2005). This literature mainly focuses on organizational and regional dimensions of collaboration and finds that similarity or proximity of organizations matters for collaboration. In the vast majority of related studies data collection is based on policy programs or predefined geographical entities (e.g., Balland 2012; Hoekman et al. 2010; Scherngell and Hu 2011). It is reported by these studies that multiple dimensions of proximity matter when explaining scientific collaborations. The most prominent result is that geographical proximity explains a good part of collaboration.

This paper addresses the question whether similar patterns are observed on the macro level. Although proximity is a fairly general concept, some dimensions of it may have a different effect on the occurrence and intensity of collaborations depending on the level of aggregation. Different factors may be relevant in explaining collaboration between individuals, organizations, regions, and countries. We argue that although some general

patterns may persist on different levels of aggregation, country-level collaboration can be determined by factors different from that on the organizational or regional level. These factors may include the institutional setting of countries, such as the support of scientific research, and the level of scientific and technological development. Hence, we analyze whether multiple dimensions of proximity play a crucial role in explaining the number of scientific collaborations proxied by co-publications between pairs of countries.

We contribute to the understanding of country-level scientific collaboration by highlighting the role of similarities and differences in the macro-environment that may encourage or discourage international scientific collaborations. Using the concept of multiple dimensions of proximity, we add different dimensions of cultural proximity as well as other factors such as differences in countries' overall R&D expenditures to the analysis which have not been included in most studies on proximity and scientific collaboration. In contrast to many other studies, our data was obtained based on an application perspective, namely pharmaceutical research related to specific medical indications. In doing so, this paper refers to pharmaceutical research in a broad sense taking basic biomedical research and clinical research into account.

The empirical analysis is performed on a unique dataset of scientific journal publications related to pharmaceutical research using different count data specifications of the gravity model. Pharmaceutical research has been widely studied in the economic literature concerned with the questions of knowledge, innovation, and science due to the importance of the field for humanity as well as high significance of basic research performed by universities and public research organizations for applied research and generation of new products by companies. Collaboration in pharmaceutical research is widespread and not restricted to the national level but extends beyond national borders. A possible explanation of this phenomenon can be found in differences in the national innovative capacities of countries that are accompanied with country-specific scientific and technological advantages (e.g., Bartholomew 1997).

The remainder of this paper is structured as follows. We describe the determinants of international collaboration in “[Related literature](#)” section. “[Data and research methodology](#)” section provides details of the methodology and the data used in our empirical analysis. “[Empirical results](#)” section reports the results of the analysis of international collaborations in the pharmaceutical research. “[Conclusion](#)” section concludes.

## Related literature

Given the increasing amount of international scientific collaboration and global knowledge sourcing, we ask about the determinants of this phenomenon on the country level. Put differently, we aim to explain what country-level factors determine why researchers from one country collaborate with colleagues in some countries, but not with those from others. One argument raised in the literature is that knowledge production and knowledge spillovers are geographically bounded within the region of creation (Audretsch and Feldman 1996; Jaffe et al. 1993). Particularly the transfer of tacit knowledge, often involved in R&D processes, is done best through face-to-face interaction, which is facilitated by close spatial distances.

Geographical proximity has been found to be an important factor determining the extent of international scientific collaboration (Luukkonen et al. 1992; Ponds et al. 2007). However, while we focus on proximity as one of the main determinants of international collaboration, we do not only refer to geographical distance between countries. Instead, we

follow Boschma (2005) who draws upon the French School of Proximity Dynamics (e.g., Torre and Rallet 2005; Torre and Gilly 2000; Kirat and Lung 1999) in suggesting that proximity is multidimensional. More precisely, we focus our analysis on geographical, cognitive, institutional, and social proximity. We also explore the linguistic and historical ties between countries in order to account for an additional dimension of proximity: cultural proximity.

In order to isolate geographical from other dimensions of proximity, its definition is restricted to spatial or physical distances between economic actors (Boschma 2005). The distance can be expressed in absolute terms, e.g., in kilometers, or in relative terms, e.g., travel times. Short distances between economic actors facilitate personal contacts, the exchange of information, and, particularly, tacit knowledge. Hence, geographical proximity may facilitate inter-organizational learning but is not a prerequisite for collaboration and learning, since other proximity dimensions may act as substitutes. On the other hand, geographical proximity may also be complementary to relational, organizational, institutional, and cognitive proximity and enhance interaction, knowledge creation and innovation more indirectly (Boschma 2005; Howells 2002). The literature suggests that geographical proximity is an important determinant of research collaborations on the regional level in Europe and is particularly important for cooperative work in life sciences (Hoekman et al. 2009, 2010; Ponds et al. 2007). On the country-level, it has been shown that geographical proximity is an important factor to explain international scientific collaboration. Among scientifically leading countries worldwide, those that are located geographically proximate tend to collaborate more intense (Nagpaul 2003). Hence, we argue that this relation should hold also for international collaboration in pharmaceuticals. Therefore, our first hypothesis is:

**Hypothesis 1** Geographical distance is negatively related to the intensity of international research collaborations.

From the evolutionary perspective, knowledge creation and innovation often imply a high degree of tacit knowledge and can be seen as cumulative outcomes of search processes conducted by boundedly rational agents. The creation of new knowledge and learning about existing knowledge depends in many instances on the combination of diverse but complementary capabilities (Nooteboom 2000). The tacit and idiosyncratic component of knowledge implies that absorptive capacity is required to identify, interpret, and exploit new knowledge. Hence, it is a precondition for effective knowledge transfer (Cohen and Levinthal 1990). Given these circumstances, cognitive proximity, also called technological proximity, is required in the exchange and combination of knowledge. Cognitive proximity describes the sufficient closeness of an actor's knowledge base towards new knowledge in order to permit successful communication, understanding, and absorption (Boschma and Lambooy 1999). An empirical test of the relation between cognitive proximity and firm cooperation demonstrated that an overlap in firms' patent stocks is associated with a higher probability of cooperation (Cantner and Meder 2007). Analyzing R&D collaborations on the regional level, Scherngell and Barber (2009, 2011) suggest that collaborations are most likely when regions are close in the cognitive dimension of proximity. In contrast to these results, Balland (2012) finds that organizations do not necessarily collaborate when their knowledge bases overlap.

Based on these findings, we ask whether cognitive proximity has a positive relation to the collaboration intensity in international collaborations on the country-level. Differences in countries' national innovative capacities imply the existence of country-specific knowledge which is the basis for the generation of new knowledge as well as for the

development and commercialization of new products (Furman et al. 2002). Different characteristics and specializations of countries' knowledge bases open up opportunities for knowledge exchange between organizations embedded in different national contexts (Bartholomew 1997; Shan and Hamilton 1991; Dosi et al. 1990). However, different country-specific knowledge bases have to be sufficiently close to each other in order to build the basis for successful research collaboration. Hence, we formulate Hypothesis 2 as follows:

**Hypothesis 2** Cognitive proximity has a positive relation with the intensity of international research collaboration.

Institutional proximity refers to the institutional environment at the macro-level. In this sense, institutions refer to as “sets of common habits, routines, established practices, rules or laws that regulate the relations and interactions between individuals and groups” (Edquist and Johnson 1997). Formal, e.g., laws, and informal, e.g., cultural norms, institutions influence the manner of and the extent of collaboration. Institutional proximity has been regarded as an enabling factor that provides stable conditions for effective interactive learning. Gertler (1995) shows that institutional differences at the macro-level, i.e., in this case job training and workplace practices, can hinder cross-border interactions among firms. More generally, empirical studies suggest that institutional proximity with respect to country-specific institutional settings may positively influence research collaboration (Hoekman et al. 2009, 2010). The intuition behind this finding is that institutions, i.e., rules and regulations affecting research projects as well as schemes to support scientific research, differ between countries. Institutional proximity, or similarity, may provide the basis for a common understanding of the values, rules, and regulations involved in the research process and may, thus, facilitate (international) research collaboration. Based on these arguments, we build our hypothesis concerning institutional proximity and research collaboration on the country-level.

**Hypothesis 3** Institutional proximity is positively related with the number of international research collaborations.

The social dimension of proximity refers to the more intangible closeness capturing the intensity of scientific collaboration between interacting parties (Basile et al. 2012; Boschma 2005; Coenen et al. 2004). This dimension of proximity draws upon the embeddedness literature, which suggests that economic activities are embedded in a social context (Granovetter 1985). Social proximity takes into account the idea that the extent of embeddedness is connected to opportunities for knowledge exchange and joint knowledge creation that may positively affect innovative activities. Shared experience can be seen as the basis for the existence of social proximity (Asheim et al. 2007). In line with these thoughts, empirical evidence points out the importance of social proximity based on past experience for research collaboration between firms (Autant-Bernard et al. 2007). The importance of joint experience for research collaboration has not been restricted to the organizational level. As Hoekman et al. (2013) suggest, past research collaboration, measured by co-publications, may increase the number of research collaborations on the regional level. Based on this evidence, we argue that social proximity based on previous collaborations increases the number of cross-country research collaborations. The observed collaboration patterns on the macro- or cross-country level may be shaped by micro-level decisions to continue and to intensify collaborations depending on the collaboration experiences made. Hence, Hypothesis 4 summarizes these thoughts:

**Hypothesis 4** Social proximity is positively related to the intensity of international research collaborations.

Similar to the institutional dimension of proximity, cultural proximity reflects a common cultural background. A common cultural space is among others formed by a common working tradition, a common language, mutual trust, and mutually respected norms of behavior (Zeller 2004). The presence of cultural differences impedes the transmission and decoding of certain types of messages, especially if tacit components are involved (Lundvall 1992). Accordingly, cultural proximity may affect interactions and research collaboration on the micro- and on the macro-level. On the micro-level, scientists' decisions may, consciously or unconsciously, favor collaboration with partners that are close in terms of cultural proximity due to advantages in communication and knowledge exchange (Zitt et al. 2000). On the macro- or country-level, the aggregation of individual decisions leads to specific interaction or collaboration patterns between countries as shown by empirical evidence. In the trade literature it has been shown that cultural proximity, often operationalized as common language, is positively connected to trade volumes among countries (e.g., Felbermayr and Toubal 2010; Melitz 2008). With respect to scientific interactions, empirical evidence reveals that linguistic and historical ties influence collaboration intensity (Hoekman et al. 2010; Zitt et al. 2000). Therefore, we formulate our hypothesis concerning the impact of cultural proximity as follows:

**Hypothesis 5** Cultural proximity is positively related with the intensity of international research collaboration.

## Data and research methodology

### Gravity model

We analyze the determinants of collaboration among different countries using a gravity model. Early applications of gravity models in economics were focused on the analysis of international trade flows (e.g., Isard 1954; Tinbergen 1962). Later, this model has been applied to a broad variety of research questions. In the context of research collaborations, it has been used to analyze the intensity of co-publications among regions (Hoekman et al. 2009, 2010). The basic idea of the gravity model can be traced back to Newton's law of universal gravitation which states that the gravitational force between two objects is proportional to the product of the masses of the objects and the distance between them. The basic gravity equation can be expressed as follows (cf., Burger et al. 2009; Hoekman et al. 2009):

$$I_{ij} = \beta_0 \frac{M_i^{\beta_1} M_j^{\beta_2}}{d_{ij}^{\beta_3}}, \quad (1)$$

where  $I_{ij}$  denotes interaction intensity, i.e., the number of research collaborations, between countries  $i$  and  $j$ .  $\beta_0$  is a proportionality constant.  $M_i$  and  $M_j$  represent the masses of country  $i$  and  $j$ , each of which in our case is the number of publications. The distance between the two countries is denoted by  $d_{ij}$ .  $\beta_{1,2}$  reflect the potential to collaborate and  $\beta_3$  reflects the effect of distance. The multiplicative form of the gravity model presented in (1) can be

transformed into a testable linear model by taking logarithms of both sides and adding a disturbance term  $\epsilon_{ij}$ :<sup>1</sup>

$$\ln I_{ij} = \ln \beta_0 + \beta_1 \ln M_i + \beta_2 \ln M_j - \beta_3 \ln d_{ij} + \epsilon_{ij}. \tag{2}$$

There has been an intense discussion in the literature on the subject of which regression models are best suited to estimate (2). Log-linearized models estimated by OLS have been frequently used in many economic applications of the gravity model. However, log-linearized models do not allow for zero interaction intensities. As a consequence, many researchers used truncated samples, which did not include the non-interacting pairs. Helpman et al. (2008) show that the restriction to positive interaction intensities may substantially bias the results. As an alternative the authors suggest to explicitly model zero interactions. Santos Silva and Tenreyro (2006) provide evidence that count data models outperform the standard log-linearized models estimated by OLS. Moreover, count data models are compatible with the existence of zero interaction in the data. Hence, we follow Santos Silva and Tenreyro (2006) and apply count data regression techniques. More precisely, we use Poisson and negative binomial models, based on maximum likelihood techniques (Burger et al. 2009; Flowerdew and Aitkin 1982). In a Poisson regression framework, the observed volume of research collaboration between  $i$  and  $j$  is Poisson distributed with conditional mean  $\mu$  which can be expressed as a function of the independent variables.

$$\Pr[I_{ij}] = \frac{\exp(-\mu_{ij})\mu_{ij}^{I_{ij}}}{I_{ij}!}, \quad (I_{ij} = 0, 1, \dots). \tag{3}$$

The conditional mean  $\mu_{ij}$  is linked to an exponential function of the regression variables:

$$\mu_{ij} = \exp(\beta_0 + \beta_1 \ln M_i + \beta_2 \ln M_j + \beta' \ln D_{ij}). \tag{4}$$

In (4),  $\beta_0$  is a constant,  $D_{ij}$  is the vector of explanatory variables representing different dimensions of distance, and  $\beta'$  is the corresponding parameter vector. An important caveat of the Poisson model is the assumption of equidispersion, which means that the variance equals the mean. In order to correct for the violation of this assumption we employ a negative binomial regression model, which can be seen as a modified Poisson model and is frequently used in count data analysis (Greene 1994):

$$\Pr[I_{ij}] = \frac{\Gamma(\alpha^{-1} + I_{ij})}{\Gamma(\alpha^{-1})I_{ij}!} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \mu_{ij}} \right)^{\alpha^{-1}} \left( \frac{\mu_{ij}}{\mu_{ij} + \alpha^{-1}} \right)^{I_{ij}}, \tag{5}$$

where  $\mu_{ij}$  is the conditional mean,  $\Gamma$  is the gamma function, and  $\alpha$  is the parameter determining the degree of dispersion, allowing that the conditional variance exceeds the conditional mean. Larger  $\alpha$  corresponds to a larger degree of overdispersion in the data. With  $\alpha$  being approximately zero, the negative binomial model reduces to the Poisson regression model.

Another problem prevalent in many practical applications of Poisson and negative binomial estimation is an excessive number of zeros in the data. In other words, the

<sup>1</sup> The trade literature suggests to add origin and destination fixed effects to the analysis (e.g., Helpman et al. 2008). However, this approach is not applicable in the case of scientific collaborations since these represent undirected interactions whereas trade flows are directed interactions. Moreover, trade cost may be better observable than the cost of collaboration which are not available in our dataset.

problem arises when the number of zero counts is greater than what the Poisson or the negative binomial distribution would predict. In order to overcome this problem, zero-inflated versions of both Poisson and negative binomial models may be applied.

The basic idea behind a zero-inflated model is that zero values are generated by a different process than the positive ones. The first part of a zero-inflated model consists of a binary process, which in our empirical application is a logit model. The dependent variable in this logit model takes value zero when there is no collaboration between respective countries in a certain therapeutic area. If the binary process equals one, the number of collaborations is equal to or greater than zero. In the second part of the estimation, a Poisson or negative binomial regression model is applied to estimate the collaboration intensity. Hence, zeros can be the outcome of both the binary process and the count process, given that the binary process takes the value one. We can express the zero-inflated Poisson model as

$$\Pr[I_{ij} = 0] = \psi_{ij} + (1 - \psi_{ij}) \exp(-\mu_{ij}), \quad (6)$$

$$\Pr[I_{ij}] = (1 - \psi_{ij}) \frac{\exp(-\mu_{ij}) \mu_{ij}^{I_{ij}}}{I_{ij}!}, \quad (7)$$

where (6) refers to the first part and (7) to the second.  $\psi$  is the proportion of observations with a strictly zero count determined by the logit model (cf., Burger et al. 2009). When  $\psi$  equals zero, the model reduces to Poisson model. Along similar lines we can define the zero-inflated negative binomial regression model:

$$\Pr[I_{ij} = 0] = \psi_{ij} + (1 - \psi_{ij}) \left( \frac{\alpha^{-1}}{\alpha^{-1} + \mu_{ij}} \right)^{\alpha^{-1}}, \quad (8)$$

$$\Pr[I_{ij}] = (1 - \psi_{ij}) \frac{\Gamma(I_{ij} + \alpha^{-1})}{I_{ij}! \Gamma(\alpha^{-1})} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \mu_{ij}} \right)^{\alpha^{-1}} \left( \frac{\mu_{ij}}{\alpha^{-1} + \mu_{ij}} \right)^{I_{ij}}. \quad (9)$$

For both versions of the zero-inflated model, the Vuong test statistic can be used to test whether the zero-inflated model is favored above the respective uninflated versions, by analyzing if there is significant evidence for excessive zero counts (Vuong 1989).

## Data

In our empirical analysis we use data from different sources. We start by drawing a list of 251 medical indications from BioPharmInsight.<sup>2</sup> Each indication represents a condition, disease or symptom. Each indication is exclusively assigned to one of 15 therapeutic areas.<sup>3</sup> Therapeutic areas are defined according to a system of an organism or a general disease group. Examples of therapeutic areas are “Central Nervous System” and “Infectious Diseases”.

This list of medical indications (or diseases) was used to search for corresponding scientific pharmaceutical publications in the Web of Science databases (WoS). The WoS consists of seven databases containing information gathered from an extensive number of

<sup>2</sup> <http://www.infinata5.com/biopharm/>.

<sup>3</sup> Table 3 provides an overview of the therapeutic areas included in the dataset.



journals, books, book series, reports, and conferences. Among these databases, the most important one is the Science Citation Index Expanded. It is multidisciplinary and indexes more than 6,500 scientific journals covering 150 scientific disciplines. The Science Citation Index Expanded covers, among other scientific fields, biochemistry, medicine, and pharmacology, which are of particular interest for our study. The WoS contains such information about each scientific publication as the title, the year of publication, the journal, cited references, a categorization of the research fields a publication can be assigned to, and further bibliographic information. In addition to this information, the Web of Science reports authors' affiliations, including the country of the location of a respected organization for most articles.

Scientific publications in the database were searched for the occurrence of each of 251 medical indications in their title. We consider all publications included in categories related to pharmaceutical research. Articles from the subcategories “Biochemistry and Molecular Biology”, “Biotechnology and Applied Microbiology”, “Chemistry, Applied”, “Chemistry, Medicinal”, “Medicine, Research and Experimental”, “Pharmacology and Pharmacy”, and “Toxicology” are included in our dataset.<sup>4</sup> We restrict our sample to journal articles and exclude publications that are labeled as meeting abstracts, editorials or reviews as well as other non-journal publications. Conference proceedings have not been considered, either since they may be of different quality compared to published papers or may be already included as published articles. For the period from 1974 to 2008 we obtain 211,661 publications. Unfortunately, the information concerning authors' affiliations was not included in the WoS prior to 1998 for a considerable number of cases. Therefore, we concentrate on the years from 1998 to 2008 which encompass 113,057 articles. After selecting articles which contained authors' affiliations, we obtained a sample of 111,096 journal articles.

In order to take into account the expansion in the number of collaborations over time, we divide the sample into two sub-periods of equal length, 1998–2002 and 2004–2008. The number of countries stays relatively similar among the sub-periods: 106 countries in the first period and 109 countries for the second period. However, the composition of countries changes a bit since there are 114 countries in total in the sample. Table 2 provides a list of the 114 countries of origin of the publications in our two periods of observation. The difference between the two sub-samples with respect to the number of zeros is more striking: on the level of therapeutic areas we have 79.1 % zeros in the first and 69.2 % zeros in the second period; without distinguishing among therapeutic areas there are 67.8 % zeros in the first and 57.3 % in the second period.

An additional 66,312 journal articles published between 1974 and 1997 for which we could identify author affiliations were used to construct a proxy for the amount of collaboration between countries in different therapeutic areas prior to the periods of observation in our sample.

Information concerning authors' affiliations is matched with WHO Regions. Since the WHO Regions do not classify all countries included in our database, we assign previously unclassified countries to additionally created regional groups. More precisely we create groups for the members of the EU-15, the United States and Canada (North America), as well as for Australia and New Zealand (Australasia), Japan, and Switzerland. We employ World Bank income groups in order to include information concerning the the wealth level of countries in our sample. Moreover, we use the World Bank Science and Technology database in order to get information about R&D activities on the country level.

<sup>4</sup> The subcategories are described in detail at <http://scientific.thomsonreuters.com/mjl/>.

We use the CHI classification of journals (Hamilton 2003) to classify each article according to the type of research prevalent in the journal it is published in. Following the CHI classification of journals 35.5 % of the journal articles in our sample are published in journals that are predominantly focussed on “basic biomedical research”, 35.2 % belong to the “clinical investigation” category. The remaining articles are classified as rather applied research in the categories “clinical mix” (13 %), “clinical observations” (3.9 %), or are not classified.

We employ the Centre d’Études Prospectives et d’Information Internationales (CEPII) database on distance measures (Mayer and Zignago 2006). The database includes different measures for geographical distances between most countries of the world. This dataset allows us to control for additional sources of proximity among countries, based on the same language, colonial linkages, or shared history as parts of the same country. In order to get additional information concerning institutional environment in different countries, we use the “Institutional Profiles Database” for the year 2009 provided by the CEPII (de Crombrugge et al. 2009). The database offers a quantitative evaluation of the institutional characteristics of 123 developed and developing countries. In order to assess the institutional profile of countries with respect to innovation and R&D activities, we concentrate on indicators that contain information concerning the existence of governmental strategies for the development of human capital, the public support for R&D activities, the technological environment and the dissemination of technologies, society’s response to innovation, as well as the level and the effectiveness of intellectual property protection.

The use of publication data implies an advantage of getting access to highly detailed information covering a long time span. Some of the major drawbacks are that research does not necessarily lead to publications, co-authorship may only partly capture scientific collaboration, the impact of publications differs considerably, and publication habits differ among scientific disciplines. Moreover, publication databases may be biased towards English language publications originating in industrialized countries. Although co-publication data is associated with the mentioned shortcomings, it has been found to be an appropriate indicator for scientific collaboration (see, e.g., Katz and Martin 1997; Laudel 2002; Lundberg et al. 2006; Hoekman et al. 2009 for a discussion).

## Variables and descriptives

The dependent variable in the gravity model is the amount of collaboration between pairs of countries.<sup>5</sup> In order to construct a proxy for the collaboration intensity, we assign each publication to the respective countries mentioned in the authors’ affiliations. The dependent variable is then calculated as the number of co-publications between each pair of countries. We distinguish collaboration intensity for each therapeutic area and each sub-period. We use full counting which implies that a publication that can be assigned to three different nations leads to an interaction intensity between each country pair of one. Since co-publications represent undirected links, we include each pair of countries only once in our analysis.

As we have pointed out, the gravity model assumes that the interaction between two countries depends on their masses. In order to derive a proxy for the mass of a country we count the total number of publications per country in the respective period. Variables *log\_PubActor* and *log\_PubPartner* represent the logs of these counts. Similar to the dependent variable, we consider publications per therapeutic area.

<sup>5</sup> Appendix 1 provides a description of the variables.

Geographical proximity (*log\_distw*) is calculated based on city-level data in order to account for the geographic distribution of the population inside each nation. The distance between two countries is calculated based on bilateral distances between the biggest cities of the two nations. These inter-city distances are weighted by the share of the city in the overall population and enter as logs in our regression models (cf., Mayer and Zignago 2006).

In correspondence with the definition of social proximity we calculate the number of collaborations lagged by one period (*log\_PrevExperience*) as a proxy for this dimension of proximity. In doing so, we account for the possibility that researchers from different organizations located in different nations have established collaborations during our two periods of observation based on previous experience from joint research projects. Furthermore, we may also account for formal and informal ties between organizations by applying this measure of previous experience. Before taking the log we add one to this proxy.

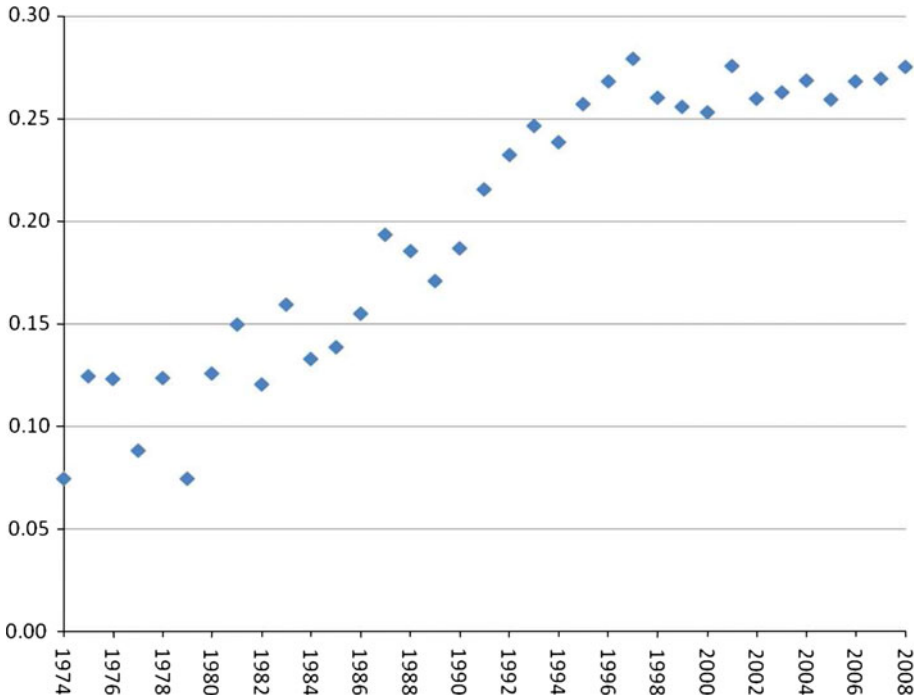
Our measure of cognitive proximity (*log\_PrevSpecialCorr*) is based on the specialization profiles of countries among therapeutic areas prior to the analyzed period. Following the idea of Jaffe (1986) and Peri (2005), we construct a vector containing the shares of publications in each therapeutic area per country prior to the analyzed period. We calculate the uncentered correlation, which corresponds to the cosine, of these vectors for each country pair and take the log.

Our measure of institutional proximity (*log\_InnoInstCorr*) is constructed along the similar lines. We calculate the log of the cosine of vectors containing different indicators from the “Institutional Profiles Database” for each pair of countries. In doing so we account for differences in the institutional environment with respect to innovation and R&D activities among countries.

We account for cultural proximity by including a set of dummy variables indicating if at least 9 % of the population in both countries share the same language (*comlang\_ethno*), if two countries have ever had a colonial link (*colony*), i.e., one country was the colonizer and the other was its colony, or were part of the same country (*smctry*). Since a common language may help to compensate for low levels of proximity, we include interactions terms between the common language dummy and the population weighted distance (*langdistw*), our proxy for social proximity (*langExp*), the measure for cognitive proximity (*langSpecial*), as well as the measure for institutional proximity in our empirical analysis (*langInnoInst*). Moreover, we control for whether two countries are adjacent (*contig*) and belong to the same Worldbank income group (*SmIncomeGr*). We control for differences in countries’ overall R&D spendings as percentage of their GDP (*log\_R&DDifference*). Prior to calculating the log, we add 0.00001 to this variable. Furthermore, we add a dummy indicating whether the collaboration took place in period 2 (2004–2008). When we distinguish the amount of collaboration among therapeutic areas, we add dummy variables for different therapeutic areas to our analysis.

Figure 1 illustrates the development of cross-country collaborated research articles over time. Similar to many other studies we find an increasing share of international research collaboration. Nevertheless, we find that by 2008 almost 72 % of the collaborations in our sample take place within national borders.<sup>6</sup> There are considerable differences in the frequency of international collaboration among WHO regions and Worldbank income groups. In contrast to publications originating in the European Union, North America and Japan, we find that articles from Switzerland and Sub-Saharan Africa show particularly

<sup>6</sup> See Figs. 2 and 3 in the Appendix 2 for a more detailed illustration.



**Fig. 1** Share of cross-country collaborations per year

high shares of international collaborations. With respect to income groups, our descriptive results suggest that organizations from low income countries are particularly engaged in international collaborations if they publish scientific articles, whereas organizations from OECD member states do not engage extensively in cross-country collaboration.

## Empirical results

### Regression results

We start our empirical analysis with Poisson regression models.<sup>7</sup> However, in contrast to the assumptions made in the Poisson framework, the variance of the dependent variable exceeds the mean for our sample, implying overdispersion. Using the test proposed by Cameron and Trivedi (1990), we find significant overdispersion in most model specifications. Therefore, we account for the possibility of overdispersion by using robust standard errors and by applying negative binomial regressions, which have been established as a standard alternative to the Poisson model.

As in many applications of count data, the data on cross-country research collaborations shows an excessive number of zeros as shown in “Data” section. We deal with this data structure by estimating a zero inflated version of the negative binomial model. The Vuong test (Vuong 1989) suggests preferring the zero-inflated models over their ordinary

<sup>7</sup> Results not presented in this paper are available upon request.

counterparts. The goodness-of-fit statistics indicates that the zero-inflated negative binomial model is most suitable for our dataset since it has the highest value of the log-pseudolikelihood and the smallest values of the Akaike criterion (AIC) and the Bayesian Information Criterion (BIC).

Tables 1 and 7 give the results of our empirical estimation. Table 1 reports estimation results on the level of therapeutic areas, whereas Table 7 does so on the aggregate level. Due to a particular time dynamics of our data, we test our hypotheses separately for each time period as well as for both time periods. Our proxies for the masses *log\_PubActor* and *log\_PubPartner*, i.e., the number of publications assigned to each country, have a positive sign and are significant, which implies that the number of publications is positively related with the intensity of scientific collaboration.

Similar to other studies in the field, we find that the population weighted distance between countries, *log\_distw*, as a proxy for geographical distance has, with an exception in Table 1 model (1), a negative and highly significant association with the number of cross-country collaborations. This finding might be influenced by the correlation between *log\_distw* and *log\_PrevExperience* variable since the coefficient of *log\_distw* turns out to be significant and increases its size considerably once the measure for social proximity is excluded from the analysis in a robustness check. In other words, we find that collaboration intensity decreases with spatial distance, which corresponds to our expectations formulated in Hypothesis 1. A possible explanation for this finding is that geographic distance impedes face-to-face interaction which is particularly important for scientific collaboration involving the transfer of tacit knowledge. Hence, our finding on the country-level may reflect the aggregated micro-level decisions of individual researchers and organizations to collaborate more intensively with partners in countries that are located nearby their country of origin and with whom they can interact more easily and frequently.

Our measure for social proximity, *log\_PrevExperience*, is positively related to the amount of collaboration among country pairs. Hence, Hypothesis 4 cannot be rejected. This result suggests that past collaborations intensify future collaborations within the same pair of countries. The reason for this effect may be both willingness of an organization to continue working with the same partner abroad, as well as establishment of new collaborations between different agents motivated by positive experience of their peers. Therefore, this finding suggests that researchers and organizations are more likely to connect to colleagues and institutions from abroad if they have some common experience or observe others having successful collaborations with them, since it reduces uncertainty and the risk of opportunism. Moreover, it might be easier for researchers and organizations that have already collaborated to acquire research grants which are likely to promote further collaborative generation of scientific output. Hence, research funding decisions and similar policy programs on the national or international level may provide an additional explanation for our result.

We do not find a robust association between cognitive proximity (*log\_PrevSpecialCorr*) and the number of cross-country research collaborations. Moreover, the coefficient has a negative sign in all model specifications. Therefore, Hypothesis 2 stating a positive relationship has to be rejected. In other words, this finding does not indicate that international research collaborations are established among countries that are specialized in similar therapeutic areas, instead showing that collaborations occur between countries with complementary knowledge. It may be that the knowledge basis of the countries in our dataset is sufficiently close since otherwise actors located in the respective countries would not be able to publish in international scientific journals. Hence, similar specialization patterns in different disease areas may not be a precondition for mutual understanding.

**Table 1** Zero-inflated negative binomial regression models with therapeutic area distinction

	(1) 1998–2002	(2) 1998–2002	(3) 1998–2002	(4) 2004–2008	(5) 2004–2008	(6) 2004–2008	(7) Pooled	(8) Pooled	(9) Pooled
Dependent variable: number of collaborations									
log_PubActor	0.3483***	0.3633***	0.3677***	0.4263***	0.4385***	0.4388***	0.3871***	0.4006***	0.4027***
log_PubPartner	0.3560***	0.3711***	0.3759***	0.4143***	0.4231***	0.4244***	0.3824***	0.3936***	0.3956***
log_distw	-0.0324	-0.0866***	-0.0957***	-0.1196***	-0.1583***	-0.1682***	-0.0788	-0.1251***	-0.1336***
log_PreVExperience	0.2851***	0.2662***	0.2771***	0.3122***	0.2930***	0.2947***	0.3005***	0.2815***	0.2886***
log_PreVSpecialCorr	-0.7701***	-0.6833***	-0.6113**	-0.2528	-0.2402	-0.1758	-0.4548***	-0.3875**	-0.2837
log_InnoInstCorr	-1.3793	-2.1022	-1.1268	-3.7126***	-4.5103***	-4.6432***	-3.1264***	-3.9805***	-3.7323***
comlang_ethno		0.4514***	-0.5517		0.3261***	-0.8129**		0.3888***	-0.6906**
colony		0.0085	0.0077		0.3148***	0.2720***		0.1889***	0.1594***
smctry		0.5429**	0.6751***		0.2154	0.2966		0.4273**	0.5564***
langdistw			0.1126**			0.1291***			0.1222***
langExp			-0.0904**			-0.0053			-0.0473*
langSpecial			-0.2832			-0.6462			-0.5031
langInnoInstCorr			-6.2937			-0.3322			-2.7442
contig	0.5621***	0.2295**	0.3595***	0.2685***	0.0833	0.1735**	0.4300***	0.1646**	0.2744***
SmIncomeGr	0.2769***	0.3317***	0.3358***	0.1578**	0.1898***	0.1959***	0.2506***	0.2901***	0.2917***
log_R&D difference	-0.2228***	-0.2192***	-0.2162***	-0.2212***	-0.2239***	-0.2226***	-0.2216***	-0.2212***	-0.2191***
Period control	No	No	No	No	No	No	Yes	Yes	Yes
Therap. area controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-1.9071***	-1.7038***	-1.6733***	-1.4110***	-1.2958***	-1.2376***	-2.1344***	-1.9770***	-1.9356***
Zero-inflated part (logit)									
log_distw	0.3630***	0.3425***	0.3376***	0.4788***	0.4613***	0.4566***	0.4160***	0.3957***	0.3915***
log_PreVExperience	-1.8792***	-1.8667***	-1.8573***	-1.6971***	-1.7057***	-1.6959***	-1.7868***	-1.7818***	-1.7743***
log_PreVSpecialCorr	-2.0335***	-1.9663***	-1.9492***	-2.6936***	-2.6561***	-2.6371***	-2.2810***	-2.2185***	-2.1823***
log_InnoInstCorr	-7.8731***	-8.0024***	-7.6861***	-4.3223***	-4.7304***	-4.7723***	-6.0175***	-6.3745***	-6.3392***

**Table 1** continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	1998–2002	1998–2002	1998–2002	2004–2008	2004–2008	2004–2008	Pooled	Pooled	Pooled
Constant	-2.4453***	-2.3296***	-2.3000***	-4.0290***	-3.9320***	-3.8948***	-3.2378***	-3.1200***	-3.0906***
Inalpha	0.2157***	0.2275***	0.2335***	0.1423***	0.1472***	0.1443***	0.2238***	0.2314***	0.2332***
N	15,703	15,703	15,703	14,712	14,712	14,712	30,415	30,415	30,415
AIC	30,159.8073	30,088.2344	30,071.3549	41,543.7634	41,458.5057	41,453.1069	72,112.1113	71,952.2177	71,927.1869
BIC	30,389.6556	30,341.0675	30,354.8344	41,771.6560	41,709.1876	41,734.1744	72,370.1147	72,235.1892	72,243.4492
Log pseudolikelihood	-15.049.904	-15.011.117	-14.998.677	-20,741.882	-20,696.253	-20,689.553	-36,025.056	-35,942.109	-35,925.593

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

The results for institutional proximity measured by *log\_InnoInstCorr* do not indicate a robust and significant positive association with the number of international research collaborations since the coefficient of this variable is not significant and has a negative sign in many specifications. Based on this finding we reject Hypothesis 3 suggesting a positive relationship between institutional proximity and the number of international research collaborations. Hence, we do not find evidence that proximity among countries in terms of their institutional setting related to R&D and innovative activities facilitates interaction and joint research projects. On the opposite, some results of the analysis seem to suggest that more collaboration occurs between countries with different institutional settings.

With respect to our measures for cultural proximity we find that links among the former colonizer and its colonies (*colony*) show a quite robust positive association with the number of research collaborations on the level of therapeutic areas. This finding is less robust for the aggregate level (Table 7). Hence, this finding provides weak evidence that collaborations between former colonial powers and their former colonies are facilitated by similarities in the (informal) institutions based on past colonial ties.<sup>8</sup>

One can argue that collaboration between nations that were part of one country in the past may be facilitated since there should be linguistic as well as cultural links and knowledge about each others' informal institutions. However, our results for *smctry* do not suggest a significantly positive relation of a joint history within one country. A common language (*comlang\_ethno*) between two countries is positively associated with the number of collaborated research articles. One could have expected that a common language is less important for international research collaboration since English has a dominant position as the language of science and dominates by far the language of the articles in our dataset. However, our results show that researchers may prefer to discuss scientific problems with their collaboration partners in their mother tongue. The aggregation of these individual preferences and collaboration decisions provides an explanation for the observed positive relationship between a common language and the collaboration intensity on the cross-country level.

The association between different dimensions of proximity may be influenced by the presence of a common language as an enabling factor for collaboration. Hence, we include the interaction terms between common language and the geographical, social, cognitive and institutional dimensions of proximity in our analysis. The interaction term of a common language and geographical distance (*langdistw*) is positive and significant in most specifications of the gravity model. This finding indicates that a common language may help to overcome the challenges and obstacles of long distance collaboration expressed by the negative relation between geographical distance and collaboration intensity. The common language dummy (*comlang\_ethno*) itself is not significant anymore in most specifications that include interaction terms. Our results do not suggest a robust and significant association between the interaction terms of language with social proximity (*langExp*), cognitive proximity (*distSpecial*), institutional proximity (*langInnoInst*), and the number of international research collaborations.

Our results for cultural proximity partly support Hypothesis 5. Linguistic ties are positively related to cross-country research collaboration. However, common language does not show a significantly positive relation to the collaboration intensity on the country level when interaction terms with different dimensions of proximity are introduced into

<sup>8</sup> Another interpretation may stress that the former colonizers may want to support scientific and economic development in their former colonies. Therefore, governments may encourage scientists to collaborate with researchers based in former colonies of the respective country.



analysis. We find weak evidence that a joint colonial history is positively related to international research collaboration.

The dummy for contiguous countries (*contig*) is not significantly associated with the amount of research collaborations, if we do not distinguish between therapeutic areas. The coefficient for *SmlIncomeGr*, indicating whether two countries are part of the same income group, is not significant in some specifications that do not distinguish among therapeutic areas. Nevertheless, we may interpret the result as weak evidence that research collaboration predominantly takes places between countries that have a similar level of economic and scientific development, such as Western Europe, the U.S. and Japan. Differences in countries' overall R&D spendings measured as percentage of the GDP (*log\_R&DDifference*) are significantly negative related to cross-country research collaborations in pharmaceuticals. This result may indicate that research collaboration is more intense among countries that share a similar level of scientific and technological knowledge. In such an environment, organizations from different countries may have the absorptive capacity required for mutual understanding and successful research collaboration.

To summarize, our results suggest that geographical distance is significantly and negatively related to the collaboration intensity between countries, whereas the amount of previous collaborations (social proximity) is positively related to the number of cross-country collaborations. Hence, our results support Hypotheses 1 and 4. With respect to the cognitive dimension of proximity, we do not find evidence for a robust and significant relation to the intensity of international research collaborations. Therefore, we reject Hypothesis 2. Similarly, we reject Hypothesis 3 since our estimations do not suggest a robust significant relationship between institutional proximity and the number of research collaborations. Our results for cultural proximity do not allow for unambiguous conclusions concerning its influence on the collaboration intensity between countries and have to be analyzed in more detail as we do not find unambiguous support for Hypothesis 5.

### Robustness checks

These results stay qualitatively the same if we restrict our sample to positive cross-country collaboration counts and if we do not add 1 to the measure for social proximity. We account for the possibility that our results are driven by collaboration among developed countries by excluding all country pairs with at least one OECD country involved from our analysis, we find that our results stay qualitatively the same in this case, too. Furthermore, there might be concerns that the same factors that drive the number of research collaboration in period  $t$  may have driven it in  $t - 1$  as well. While it is beyond the scope of the paper to address the issue of co-determination in detail, we account for this concern by estimating our models without including the number of previous collaborations as proxy for social proximity. The results are qualitatively similar to those obtained above. Since previous research has shown that being part of a multinational research area such as the European Union influences international scientific collaboration (e.g., Mattsson et al. 2008), we use a dummy indicating whether two countries belong to the European Union or are connected to it through a signed science and technology agreement as alternative proxy for institutional proximity.<sup>9</sup> We construct this dummy based on the EU-15 as well as on the EU-27 countries. The regression results stay qualitatively the same for this alternative proxy for institutional proximity. In particular, we do not find evidence that being part of

<sup>9</sup> A list of science and technology agreements signed by the European Union can be obtained from the European Commission's web site: <http://ec.europa.eu/research/iscp/index.cfm?lg=en&pg=countries>.

the EU or having signed a science and technology agreement with the EU is positively related to the number of cross-country research collaborations. The use of the average number of R&D employees per one million inhabitants as a different control for R&D orientation of countries provides qualitatively similar results.

As a further robustness check we use log-linearized models to estimate different specifications of our gravity model. More precisely, we use the natural logarithm of the number of collaborations plus one as dependent variable and estimate OLS and Tobit regression models. In the case of OLS, we also estimate the models without adding one to the number of collaborations. The results of these models are similar to those of the count data models.

There are noticeable differences in the results reported for different time periods. For example, the magnitude of the coefficient for geographical distance seems to increase over time as in Hoekman et al. (2010) while it seems to decrease for the common language dummy. Another example is institutional proximity which becomes significant in period 2, but is not significant in period 1. These differences may suggest that not only the number of collaborations increases over time, but also the type of collaborations changes. However, we should also keep in mind that the data was separated into two periods due to noticeable differences in its structure. Therefore, a more profound analysis of these changes might be necessary before sound conclusions about these changes are drawn.

## Conclusion

In this paper we tested empirically the determinants of cross-country collaborations in pharmaceutical research. We focused our attention on different dimensions of proximity in order to explain the intensity of international collaborations. As our results suggest, some dimensions of proximity are important in explaining collaboration on the country level. For example, countries close in the geographical dimension of proximity show higher levels of international collaboration. Our findings give further support to previous studies on the macro-level scientific collaboration between OECD countries (Choi 2012) as well as to contributions on the regional and organizational level (see e.g., Acosta et al. 2011; Balland 2012; Hoekman et al. 2010; Frenken et al. 2009).

With respect to cognitive proximity, our results do not indicate a robust association with the number of international research collaborations. This dimension of proximity has been found to be relevant in some studies on the regional and on the firm level (e.g., Scherngell and Barber 2009, 2011; Scherngell and Hu 2011; Cantner and Meder 2007), while Balland (2012) does not find evidence for a significant role of cognitive proximity on the organizational level. For institutional proximity, our empirical analysis does not suggest a significant positive relationship to international research collaborations. This result contradicts earlier contributions suggesting that similar institutional settings, as in the case of EU-countries, could promote scientific collaboration (Mattsson et al. 2008).

Social proximity measured by previous collaborations among countries is positively related to the intensity of international research collaborations. This finding suggests that collaborations between countries rely upon some past experience, which was previously stressed in some micro-level research (e.g., Broekel and Boschma 2012; Agrawal et al. 2006).

We introduce different proxies for cultural proximity, such as colonial and linguistic links, as well as being a part of the same country in the past. We find weak evidence for a positive association between colonial ties among the former colonizer and its colonies and the amount of cross-country research collaboration. This finding supports descriptive evidence stressing the important role of former colonial relationships for scientific

(co-)publications and intraregional scientific collaboration in Central Africa (Boshoff 2009). However, we do not find a significant evidence that research collaboration is enhanced between countries formerly belonging to the same country.

According to our analysis, linguistic linkages enhance collaboration intensity among countries. This result is in line with the regional study by Hoekman et al. (2010) as well as the descriptive evidence by Zitt et al. (2000) and Schubert and Glänzel (2006). However, linguistic proximity might be a less important driver for scientific collaboration between OECD countries (Choi 2012). We further contribute to this literature by including interaction terms between language and different dimensions of proximity in the analysis. The coefficient of the common language dummy loses its significance in most specifications when we introduce interaction terms. We find a significantly positive interaction term of common language and geographical distance. This result may indicate that a common language may counterbalance the negative relation between geographical distance and the collaboration intensity to some extent. In other words, different dimensions of proximity may be seen as substitutes (Ponds et al. 2007; Boschma 2005): language similarities may encourage collaboration even over long distances. The interaction terms of a common language among countries with social proximity, cognitive proximity and institutional proximity do not indicate a robust significant relation to the number of international research collaborations.

The negative relationship between differences in countries' overall R&D spendings and international scientific collaboration are largely supported by evidence on the regional level provided by Acosta et al. (2011). In contrast to studies focusing on scientific collaborations among advanced economies, e.g., Choi (2012) and Acosta et al. (2011), we find evidence for countries' income affecting international scientific collaboration. This finding is in-line with firm-level evidence suggesting that most R&D-partnerships are formed between companies from high income countries, particularly from North America, Western Europe, Japan, and South Korea (Hagedoorn 2002).

Notwithstanding the need for further research addressing the relation between cultural proximity and international research collaboration with more detailed data on cultural linkages among countries, our findings may provide some recommendations for policy makers. Improved language education for students in schools and universities as well as exchange programs for students and researchers may help to overcome to some extent the challenges and obstacles of geographical distance and help to establish and maintain international research collaboration. The further strengthening of existing linguistic ties among parts of the population of two countries may help to intensify research collaboration.

Our study should be supplemented by additional investigations addressing the determinants of cross-country research collaboration in a wider set of research fields. Another limitation of this study arises from our dataset, which does not allow of taking policy programs established to stimulate international research collaboration explicitly into account. The objectives of these programs may be quite diverse. They may encompass the establishment of an integrated research area in the case of the European Union as well as the support of scientific research in developing countries.

Moreover, future research may study international research collaboration on disaggregated levels of analysis. Different types of organizations, e.g., universities and public research institutions as well as firms, may differ in their collaboration patterns. In contrast to universities, firms may be more likely to engage in international research collaboration that takes place within one organization with R&D facilities in different countries. This mode of international research collaboration differs from collaboration involving different organizations. Future research should therefore address different types of international scientific collaborations on the firm level. Scientific collaboration may be driven by

individual researchers rather than by inter-organizational agreements. Hence, data on individual researchers' publications and their career paths may help to further analyze the role of different dimension of proximity for scientific collaboration in terms of co-publication.

**Acknowledgements** This research was partly done while the authors were members of the Graduate College “The Economics of Innovative Change” at the Friedrich Schiller University Jena. We thank the German Science Foundation (DFG) and the DIME network for financial support. We are grateful to Uwe Cantner, the participants of the Jena Economic Research Workshop in February 2011, especially to Ljubica Nedelkoska, Fang Wang and Sebastian Wilfling, the participants of the DRUID-DIME Academy Winter Conference 2011, particularly Lars Alkærising and Tim Pohlmann, and the participants of the DIME Final Conference 2011 and to an anonymous referee for useful comments, expressed interest, and concerns. We thank Bart Leten for helping with the CHI classification. The usual caveats apply.

## Appendix 1: List of countries, therapeutic areas and description of variables

See Tables 2, 3, and 4.

**Table 2** List of countries

Algeria	Finland	Lithuania	Saudi Arabia
Argentina	France	Luxembourg	Senegal
Armenia	French Polynesia	Madagascar	Seychelles
Australia	Gabon	Malawi	Singapore
Austria	Gambia	Malaysia	Slovak Republic
Azerbaijan	Georgia	Mali	Slovenia
Bangladesh	Germany	Mexico	South Africa
Belarus	Ghana	Moldova	Spain
Belgium	Greece	Mongolia	Sri Lanka
Belize	Guatemala	Morocco	Sudan
Bolivia	Guinea	Mozambique	Sweden
Bosnia and Herzegovina	Hong Kong China	Namibia	Switzerland
Brazil	Hungary	Nepal	Tanzania
Bulgaria	Iceland	Netherlands	Thailand
Cameroon	India	New Zealand	Trinidad and Tobago
Canada	Indonesia	Niger	Tunisia
Chile	Iraq	Nigeria	Turkey
China	Ireland	Norway	Uganda
Colombia	Israel	Oman	Ukraine
Costa Rica	Italy	Pakistan	United Arab Emirates
Croatia	Jamaica	Panama	United Kingdom
Cuba	Japan	Papua New Guinea	United States
Cyprus	Jordan	Peru	Uruguay
Czech Republic	Kazakhstan	Philippines	Venezuela
Denmark	Kenya	Poland	Vietnam
Dominica	Korea Rep	Portugal	Zambia
Ecuador	Kuwait	Romania	Zimbabwe
Estonia	Latvia	Russian Federation	
Fiji	Lebanon	Rwanda	

**Table 3** List of therapeutic areas

Therapeutic area	Therapeutic area ID
Cancer	1
Cardiovascular	2
Central Nervous System	3
Dermatology	4
Eye and ear	6
Gastrointestinal tract	7
Genitourinary system	8
Hematological	9
HIV infections	10
Hormonal systems	11
Immune system	12
Infectious diseases	13
Musculoskeletal	15
Pain	16
Respiratory	17

**Table 4** Description of variables

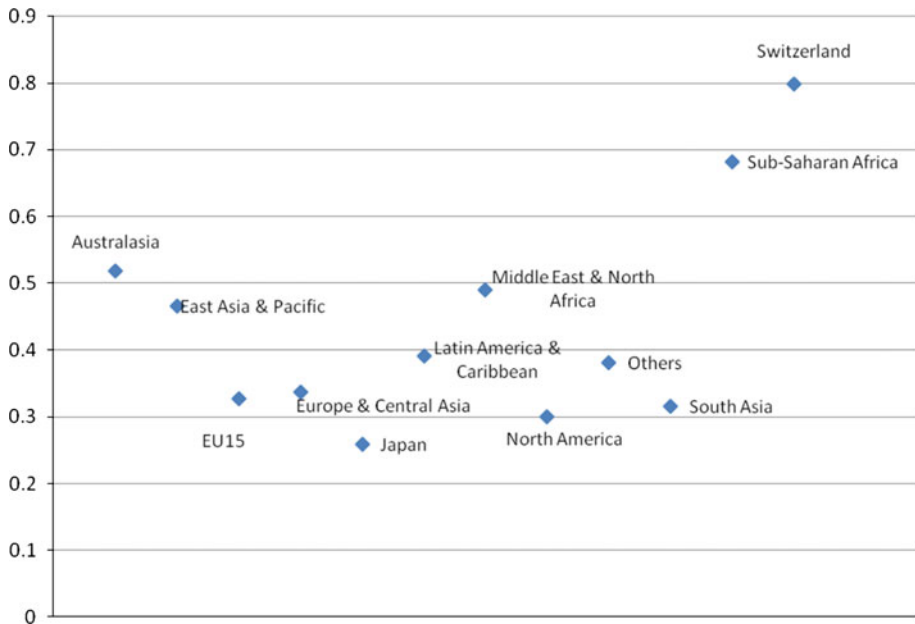
<i>Dependent variable</i>		
Collaborations		Number of collaborations between two countries
<i>Explanatory variables</i>	<i>Proximity</i>	
log_PubActor	Mass	Log of the number of publications of the actor country
log_PubPartner	Mass	Log of the number of publications of the partner country
log_distw	Geographical	Log of population weighted geographic distance
log_PrevExperience	Social	Log of the number of previous collaborations between countries
log_PrevSpecialCorr	Cognitive	Log of the cosine of country vectors containing the share of publications per therapeutic area prior to the analyzed period
log_InnoInstCorr	Institutional	Log of the cosine of country vectors containing countries' institutional characteristics related to R&D activities and innovation
comlang_ethno	Cultural	Equals 1 if at least 9 % of the population in both countries share the same language
colony	Cultural	Equals 1 if two countries had ever a colonial link
smctry	Cultural	Equals 1 if two countries were part of the same country during their history
langdistw		Interaction term between comlang_ethno and log_distw
langExp		Interaction term between comlang_ethno and log_PrevExperience
langSpecial		Interaction term between comlang_ethno and log_PrevSpecialCorr
langInnoInst		Interaction term between comlang_ethno and log_InnoInstCorr

**Table 4** continued

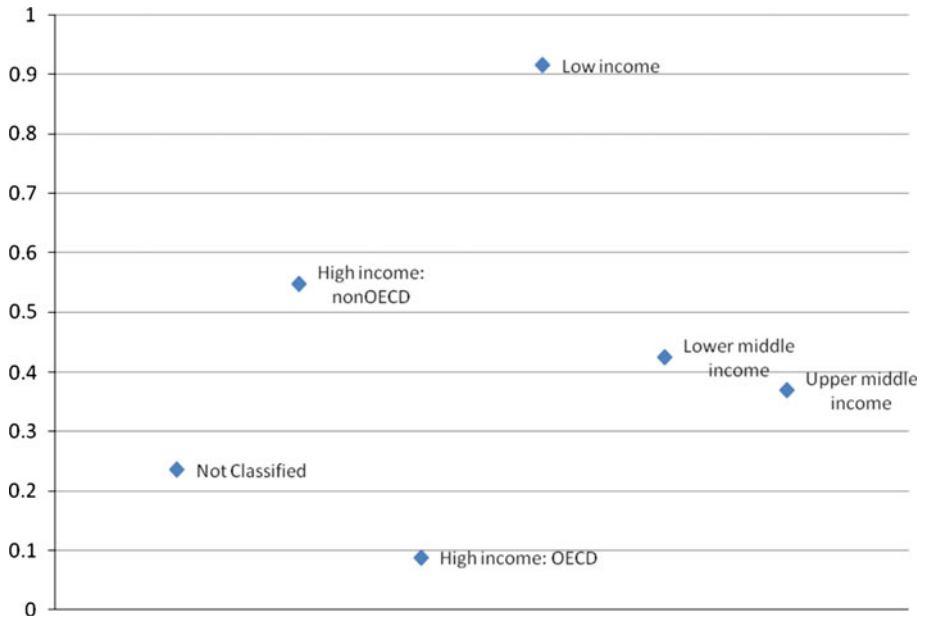
<i>Dependent variable</i>	
Collaborations	
Number of collaborations between two countries	
<i>Explanatory variables</i>	<i>Proximity</i>
<i>Controls</i>	
contig	Equals 1 if two countries are contiguous
SmIncomeGr	Equals 1 if two collaborating countries belong the same Worldbank income group
log_R&DDifference	Log of the difference in countries' R&D spendings as percentage of GDP
PeriodControl	Equals 1 if collaboration is observed in period 2
Therap. area controls	Dummy variables for the different therapeutic areas

**Appendix 2: Descriptives, correlations, and additional regressions**

See Figs. 2, 3 and Tables 5, 6, and 7.



**Fig. 2** Share of cross-country collaborations per region



**Fig. 3** Share of cross-country collaborations per income group

**Table 5** Cross-correlation table without therapeutic areas distinction

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) log_PubActor	1.0000						
(2) log_PubPartner	0.0742***	1.0000					
(3) log_distw	-0.0107**	-0.0220***	1.0000				
(4) log_PrevExperience	0.3374***	0.3379***	-0.3119***	1.0000			
(5) log_PrevSpecialCorr	0.2120***	0.2118***	-0.1694***	0.1950***	1.0000		
(6) log_InnovInstCorr	0.1016***	0.1004***	-0.1998***	0.2008***	0.2543***	1.0000	
(7) comlang_ethno	0.0339***	0.0257***	0.0232***	0.0599***	-0.0098**	0.0127**	1.0000
(8) colony	0.0653***	0.0751***	-0.0431***	0.0967***	0.0328***	0.0207***	0.2448***
(9) smctry	-0.0256***	-0.0297***	-0.2072***	0.0115**	0.0208***	0.0283***	0.0807***
(10) contig	0.0289***	0.0398***	-0.3121***	0.1087***	0.0705***	0.0467***	0.1505***
(11) SmlIncomeGr	0.1097***	0.1095***	-0.3114***	0.3182***	0.2285***	0.2715***	0.0144***
(12) log_R&DDifference	0.0828***	0.0810***	0.4367***	-0.4338***	-0.1276***	-0.1682***	0.0444***
(13) PeriodControl	-0.0462***	-0.0453***	0.0027	-0.0084*	-0.0379***	-0.0236***	-0.0047

	(8)	(9)	(10)	(11)	(12)	(13)
(8) colony	1.0000					
(9) smctry	0.0985***	1.0000				
(10) contig	0.1842***	0.3561***	1.0000			
(11) SmlIncomeGr	0.0073	0.0572***	0.1406***	1.0000		
(12) log_R&DDifference	0.0237***	-0.0191***	-0.0050	-0.3018***	1.0000	
(13) PeriodControl	-0.0039	0.0031	-0.0055	-0.0218***	-0.0023	1.0000

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$



**Table 6** Cross-correlation table with distinguishing therapeutic areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) log_PubActor	1.0000						
(2) log_PubPartner	0.0742***	1.0000					
(3) log_distw	-0.0107**	-0.0220***	1.0000				
(4) log_PrevExperience	0.3374***	0.3379***	-0.3119***	1.0000			
(5) log_PrevSpecialCorr	0.2120***	0.2118***	-0.1694***	0.1950***	1.0000		
(6) log_InnoInstCorr	0.1016***	0.1004***	-0.1998***	0.2008***	0.2543***	1.0000	
(7) comlang_ethno	0.0339***	0.0257***	0.0232***	0.0599***	-0.0098**	0.0127**	1.0000
(8) colony	0.0653***	0.0751***	-0.0431***	0.0967***	0.0328**	0.0207***	0.2448***
(9) smctry	-0.0256***	-0.0297***	-0.2072***	0.0115**	0.0208***	0.0283***	0.0807***
(10) contig	0.0289***	0.0398***	-0.3121***	0.1087***	0.0705**	0.0467***	0.1505***
(11) SmlIncomeGr	0.1097***	0.1095***	-0.3114***	0.3182***	0.2285***	0.2715***	0.0144***
(12) log_R&DDifference	0.0828***	0.0810***	0.4367***	-0.4338***	-0.1276***	-0.1682***	0.0444***
(13) PeriodControl	-0.0462***	-0.0453***	0.0027	-0.0084*	-0.0379***	-0.0236***	-0.0047

	(8)	(9)	(10)	(11)	(12)	(13)
(8) colony	1.0000					
(9) smctry	0.0985***	1.0000				
(10) contig	0.1842***	0.3561***	1.0000			
(11) SmlIncomeGr	0.0073	0.0572***	0.1406***	1.0000		
(12) log_R&DDifference	0.0237***	-0.0191***	-0.0050	-0.3018***	1.0000	
(13) PeriodControl	-0.0039	0.0031	-0.0055	-0.0218***	-0.0023	1.0000

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

**Table 7** Zero-inflated negative binomial regression models without therapeutic area distinction

	(1) 1998–2002	(2) 1998–2002	(3) 1998–2002	(4) 2004–2008	(5) 2004–2008	(6) 2004–2008	(7) Pooled	(8) Pooled	(9) Pooled
Dependent variable: number of collaborations									
log_PubActor	0.3244***	0.3691***	0.3711***	0.3546***	0.3739***	0.3739***	0.3418***	0.3694***	0.3714***
log_PubPartner	0.3249***	0.3599***	0.3673***	0.3797***	0.3970***	0.3983***	0.3585***	0.3821***	0.3862***
log_distw	-0.1290***	-0.1891***	-0.1871***	-0.2701***	-0.3101***	-0.3231***	-0.2028***	-0.2516***	-0.2604***
log_PrevExperience	0.4832***	0.4328***	0.4592***	0.4695***	0.4390***	0.4496***	0.4766***	0.4388***	0.4563***
log_PrevSpecialCorr	-0.0624	-0.0078	-0.0889	-0.3135*	-0.3329*	-0.4963**	-0.1681	-0.1424	-0.2446*
log_InnoInstCorr	0.9482	0.7657	2.2511	-1.5117	-1.6946	-1.7106	-0.4561	-0.7659	-0.3481
comlang_ethno		0.6815***	-0.0154		0.3640***	-0.7059		0.5029***	-0.4313
colony		-0.0693	-0.0082		0.2666**	0.2553**		0.1054	0.1277
smctry		0.0979	0.1520		-0.2338	-0.1624		-0.0641	0.0225
langdistw			0.1023			0.1614**			0.1396**
langExp			-0.1683***			-0.0658			-0.1147***
langSpecial			0.2423			0.7688**			0.3610
langInnoInstCorr			-6.1043			0.4017			-1.6152
contig	0.3356*	0.1378	0.2745	0.0034	-0.0625	0.0155	0.1603	0.0353	0.1427
SmIncomeGr	0.2725***	0.3395***	0.3265***	-0.0063	0.0200	0.0192	0.1307**	0.1672***	0.1590**
log_R&DDifference	-0.1359***	-0.1432***	-0.1390***	-0.1479***	-0.1542***	-0.1533***	-0.1422***	-0.1486***	-0.1463***
PeriodControl	No	No	No	No	No	No	Yes	Yes	Yes
Constant	-2.2460***	-2.3596***	-2.4568***	-0.2243	-0.1439	-0.0948	-2.0338***	-2.0385***	-2.0468***
Zero-inflated part (logit)									
log_distw	0.3467***	0.3167***	0.3109***	0.7677***	0.7342***	0.7281***	0.5668***	0.5319***	0.5258***
log_PrevExperience	-1.5126***	-1.5235***	-1.4754***	-1.7943***	-1.8086***	-1.7951***	-1.6530***	-1.6545***	-1.6270***
log_PrevSpecialCorr	-0.3968**	-0.3577*	-0.4013**	-2.2566***	-2.2554***	-2.3123***	-1.0612***	-1.0301***	-1.0691***
log_InnoInstCorr	-6.8214**	-6.4818**	-5.7732**	-8.1065***	-8.1117***	-8.0390***	-7.4628***	-7.4858***	-7.2688***
Constant	-2.6047***	-2.4169***	-2.3726***	-7.4379***	-7.1826***	-7.1422***	-5.0682***	-4.8153***	-4.7768***

**Table 7** continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	1998–2002	1998–2002	1998–2002	2004–2008	2004–2008	2004–2008	Pooled	Pooled	Pooled
Inalpha	-0.2250*	-0.2206*	-0.2600**	-0.2267***	-0.2393***	-0.2530***	-0.1856***	-0.1929***	-0.2052***
N	2820	2820	2820	2866	2866	2866	5686	5686	5686
AIC	8356.0223	8307.8402	8291.5678	11295.1448	11266.9568	11263.0026	19732.4595	19660.0275	19643.7291
BIC	8451.1342	8420.7855	8428.2912	11390.5156	11380.2096	11400.0980	19845.4375	19792.9427	19803.2274
Log pseudolikelihood	-4162.011	-4134.920	-4122.784	-5631.572	-5614.478	-5608.501	-9849.230	-9810.014	-9797.865

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## References

- Acosta, M., Coronado, D., Ferrádiz, E., & León, M. (2011). Factors affecting inter-regional academic scientific collaboration within Europe: The role of economic distance. *Scientometrics*, 87(1), 63–74.
- Adams, J.D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from U.S. Universities, 1981–1999. *Research Policy*, 34(3), 259–285.
- Agrawal, A., Cockburn, I., & McHale, J. (2006). Gone but not forgotten: Knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*, 6(5), 571–591.
- Asheim, B., Coenen, L., & Vang, J. (2007). Face-to-face, buzz, and knowledge bases: Sociospatial implications for learning, innovation, and innovation policy. *Environment and Planning C: Government and Policy*, 25(5), 655–670.
- Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review*, 86(3), 630–640.
- Autant-Bernard, C., Billand, P., Frachisse, D., & Massard, N. (2007). Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies. *Papers in Regional Science*, 86(3), 495–519.
- Balland, P.-A. (2012). Proximity and the evolution of collaboration networks: Evidence from research and development projects within the global navigation satellite system (GNSS) industry. *Regional Studies*, 46(6), 741–756. forthcoming.
- Bartholomew, S. (1997). National systems of biotechnology innovation: Complex interdependence in the global system. *Journal of International Business Studies*, 28(2), 241–266.
- Basile, R., Capello, R., & Caragliu, A. (2012). Technological interdependence and regional growth in Europe: Proximity and synergy in knowledge spillovers. *Papers in Regional Science*, 91(4), 697–722.
- Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39(1):61–74.
- Boschma, R., & Lambooy, J. G. (1999). Evolutionary economics and economic geography. *Journal of Evolutionary Economics*, 9(4), 411–429.
- Boshoff, N. (2009). Neo-colonialism and research collaboration in Central Africa. *Scientometrics*, 81(2), 413–434.
- Broekel, T., & Boschma, R. (2012). Knowledge networks in the Dutch aviation industry: The proximity paradox. *Journal of Economic Geography*, 12(2), 409–433.
- Burger, M., van Oort, F., & Linders, G.-J. (2009). On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis*, 4(2), 167–190.
- Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics*, 46(3), 347–364.
- Cantner, U., & Meder, A. (2007). Technological proximity and the choice of cooperation partner. *Journal of Economic Interaction and Coordination*, 2(1), 45–65.
- Choi, S. (2012). Core-periphery, new clusters, or rising stars? International scientific collaboration among ‘advanced’ countries in the era of globalization. *Scientometrics*, 90(1), 25–41.
- Coenen, L., Moodysson, J., & Asheim, B. T. (2004). Nodes, networks and proximities: On the knowledge dynamics of the Medicon Valley biotech cluster. *European Planning Studies*, 12(7), 1003–1018.
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128–152.
- de Crombrughe, D., Farla, K., Meisel, N., de Neubourg, C., Aoudia, J. O., & Szirmai, A. (2009). Institutional profiles database III—Presentation of the institutional profiles database 2009 (IPD 2009). *Documents des Travaux de la DGPE*, 2009/14.
- de Solla Price, D. (1963). *Little science, big science*. New York: Columbia University Press.
- Dosi, G., Pavitt, K., & Soete, L. (1990). *The economics of technical change and international trade*. New York: Harvester Wheatsheaf.
- Edquist, C., & Johnson, B. (1997). Institutions and organizations in systems of innovation. in *Systems of innovation: Technologies, institutions, and organizations* (pp. 41–63). London: Pinter.
- Felbermayr, G. J., & Toubal, F. (2010). Cultural proximity and trade. *European Economic Review*, 54(2), 279–293.
- Flowerdew, R., & Aitkin, M. (1982). A method of fitting the gravity model based on the Poisson distribution. *Journal of Regional Science*, 22(2), 191–202.
- Frenken, K., Hoekman, J., Kok, S., Ponds, R., van Oort, F., & van Vliet, J. (2009). Death of distance in science? A gravity approach to research collaboration. In A. Pyka, & A. Scharnhorst (Eds.), *Innovation networks* (pp. 43–57). Berlin: Springer.
- Furman, J. L., Porter, M. E., & Stern, S. (2002). The determinants of national innovative capacity. *Research Policy*, 31(6), 899–933.

- Gertler, M. S. (1995). “Being there”: Proximity, organization, and culture in the development and adoption of advanced manufacturing technologies. *Economic Geography*, 71(1), 1–26.
- Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *The American Journal of Sociology*, 91(3), 481–510.
- Greene, W. H. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. *NYU Working Paper*, No. EC-94-10.
- Hagedoorn, J. (2002). Inter-firm R&D partnerships: An overview of major trends and patterns since 1960. *Research Policy*, 31(4), 477–492.
- Hamilton, K. S. (2003). Subfield and level classification of journals. *CHI Research Inc.*, CHI No. 2012-R.
- Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics*, 123(2), 441–487.
- Hoekman, J., Frenken, K., & Tijssen, R. J. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, 39(5), 662–673.
- 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.
- Hoekman, J., Scherngell, T., Frenken, K., & Tijssen, R. (2013). Acquisition of European research funds and its effect on international scientific collaboration. *Journal of Economic Geography*, 13(1), 23–52.
- Howells, J. R. L. (2002). Tacit knowledge, innovation and economic geography. *Urban Studies*, 39(5–6), 871–884.
- Isard, W. (1954). Location theory and trade theory: Short-run analysis. *The Quarterly Journal of Economics*, 68(2), 305–320.
- Jaffe, A. B. (1986). Technological opportunity and spillovers of R & D: Evidence from firms’ patents, profits, and market value. *The American Economic Review*, 76(5), 984–1001.
- Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3), 577–598.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1), 1–18.
- Kirat, T., & Lung, Y. (1999). Innovation and proximity: Territories as loci of collective learning processes. *European Urban and Regional Studies*, 6(1), 27–38.
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, 11(1), 3–15.
- Lundberg, J., Tomson, G., Lundkvist, I., Skar, J., & Brommels, M. (2006). CCollaboration uncovered: Exploring the adequacy of measuring university–industry collaboration through co-authorship and funding. *Scientometrics*, 69(3), 575–589.
- Lundvall, B. (1988). Innovation as an interactive process: From user–producer interaction to the national system of innovation. In: G. Dosi, C. Freeman, R. Nelson, G. Silverberg, & L. Soete (Eds.), *Technical change and economic theory* (pp. 349–369). London: Pinter.
- Lundvall, B.-A. (1992). *National systems of innovation: Towards a theory of innovation and interactive learning*. London: Pinter.
- Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology & Human Values*, 17(1), 101–126.
- Mattsson, P., Laget, P., Nilsson, A., & Sundberg, C.-J. (2008). Intra-EU vs. extra-EU scientific co-publication patterns in EU. *Scientometrics*, 75(3), 555–574.
- Mayer, T., & Zignago, S. (2006). *Notes on CEPII’s distances measures*. CEPII (Centre d’Études Prospectives et d’Information Internationales), Paris.
- Melitz, J. (2008). Language and foreign trade. *European Economic Review*, 52(4), 667–699.
- Miquel, J., & Okubo, Y. (1994). Structure of international collaboration in science, part II: Comparisons of profiles in countries using a link indicator. *Scientometrics*, 29(2), 271–297. doi:[10.1007/BF02017977](https://doi.org/10.1007/BF02017977).
- Nagpaul, P. (2003). Exploring a pseudo-regression model of transnational cooperation in science. *Scientometrics*, 56(3), 403–416.
- Narin, F., Stevens, K., & Whitlow, E. (1991). Scientific co-operation in Europe and the citation of multi-nationally authored papers. *Scientometrics*, 21(3), 13–323.
- Nooteboom, B. (2000). *Learning and innovation in organizations and economies*. Oxford: Oxford University Press.
- Peri, G. (2005). Determinants of knowledge flows and their effect on innovation. *Review of Economics and Statistics*, 87(2), 308–322.
- Ponds, R., van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science*, 86(3), 423–443.
- Santos Silva, J., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4), 641–658.
- Scherngell, T., & Barber, M. J. (2009). Spatial interaction modelling of cross-region R&D collaborations: Empirical evidence from the 5th EU Framework Programme. *Papers in Regional Science*, 88(3), 531–546.

- Scherngell, T., & Barber, M. J. (2011). Distinct spatial characteristics of industrial and public research collaborations: Evidence from the fifth EU Framework Programme. *The Annals of Regional Science*, 46(2), 247–266.
- Scherngell, T., & Hu, Y. (2011). Collaborative knowledge production in China: Regional evidence from a gravity model approach. *Regional Studies*, 45(6), 755–772.
- Schubert, A., & Glänzel, W. (2006). Cross-national preference in co-authorship, references and citations. *Scientometrics*, 69(2), 409–428.
- Shan, W., & Hamilton, W. (1991). Country-specific advantage and international cooperation. *Strategic Management Journal*, 12(6), 419–432.
- Tinbergen, J. (1962). *Shaping the world economy: Suggestions for an international economic policy*. New York: The Twentieth Century Fund.
- Torre, A., & Gilly, J.-P. (2000). On the analytical dimension of proximity dynamics. *Regional Studies*, 34(2), 169–180.
- Torre, A., & Rallet, A. (2005). Proximity and localization. *Regional Studies*, 39(1), 47–59.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2), 307–333.
- Wagner, C. S., & Leydesdorff, L. (2005). Mapping the network of global science: Comparing international co-authorships from 1990 to 2000. *International Journal of Technology and Globalisation*, 1(2), 185–208.
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–1039.
- Zeller, C. (2004). North Atlantic innovative relations of Swiss pharmaceuticals and the proximities with regional biotech arenas. *Economic Geography*, 80(1), 83–111.
- Zitt, M., Bassecoulard, E., & Okubo, Y. (2000). Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics*, 47, 627–657.