JENA ECONOMIC RESEARCH PAPERS

2011 – 026

Collaboration in pharmaceutical research: Exploration of country-level determinants.

by

Tatiana Plotnikova
Bastian Rake

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

Jena Economic Research Papers 2011 - 026

Collaboration in pharmaceutical research: Exploration of country-level determinants. *

Tatiana Plotnikova† and Bastian Rake‡

June, 2011

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. Moreover, there is no robust and significant relation between the interaction terms of geographical distance with social, cognitive, or institutional proximity, and international research collaboration. Our findings for cultural proximity do not allow of unambiguous conclusions concerning their influence on the collaboration intensity between countries. Linguistic ties among countries are associated with a higher amount of cross-country research collaboration but we find no clear association for historical and colonial linkages.

Keywords: International Cooperation, Pharmaceuticals, Proximity
JEL Classification: R10, O31

*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 Alkærsing and Tim Pohlmann, and the participants of the DIME Final Conference 2011 for useful comments, expressed interest, and concerns. We thank Bart Leten for helping with CHI classification. The usual caveats apply.

†Friedrich-Schiller-University, Jena, Graduate College "The Economics of Innovative Change" (DFG-GK-1411), Carl-Zeiss-Strasse 3, D-07743 Jena, Germany, phone: +49-(0)3641-943-276, e-mail: tatiana.plotnikova@gmail.com

‡Friedrich-Schiller-University Jena, Graduate College "The Economics of Innovative Change" (DFG-GK-1411), Carl-Zeiss-Strasse 3, D-07743 Jena, Germany, phone: +49-(0)3641-943-274, e-mail: bastian.rake@uni-jena.de, corresponding author

1 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 a rising number of authors 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 the number of authors of scientific papers originating from the 110 top U.S. research universities increased during the 19-year period from 1981 to 1999 by 50%. Using an extended dataset of 19.9 million papers and 2.1 million patents, Wuchty et al. (2007) show that the phenomenon of an increasing team size in knowledge production is particularly prevalent in science and engineering, in social sciences publications, as well as in patents, whereas it is much less evident in arts and humanities.

The general trend of an increasing number of team members in the production of knowledge is accompanied with a growing amount of international research. Institutional collaboration on the international level has become more important during the last decades as indicated by the growing number of collaborations in research. For instance, 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). Knowledge produced by international teams is more frequently cited than research whose authors are affiliated to different institutions in the same country, and it receives particularly more citations than papers originating in a single institution (Narin et al., 1991).

The increasing importance of the phenomena of international collaboration in scientific research motivated us to explore its determinants. In doing so, we apply the concept of multiple dimensions of proximity (cf. Boschma, 2005). More precisely, we aim to explain the intensity of international collaborations in pharmaceutical research by geographical, cognitive, institutional, social, and cultural proximity. Empirical analysis is performed on a sample of scientific journal publications related to pharmaceutical research using different count data specifications of the gravity model. Our findings suggest a significant negative relationship between geographical distance and the collaboration intensity between countries. The amount of previous collaborations, as a proxy for social proximity,

is positively related to the amount of cross-country collaborations. We do not find robust significant associations between cognitive proximity or institutional proximity with the number of international research collaborations. Moreover, there is no robust and significant relation between the interaction terms of geographical distance with social, cognitive or institutional proximity, and international research collaboration. Our findings for cultural proximity do not allow of unambiguous conclusions concerning their influence on the collaboration intensity between countries. Linguistic ties among countries are associated with a higher amount of cross-country research collaboration but we find no clear effect for historical and colonial linkages.

The remainder of this paper is structured as follows. We describe the tendencies of internationalization in pharmaceutical research as well as the determinants of international collaboration in Section 2. Section 3 provides details of the methodology and the data used in our empirical analysis. Section 4 reports the results of the analysis of international collaborations in the pharmaceutical research. Section 5 concludes.

2 Related Literature

2.1 Research collaboration in the pharmaceutical industry

The number of research collaborations in the pharmaceutical industry has increased after the emergence of biotechnology in the mid 1970s for several reasons. First, the nature of the industry's R&D process changed from random screening to guided research. This change has increasingly involved the application of scientific advances made by universities and public research organizations in the pharmaceutical companies' innovation process (McKelvey et al., 2004, Gambardella, 1995).

Second, many scientific achievements have been commercialized by new, specialized biotechnology companies. Therefore, in-house and collaborative research became two important ways to build up absorptive capacity, which is required to successfully incorporate knowledge produced in academia and the biotechnology sector in the R&D projects of pharmaceutical companies (Cockburn and Henderson, 1996, Gambardella, 1992). Consequently, private sector actors can be found among the important contributors to scientific publications in the relevant research areas, often in collaboration with academic partners

(Tijssen, 2004).
Third, as a result of the industry's expanding knowledge base, it became too broad to be available in all details to one single organization. Therefore, collaboration is seen now as a channel through which external knowledge can be accessed (cf. Orsenigo et al., 2001).

Collaboration in pharmaceutical R&D extends beyond national borders. For instance, European pharmaceutical companies not only have established R&D centers outside their home countries, but are connected to a considerable number of international research partners (Tijssen, 2009). Similarly, in almost one quarter of corporate research collaborations, firms and institutions from more than three world regions are involved (Calero et al., 2007).
This trend of internationalization of research collaboration has been promoted particularly by large technology-based multinationals intending to source knowledge at a global scale through the location of their R&D activities at a few technologically and scientifically leading global centers of excellence (Gassmann and von Zedtwitz, 1999). In this way, companies have established a network of specialized research activities across countries in order to get access to locally based technological and scientific expertise (Cantwell and Janne, 2000).

2.2 Determinants of Collaboration

Given the increasing importance of collaboration in the pharmaceutical industry, we ask what determines this phenomenon on the international level. Put differently, we aim to explain 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 (Ponds et al., 2007, Luukkonen et al., 1992). Ho-

wever, 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 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 the 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 social, 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 the life sciences (Hoekman et al., 2009, 2010, Ponds et al., 2007). We argue that this relation should hold also on the level of international collaborations. Therefore, our first hypothesis is

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

In an evolutionary perspective, knowledge creation and innovation often imply a high degree of tacit knowledge and can be seen as the cumulative outcome of search processes conducted by boundedly rational agents. The creation of new knowledge and the learning about existing knowledge depends in many instances on the combination of diverse but complementary capabilities within and between organizations (Nooteboom, 2000). The tacit and idiosyncratic component of knowledge implies that besides access, absorptive capacity enables actors 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 is required in the exchange and combination of knowledge. Cognitive proximity means the sufficient closeness of an actor's cognitive 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' knowledge stocks is associated with a higher probability of cooperation (Cantner and Meder, 2007). We would like to ask whether cognitive proximity has a similar importance in international collaborations.

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 refers to '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. Based on these arguments we build our hypothesis concerning the impact of institutional proximity on the country-level.

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

The notion of social proximity draws upon the embeddedness literature, which suggests that economic activities are embedded in a social context (Granovetter, 1985). The extent of this embeddedness is connected to organizational learning and innovative performance. Building upon these thoughts, social proximity is defined in terms of socially embedded relations, i.e., it involves trust based on friendship, kinship, and experience, between agents at the micro-level (Boschma, 2005). On the one hand, social proximity may enhance interactive learning and knowledge creation through mutual trust and commitment. On the other hand, there is the danger of lock-in and the risk of opportunism. On the firm level, it has been argued that firms often connect due to a common work experience of their

employees (Agrawal et al., 2006). We suspect that these results hold on the country-level, too.

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 behaviour (Zeller, 2004). Cultural proximity has been found to facilitate social proximity. The presence of cultural differences impedes the transmission and decoding of certain types of messages, especially if tacit components are involved (Lundvall, 1992). Empirical studies reveal that linguistic and historical ties influence the intensity of scientific collaboration (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.*

3 Data and Research Methodology

3.1 Gravity Model

We analyse 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 analyse the intensity of co-publications among regions (Hoekman et al., 2010, 2009). 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):

$$I_{ij} = \beta_0 \frac{M_i^{\beta_1} M_j^{\beta_2}}{d_{ij}^{\beta_3}} \tag{1}$$

I_{ij} denotes the the interaction intensity, i.e., the number of research collaborations, between countries i and j. β_0 is a proportionality constant. M_i and M_j represent the masses of country i and j, which are in our case the number of publications. The distance between the two countries is denoted by d_{ij}. $\beta_{1,2}$ reflect the potential to collaborate and β_3 reflects the effect of distance. The multiplicative form of the gravity model presented in (1) can be transformed to a testable linear model by taking the logarithms of both sides and adding a disturbance term ϵ_{ij}:

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

Concerning the estimation of (2) we have to take into account that we are dealing with count data. Hence, the OLS framework is not appropriate and we apply alternative regression techniques: the 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 μ which can be expressed as a function of the independent variables.

$$\Pr\left[I_{ij}\right] = \frac{\exp\left(-\mu_{ij}\right)\mu_{ij}^{I_{ij}}}{I_{ij}!}, (I_{ij} = 0, 1, \ldots) \tag{3}$$

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

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

In (4), β_0 is a constant, D_{ij} is the vector of explanatory variables representing different dimensions of distance, and β' 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):

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

where μ_{ij} is the conditional mean, Γ is the gamma function, and α is the parameter determining the degree of dispersion, allowing that the conditional variance exceeds the conditional mean. Larger α corresponds to a larger degree of overdispersion in the data.

With α 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 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 the value zero when there is no collaboration between the 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

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

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

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

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

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

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

3.2 Data

In our empirical analysis we use data from different sources. We start by drawing a list of 251 medical indications from BioPharmInsight.[1] Each indication represents a condition, disease or symptom. Each indication is exclusively assigned to one of 15 therapeutic areas.[2] 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 journals, books, book series, reports, and conferences. In the case of the Friedrich-Schiller-University of Jena, it is hosted by Thomson Reuters. Among these databases, the most important one is the Science Citation Index Expanded. It is multidisciplinary and indexes more than 6,500 scientific journals and covers 150 scientific disciplines. The Science Citation Index Expanded covers, among others scientific fields, biochemistry, medicine, and pharmacology, which are of particular interest for our study.
The WoS contains information concerning the scientific publications themselves, such 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 for most articles the authors' affiliations, including the country of the location of a respected organization.

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 & Molecular Biology", "Biotechnology and Applied Microbiology", "Chemistry, Applied", "Chemistry, Medicinal", "Medicine, Research & Experimental", "Pharmacology & Pharmacy", and "Toxicology" are included in our dataset.[3] We restrict our sample to journal articles and exclude publications that are labeled as meeting abstracts, editorials or reviews as well as

[1] http://www.infinata5.com/biopharm/

[2] Table 3 provides an overview of the therapeutic areas included in the dataset.

[3] The subcategories are described in detail at http://scientific.thomsonreuters.com/mjl/.

other non-journal publications. Conference proceedings have not been considered, either since they might 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, prior to 1998 information concerning the authors' affiliations was not included in the WoS 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 the authors' affiliations, we had a sample of 111,096 journal articles. 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 the different therapeutic areas prior to the periods of observation in our sample.

Information concerning the authors' affiliations is matched with WHO Regions and Worldbank income groups in order to include the geographical region a country is located in and the wealth level of the countries in our sample. 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 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. By using this classification scheme we can distinguish "clinical observation", "clinical mix", "clinical investigation", and "basic biomedical research".

We employ the CEPII (Centre d'Études Prospectives et d'Information Internationales) 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 the institutional environment in different countries, we use the "Index of Economic Freedom" for the years 1996 to 2008, created

by The Heritage Foundation and The Wall Street Journal.[4] The index aims at measuring the degree of economic freedom in ten subfields ranging from property rights to entrepreneurship for (currently) 183 countries.

The use of publication data implies the 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 publication 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), and Hoekman et al. (2009) for a discussion).

4 Empirical Results Gravity Regression Model

4.1 Variables and Descriptives

The dependent variable in the gravity model is the amount of collaboration between two countries.[5] 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 the 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. The

[4] http://www.heritage.org/index/

[5] Appendix A.1 provides a description of the variables.

variables *log_PubActor* and *log_PubPartner* represent the logs of these counts. Similar to the dependent variable, we consider publications per therapeutic area.
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 cumulated previous collaborations (*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. The variable *distExp* represents the interaction term between the population weighted geographical distance and our proxy for social proximity.

Our measure of cognitive proximity (*log_PrevSpecialCorr*) is based on the specialization profiles of countries among therapeutic areas prior to the analysed 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 analysed period. We calculate the uncentred correlation, which corresponds to the cosine, of these vectors for each country pair and take the log. We additionally include the interaction term between the population weighted geographic distance and our measure for cognitive proximity (*distSpecial*).

Our measure of institutional proximity (*log_PrevResTypeCorr*) is constructed along similar lines. We calculate the uncentred correlation of the country vectors containing the share of different research types according to the CHI classification, i.e., "clinical observation", "clinical mix", "clinical investigation", and "basic biomedical research", lag the respective vectors by one period and take the log. In doing so we account for differences

in the institutional environment of the research sectors among countries. Moreover, we include the interaction term between the population weighted geographical distance and our measure for institutional proximity in our analysis (*distResType*).

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*), had the same colonizer after 1945 (*comcol*), or were part of the same country (*smctry*). Moreover, we control for whether two countries are adjacent (*contig*) and belong to the same Worldbank income group (*SmIncomeGr*). Furthermore, we add a dummy indicating whether the collaboration took place in period 2 (2004 to 2008). When we distinguish the amount of collaboration among therapeutic areas, we add dummy variables for the 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[6]. There are considerable differences in the frequency of international collaboration among WHO regions and Worldbank income groups. In contrast to publications from Europe, North America and Japan, we find that articles from Switzerland and Sub-Saharan Africa show particularly 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.

4.2 Regression Results

We start our empirical analysis with Poisson regression models[7]. 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 specifica-

[6] See Figures 2 and 3 in the Appendix for a more detailed illustration.

[7] Results not presented in this paper are available upon request.

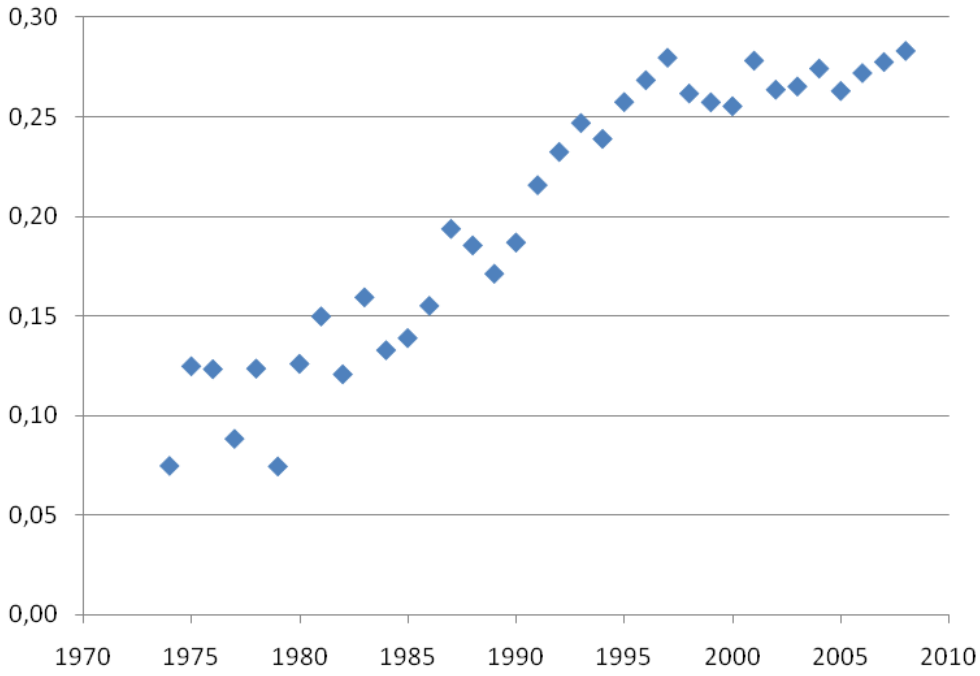

Figure 1: Share of Cross-Country Collaborations per Year

tions. Therefore, we account for the possibility of overdispersion by using robust standard errors and by applying negative binomial regressions, which have been established as the standard alternative to the Poisson model.

In Table 1 we distinguish the number of publications and collaborations according to the 15 therapeutic areas included in our study. As in many applications of count data, the data on cross-country research collaborations shows an excessive number of zeros. Without distinguishing among therapeutic areas, 10,975 out of 14,016 observations are zero counts, with the distinction 34,131 out of 42,999 observations are zeros. These are, respectively, 78.3% and 79.4%. We deal with this data structure by estimating a zero inflated version of the negative binomial model, presented in Table 2. The Vuong test (Vuong, 1989) suggests preferring the zero-inflated models over their ordinary 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 Baysian Information Criterion (BIC).

Tables 1 and 2 give the results of our empirical estimation. We test our hypothesis

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 a negative and highly significant association with the number of cross-country collaborations. In other words, we find that the 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, researchers collaborate more intensively with partners in countries that are close to their country of origin.

Our measure for social proximity, *log_PrevExperience*, is positively related to the amount of collaboration among country pairs. This finding indicates that previous experience through research collaborations with partners in the respective country is positively related to the extent of collaboration in the two periods of observation. Hence, hypothesis 4 cannot be rejected. 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 with them, since it reduces uncertainty and the risk of opportunism.

The coefficient for cognitive proximity (*log_PrevSpecialCorr*) changes its sign. It is positive, although not always significant, in the negative binomial models. In the zero-inflated models, however, we do not find a robust positive association since the coefficient is not always significant, and is negative in several 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 the same therapeutic areas. It may be that the cognitive 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.

The results for institutional proximity proxied by *log_PrevResTypeCorr* do not indicate a robust and significant positive association with the number of international research collaborations. The coefficient of this variable is not significant in most model specifications and changes its sign quite frequently. Based on this finding we reject hypothesis 3 suggesting a positive relationship between cognitive proximity and the number of international research collaborations. Our findings indicate no clear relation between similar institutional settings in the research sector, i.e., a specialization in a certain type of pharmaceutical research, and the intensity of collaboration. Hence, institutional proximity among countries may not necessarily facilitate interaction and joint research projects.

With respect to our measures for cultural proximity we find that 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.

We find inconclusive evidence concerning the variables accounting for colonial ties expressed by *colony* and *comcol* on the number of research collaborations. Similarities in countries' (informal) institutional settings based on a joint colonial history may facilitate collaboration, but we do not find support for this idea in our regression models. Along similar lines one could 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 robust, significant positive relation of a joint history within one country. With respect to hypothesis 5, we find support for the positive relation of cultural proximity on cross-country research collaboration only for linguistic ties. We find no robust evidence that colonial and other historical ties are positively related to international research collaboration.

Surprisingly, the dummy for contiguous countries (*contig*) seems to be negatively associated with the amount of research collaborations, particularly if we distinguish between therapeutic areas. The result may indicate that researchers may aim to connect to global centers of excellence which may not be located in an contiguous nation. The coefficient for *SmIncomeGr*, indicating whether two countries are part of the same income group, is positive and highly significant in all model specifications. This may indicate that countries on a similar stage of economic development share a similar knowledge base and therefore intensify collaboration. It may also be that similar economic conditions and livings standards are associated with the prevalence of similar diseases in the respective countries leading to intensified collaboration.

We additionally included the interaction terms between geographical distance and the social, cognitive and institutional dimensions of proximity in our analysis. Our results do not suggest a robust and significant association between the interaction terms, *distExp, distSpecial, distResType*, and the number of international research collaborations. Consequently, our results do not suggest a complementarity between geographical and other dimensions of proximity.

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 their influence on the collaboration intensity between countries and have to be analysed in more detail as we do not find unambiguous support for hypothesis 5.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	98-02	98-02	98-02	04-08	04-08	04-08	Pooled	Pooled	Pooled
Dependent Variable: Number of Collaborations									
log_PubActor	0.517***	0.522***	0.529***	0.549***	0.551***	0.560***	0.536***	0.540***	0.547***
log_PubPartner	0.500***	0.502***	0.509***	0.535***	0.535***	0.543***	0.522***	0.523***	0.529***
log_distw	-0.625***	-0.667***	-0.785***	-0.629***	-0.653***	-0.799***	-0.631***	-0.664***	-0.796***
log_PrevExperience	0.671***	0.638***	0.236*	0.633***	0.608***	0.286***	0.641***	0.611***	0.263***
log_PrevSpecialCorr	0.277	0.360	9.765***	0.764***	0.777***	5.724***	0.583***	0.622***	7.913***
log_PrevResTypeCorr	2.689	2.530	-0.375	-0.0353	-0.00226	3.245***	-0.0212	0.0113	2.857***
comlang_ethno		0.927***	0.927***		0.474***	0.467***		0.694***	0.685***
colony		0.274	0.328*		0.472***	0.502***		0.373***	0.415***
comcol		-0.493	-0.443		0.0258	0.117		-0.171	-0.0905
smctry		0.333	0.344		0.309	0.396		0.312	0.355*
distExp			0.0517***			0.0373***			0.0422***
distSpecial			-1.108***			-0.579**			-0.855***
distResType			0.296			-0.370***			-0.321***
contig	-0.185	-0.557**	-0.522**	-0.481***	-0.726***	-0.663***	-0.351***	-0.656***	-0.597***
SmIncomeGr	0.550***	0.626***	0.580***	0.314***	0.347***	0.319***	0.430***	0.478***	0.438***
Period Control	No	No	No	No	No	No	Yes	Yes	Yes
Therap. Area Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
_cons	0.399	0.591**	1.496***	0.0866	0.193	1.374***	0.198	0.344*	1.397***
lnalpha									
_cons	1.377***	1.337***	1.305***	1.022***	0.998***	0.977***	1.185***	1.156***	1.130***
N	23182	23182	23182	19817	19817	19817	42999	42999	42999
AIC	36406.1	36120.8	36034.2	48123.8	47960.8	47845.4	84807.4	84363.3	84156.0
BIC	36599.3	36346.2	36283.8	48313.3	48181.9	48090.1	85024.1	84614.7	84433.4
Log pseudolikelihood	-18179.061	-18032.404	-17986.103	-24037.918	-23952.419	-23891.69	-42378.681	-42152.638	-42046.005

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: Negative Binomial Regression Models With Therapeutic Area Distinction

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	98-02	98-02	98-02	04-08	04-08	04-08	Pooled	Pooled	Pooled
Dependent Variable: Number of Collaborations									
log_PubActor	0.385***	0.398***	0.409***	0.433***	0.439***	0.450***	0.418***	0.428***	0.437***
log_PubPartner	0.367***	0.379***	0.394***	0.422***	0.428***	0.438***	0.405***	0.413***	0.423***
log_distw	-0.469***	-0.511***	-0.583***	-0.491***	-0.512***	-0.550***	-0.489***	-0.520***	-0.571***
log_PrevExperience	0.476***	0.462***	0.457***	0.512***	0.498***	0.566***	0.485***	0.471***	0.504***
log_PrevSpecialCorr	-0.551**	-0.519**	13.73***	0.121	0.119	7.366***	-0.190	-0.170	9.963***
log_PrevResTypeCorr	-0.0175	-0.334	1.379	-0.223***	-0.202***	2.939***	-0.121	-0.103	2.668***
comlang_ethno		0.599***	0.585***		0.252***	0.248***		0.425***	0.411***
colony		-0.0377	0.0205		0.249**	0.302***		0.124	0.183**
comcol		-0.127	-0.0801		0.334*	0.385*		0.154	0.224
smctry		0.430*	0.500*		0.372*	0.489*		0.375**	0.459**
distExp			-0.00138			-0.0137			-0.00824
distSpecial			-1.726***			-0.856***			-1.208***
distResType			-0.324			-0.355***			-0.311***
contig	-0.412***	-0.728***	-0.618***	-0.557***	-0.777***	-0.662***	-0.483***	-0.745***	-0.635***
SmIncomeGr	0.438***	0.512***	0.508***	0.267***	0.297***	0.304***	0.360***	0.403***	0.399***
Period Control	No	No	No	No	No	No	Yes	Yes	Yes
Therap. Area Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
constant	1.233***	1.320***	1.726***	0.708***	0.763***	1.013***	0.926***	1.007***	1.340***
Zero-inflated Part (logit)									
log_distw	0.221***	0.200***	0.279***	0.239***	0.220***	0.347***	0.221***	0.202***	0.297***
log_PrevExperience	-1.895***	-1.876***	-1.883***	-1.350***	-1.344***	-1.369***	-1.636***	-1.628***	-1.637***
log_PrevSpecialCorr	-2.199***	-2.182***	-2.418***	-1.641***	-1.646***	-1.705***	-1.904***	-1.899***	-2.000***
log_PrevResTypeCorr	-5.276***	-5.560***	-5.934***	-0.594***	-0.582***	-0.497***	-0.255***	-0.252**	-0.197*
constant	-1.150***	-1.056***	-1.758***	-1.841***	-1.710***	-2.759***	-1.363***	-1.257***	-2.056***
lnalpha									
constant	0.400***	0.413***	0.381***	0.247***	0.239***	0.219***	0.344***	0.348***	0.319***
N	23182	23182	23182	19817	19817	19817	42999	42999	42999
AIC	34939.5	34816.9	34684.4	46793.7	46712.1	46556.2	82072.0	81865.6	81588.3
BIC	35172.9	35082.6	34974.2	47022.7	46972.6	46840.4	82332.1	82160.4	81909.0
Log pseudolikelihood	-17440.73	-17375.44	-17306.19	-23367.87	-23323.06	-23242.12	-41006.02	-40898.8	-40757.14

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Zero-inflated Negative Binomial Regression Models With Therapeutic Area Distinction

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. Since our measure for institutional proximity is restricted to the research environment and may face classification problems we use the "Index of Economic Freedom" as an alternative measure for the institutional settings on the macro level. In order to account for institutional proximity, we calculate the uncentred correlation of country vectors containing the average of the subindices of economic freedom for the respective years and take the log. The regression results stay qualitatively the same for this alternative proxy.

5 Conclusion

In this paper we tested empirically the determinants of cross-country collaborations in pharmaceutical research. We focused our attention on the 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. Countries close in geographical and social proximity dimensions show higher levels of international collaboration. Our empirical analysis does not indicate a robust significant association between the cognitive and institutional dimension of proximity the number of international research collaborations. Furthermore, the interaction terms of geographical distance and social, cognitive and institutional distance do not indicate a significant relation to the number of research collaborations on the country level. With respect to cultural proximity our results are inconclusive. Linguistic ties among countries are associated with a higher amount of cross-country research collaboration whereas we find no clear effect for historical and colonial linkages.

Our findings for the geographical and social dimensions of proximity are in line with previous findings in the literature (see e.g. Frenken et al., 2009). In contrast to theory, see e.g., Boschma (2005), the results of our empirical analysis do not indicate a significant association between cognitive and institutional proximity with the number of internatio-

nal research collaborations in pharmaceuticals. Furthermore, we do not find hints for the theoretically suggested complementarity between geographical and other dimensions of proximity. Similarly to Zitt et al. (2000), we find that linguistic ties are positively related to international scientific collaboration. However, we find no robust support for a positive relation between (past) colonial relations and research collaboration.

Our study should be supplemented by additional investigations addressing the determinants of cross-country research collaboration in different industries. Evidence from pharmaceutical research may not be representative since pharmaceuticals differ from other industries in many respects, e.g., their scientific foundation and their internationalization. Another limitation of this study arises from our dataset, which does not allow of taking policy programs established to stimulate international research collaboration 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 enhancement of scientific research in developing countries. Moreover, 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.

A Appendix

A.1 List of Therapeutic Areas and Description of Variables

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 3: List of Therapeutic Areas

Dependent Variable		
Collaborations		Number of Collaborations between two countries
Explanatory Variables	*Proximity*	
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 cumulated 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 analysed period
log_PrevResTypeCorr	institutional	log of country vector uncentered correlation containing the share of publications per CHI level prior to the analysed period
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
comcol	cultural	equals 1 if two countries had the same colonizer after 1945
smctry	cultural	equals 1 if two countries were part of the same country during their history
distExp		Interaction term between log_distw and log_PrevExperience
distSpecial		Interaction term between log_distw and log_PrevSpecialCorr
distResType		Interaction term between log_distw and log_ResTypeCorr
Controls		
contig		equals 1 if two countries are contiguous
SmIncomeGr		equals 1 if two collaborating countries belong the same Worldbank income group
Period Control		equals 1 if collaboration is observed in period 2
Therap. Area Controls		dummy variables for the different therapeutic areas

Table 4: Description of Variables

A.2 Descriptives and Correlations and Additional Regressions

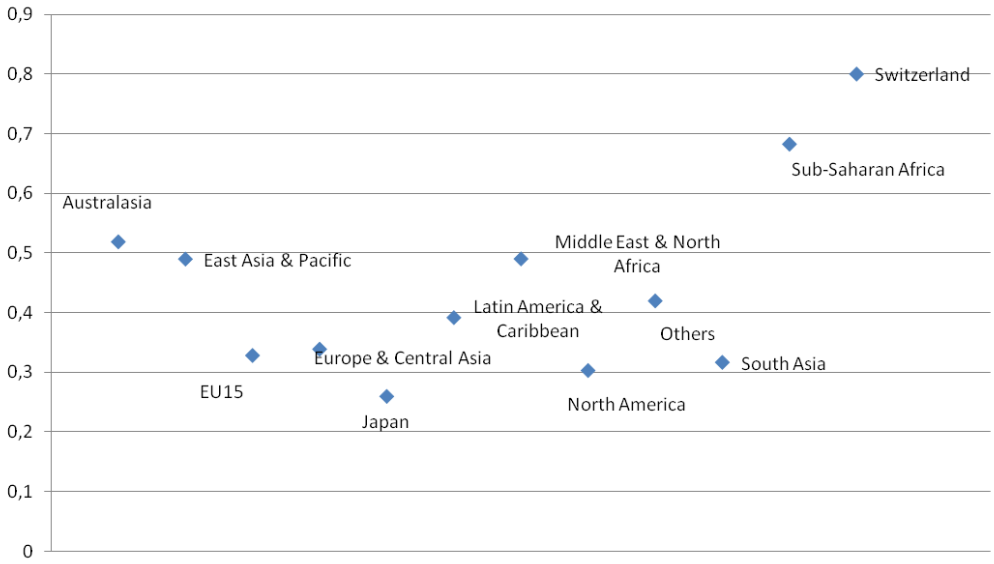

Figure 2: Share of Cross-Country Collaborations per Region

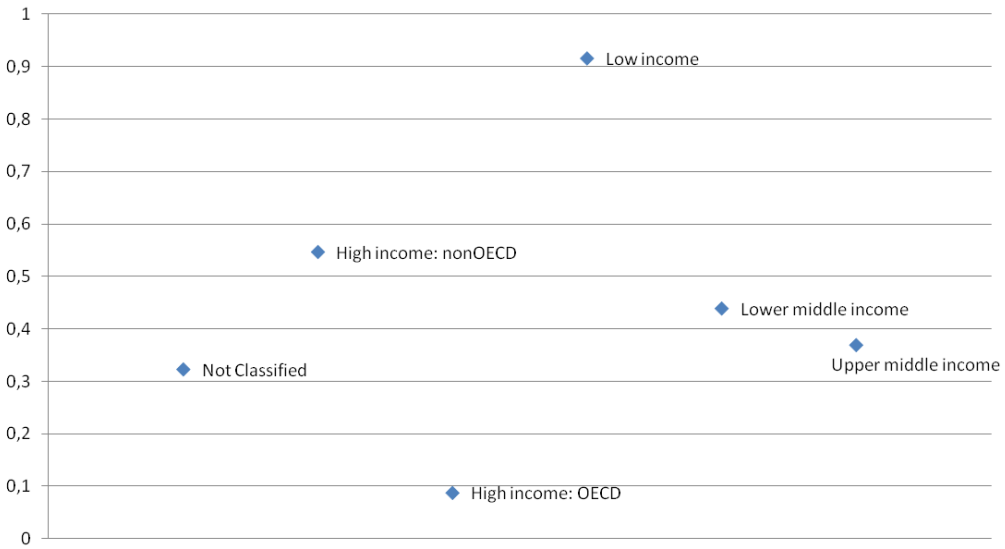

Figure 3: Share of Cross-Country Collaborations per Income Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) log_PubActor	1											
(2) log_PubPartner	0.0174*	1										
(3) log_distw	-0.0439***	-0.0421***	1									
(4) log_PrevExperience	0.379***	0.391***	-0.298***	1								
(5) log_PrevSpecialCorr	0.156***	0.146***	-0.0844***	0.214***	1							
(6) log_PrevResTypeCorr	0.193***	0.189***	-0.145***	0.179***	0.144***	1						
(7) comlang_ethno	-0.0406***	-0.0366***	-0.0488***	0.0474***	-0.0149	-0.0524***	1					
(8) colony	0.0790***	0.0985***	-0.0573***	0.159***	0.0144	0.0251**	0.163***	1				
(9) comcol	-0.137***	-0.135***	-0.0534***	-0.0815***	-0.0310***	-0.0629***	0.334***	-0.0351***	1			
(10) smctry	-0.00216	-0.0103	-0.211***	0.0290***	0.0124	0.0113	0.0977***	0.0658***	0.101***	1		
(11) contig	0.0288***	0.0279***	-0.310***	0.129***	0.0447***	0.0356***	0.130***	0.129***	0.0731***	0.369***	1	
(12) SmIncomeGr	0.0134	0.00394	-0.242***	0.252***	0.0979***	0.0533***	0.0211**	-0.00674	0.0349***	0.0885***	0.139***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Cross-correlation Table Without Therapeutic Areas Distinction

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) log_PubActor	1											
(2) log_PubPartner	0.0962***	1										
(3) log_distw	0.0000510	-0.00109	1									
(4) log_PrevExperience	0.324***	0.329***	-0.319***	1								
(5) log_PrevSpecialCorr	0.173***	0.181***	-0.163***	0.201***	1							
(6) log_PrevResTypeCorr	0.0130**	0.0174***	-0.107***	0.107***	0.110***	1						
(7) comlang_ethno	0.0159***	0.0193***	0.0135**	0.0739***	-0.00293	-0.00203	1					
(8) colony	0.0684***	0.0624***	-0.0513***	0.105***	0.0353***	0.0104*	0.233***	1				
(9) comcol	-0.0809***	-0.0787***	-0.0230***	-0.0469***	-0.0707***	0.000916	0.256***	-0.0278***	1			
(10) smctry	-0.0151***	-0.0189***	-0.209***	0.0172***	0.0243***	0.0111*	0.102***	0.102***	0.0615***	1		
(11) contig	0.0263***	0.0277***	-0.314***	0.116***	0.0689***	0.0299***	0.153***	0.176***	0.0454***	0.345***	1	
(12) SmIncomeGr	0.0861***	0.0811***	-0.306***	0.326***	0.213***	0.0841***	0.0173***	0.0108*	-0.0144**	0.0502***	0.141***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Cross-correlation Table with Distinguishing Therapeutic Areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	98-02	98-02	98-02	04-08	04-08	04-08	Pooled	Pooled	Pooled
Dependent Variable: Number of Collaborations									
log_PubActor	0.485***	0.480***	0.484***	0.462***	0.479***	0.480***	0.476***	0.484***	0.486***
log_PubPartner	0.486***	0.488***	0.493***	0.498***	0.507***	0.505***	0.494***	0.501***	0.501***
log_distw	-0.584***	-0.600***	-0.731***	-0.724***	-0.754***	-0.920***	-0.670***	-0.699***	-0.829***
log_PrevExperience	0.602***	0.559***	0.272	0.579***	0.533***	0.0108	0.586***	0.539***	0.136
log_PrevSpecialCorr	0.203	0.292*	3.051**	0.273*	0.344**	1.137	0.247*	0.333***	1.613
log_PrevResTypeCorr	1.070*	1.523***	6.842	0.164	0.237	2.933	0.317*	0.455**	2.742
comlang_ethno		0.941***	0.960***		0.721***	0.704***		0.803***	0.798***
colony		0.727***	0.751***		0.703***	0.660**		0.679***	0.659***
comcol		-0.511	-0.444		0.0766	0.0808		-0.122	-0.100
smctry		-0.180	-0.241		-0.428	-0.366		-0.332	-0.290
distExp			0.0352			0.0660***			0.0510***
distSpecial			-0.325**			-0.0958			-0.153
distResType			-0.598			-0.311			-0.263
contig	0.0426	-0.115	0.0558	-0.196	-0.381*	-0.360	-0.0847	-0.263	-0.220
SmIncomeGr	0.458***	0.525***	0.476***	0.139	0.164	0.166	0.291***	0.332***	0.316***
Period Control	No	No	No	No	No	No	Yes	Yes	Yes
_cons	-0.0674	0.00484	1.071	1.250***	1.319***	2.704***	0.660*	0.741**	1.821***
lnalpha									
_cons	1.179***	1.103***	1.087***	0.835***	0.783***	0.775***	0.983***	0.925***	0.918***
N	7319	7319	7319	6697	6697	6697	14016	14016	14016
AIC	12251.7	12112.7	12095.2	17017.5	16883.4	16858.4	29314.6	29049.0	29019.0
BIC	12320.7	12209.3	12212.5	17085.6	16978.7	16974.1	29397.6	29162.2	29154.8
Log pseudolikelihood	-6115.851	-6042.343	-6030.601	-8498.75	-8427.694	-8412.192	-14646.294	-14509.484	-14491.487

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Negative Binomial Regression Models Without Therapeutic Area Distinction

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	98-02	98-02	98-02	04-08	04-08	04-08	Pooled	Pooled	Pooled
Dependent Variable: Number of Collaborations									
log_PubActor	0.342***	0.357***	0.364***	0.360***	0.378***	0.385***	0.358***	0.374***	0.381***
log_PubPartner	0.340***	0.357***	0.367***	0.374***	0.391***	0.395***	0.366***	0.384***	0.389***
log_distw	-0.417***	-0.460***	-0.479***	-0.495***	-0.535***	-0.588***	-0.464***	-0.507***	-0.538***
log_PrevExperience	0.520***	0.489***	0.587***	0.536***	0.512***	0.461***	0.525***	0.497***	0.518***
log_PrevSpecialCorr	-0.140	-0.0536	3.171**	-0.362**	-0.311*	0.701	-0.304***	-0.228*	1.520
log_PrevResTypeCorr	-0.0861	0.342	4.189	-0.238	-0.178	2.802	-0.114	0.0152	2.242
comlang_ethno		0.724***	0.761***		0.394***	0.422***		0.535***	0.561***
colony		0.114	0.137		0.257*	0.241*		0.201*	0.192*
comcol		-0.494	-0.423		0.150	0.195		-0.0756	-0.0244
smctry		-0.119	-0.128		-0.280	-0.218		-0.215	-0.156
distExp			-0.0151			0.00520			-0.00429
distSpecial			-0.387**			-0.125			-0.212
distResType			-0.420			-0.339			-0.252
contig	-0.240	-0.357	-0.204	-0.431***	-0.517***	-0.450**	-0.340**	-0.442**	-0.360*
SmIncomeGr	0.461***	0.509***	0.503***	0.246**	0.268***	0.264***	0.359***	0.389***	0.383***
Period Control	No	No	No	No	No	No	Yes	Yes	Yes
constant	0.570*	0.695*	0.769	0.876***	0.969***	1.342**	0.651***	0.777***	0.957**
Zero-inflated Part (logit)									
log_distw	0.156*	0.129*	0.215**	0.659***	0.625***	0.673***	0.409***	0.376***	0.434***
log_PrevExperience	-1.809***	-1.789***	-1.802***	-1.518***	-1.518***	-1.535***	-1.617***	-1.608***	-1.623***
log_PrevSpecialCorr	-0.580***	-0.528**	-0.577**	-1.622***	-1.597***	-1.624***	-1.161***	-1.126***	-1.151***
log_PrevResTypeCorr	-3.031***	-2.770***	-2.623***	-1.010***	-0.985***	-0.944***	-1.167***	-1.097***	-1.071***
constant	-0.877	-0.713	-1.444*	-5.925***	-5.682***	-6.111***	-3.369***	-3.146***	-3.652***
lnalpha									
constant	0.239*	0.240*	0.223*	0.0530	0.0554	0.0609	0.148**	0.153**	0.151**
N	7319	7319	7319	6697	6697	6697	14016	14016	14016
AIC	11802.2	11737.7	11719.1	16299.1	16258.7	16252.3	28224.0	28119.0	28096.7
BIC	11905.6	11868.7	11870.8	16401.3	16388.0	16402.1	28344.8	28269.9	28270.3
Log pseudolikelihood	-5886.08	-5849.838	-5837.532	-8134.573	-8110.328	-8104.163	-14096.02	-14039.48	-14025.34

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Zero-inflated Negative Binomial Regression Models Without Therapeutic Area Distinction

References

Adams, J. D., Black, G. C., Clemmons, J. R., and 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., and McHale, J. (2006). Gone but not Forgotten: Knowledge Flows, Labor Mobility, and Enduring Social Relationships. *Journal of Economic Geography*, 6(5):571–591.

Audretsch, D. B. and Feldman, M. P. (1996). R&D Spillovers and the Geography of Innovation and Production. *The American Economic Review*, 86(3):630–640.

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, 39(1):61–74.

Boschma, R. and Lambooy, J. G. (1999). Evolutionary Economics and Economic Geography. *Journal of Evolutionary Economics*, 9(4):411–429.

Burger, M., van Oort, F., and 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.

Calero, C., van Leeuwen, T. N., and Tijssen, R. J. W. (2007). Research Cooperation Within the Bio-pharmaceutical Industry: Network Analyses of Co-publications Within and Between Firms. *Scientometrics*, 71(1):87–99.

Cameron, A. C. and Trivedi, P. K. (1990). Regression-based Tests for Overdispersion in the Poisson Model. *Journal of Econometrics*, 46(3):347–364.

Cantner, U. and Meder, A. (2007). Technological Proximity and the Choice of Cooperation Partner. *Journal of Economic Interaction and Coordination*, 2:45–65.

Cantwell, J. and Janne, O. (2000). The Role of Multinational Corporations and National States in the Globalization of Innovatory Capacity: The European Perspective. *Technology Analysis & Strategic Management*, 12(2):243–262.

Cockburn, I. and Henderson, R. (1996). PublicPrivate Interaction in Pharmaceutical Research. *Proceedings of the National Academy of Sciences*, 93(23):12725–12730.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1):128–152.

de Solla Price, D. (1963). *Little Science, Big Science*. Columbia University Press, New York et al.

Edquist, C. and Johnson, B. (1997). Institutions and Organizations in Systems of Innovation. In *Systems of Innovation: Technologies, Institutions, and Organizations*, pages 41–63. Pinter, London et al.

Flowerdew, R. and 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., and van Vliet, J. (2009). Death of Distance in Science? A Gravity Approach to Research Collaboration. In Pyka, A. and Scharnhorst, A., editors, *Innovation Networks*, pages 43–57. Springer, Berlin and Heidelberg.

Gambardella, A. (1992). Competitive Advantages from In-House Scientific Research: The U.S. Pharmaceutical Industry in the 1980s. *Research Policy*, 21(5):391–407.

Gambardella, A. (1995). *Science and Innovation: The U.S. Pharmaceutical Industry During the 1980s*. Cambridge University Press, Cambridge et al.

Gassmann, O. and von Zedtwitz, M. (1999). New Concepts and Trends in International R&D Organization. *Research Policy*, 28(2-3):231–250.

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.

Hamilton, K. S. (2003). Subfield and Level Classification of Journals. *CHI Research Inc.*, CHI No. 2012-R.

Hoekman, J., Frenken, K., and 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., and van Oort, F. (2009). The Geography of Collaborative Knowledge Production in Europe. *The Annals of Regional Science*, 43(3):721–738.

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., and 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. and Martin, B. R. (1997). What is Research Collaboration? *Research Policy*, 26(1):1–18.

Laudel, G. (2002). What do we Measure by Co-authorships? *Research Evaluation*, 11(1):3–15.

Lundberg, J., Tomson, G., Lundkvist, I., Skar, J., and Brommels, M. (2006). Collaboration Uncovered: Exploring the Adequacy of Measuring University-Industry Collaboration Through Co-Authorship and Funding. *Scientometrics*, 69(3):575–589.

Lundvall, B.-A. (1992). *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. Pinter, London.

Luukkonen, T., Persson, O., and Sivertsen, G. (1992). Understanding Patterns of International Scientific Collaboration. *Science, Technology & Human Values*, 17(1):101–126.

Mayer, T. and Zignago, S. (2006). Notes on CEPIIs Distances Measures. *CEPII (Centre d'Études Prospectives et d'Information Internationales), Paris*.

McKelvey, M., Orsenigo, L., and Pammolli, F. (2004). Pharmaceuticals Analyzed Through the Lens of a Sectoral Innovation System. In Malerba, F., editor, *Sectoral Systems of Innovation: Concepts, Issues and Analyses of Six Major Sectors in Europe*, pages 73–120. Cambridge University Press, Cambridge et al.

Narin, F., Stevens, K., and Whitlow, E. (1991). Scientific Co-Operation in Europe and the Citation of Multinationally Authored Papers. *Scientometrics*, 21:313–323.

Nooteboom, B. (2000). *Learning and Innovation in Organizations and Economies*. Oxford University Press, Oxford et al.

Orsenigo, L., Pammolli, F., and Riccaboni, M. (2001). Technological Change and Network Dynamics: Lessons from the Pharmaceutical Industry. *Research Policy*, 30(3):485–508.

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., and Frenken, K. (2007). The Geographical and Institutional Proximity of Research Collaboration. *Papers in Regional Science*, 86(3):423–443.

Tijssen, R. J. W. (2004). Is the Commercialisation of Scientific Research Affecting the Production of Public Knowledge?: Global Trends in the Output of Corporate Research Articles. *Research Policy*, 33(5):709–733.

Tijssen, R. J. W. (2009). Measuring the Corporate Web of Science: Research and Partnership Networks Within the European Pharmaceutical Industry. In Malerba, F. and Vonortas, N. S., editors, *Innovation Networks in Industries*, pages 81–104. Edward Elgar Publishing, Cheltenham et al.

Tinbergen, J. (1962). *Shaping the World Economy : Suggestions for an International Economic Policy*. The Twentieth Century Fund, New York.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57(2):307–333.

Wuchty, S., Jones, B. F., and 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., and Okubo, Y. (2000). Shadows of the Past in International Cooperation: Collaboration Profiles of the Top Five Producers of Science. *Scientometrics*, 47:627–657.