From the Unit of Bioentrepreneurship, Department of Learning, Informatics, Management and Ethics
Karolinska Institutet, Stockholm, Sweden

European knowledge transfer reflected by research collaboration and patent citations indicators

Pauline Mattsson

Stockholm 2011

Published by Karolinska Institutet.

ISBN 978-91-7457-192-9

Printed by
REPROPRINT AB
Stockholm 2010
www.reproprint.se
Gårdsvägen 4, 169 70 Solna

To my growing family

ABSTRACT

Knowledge transfer consists of activities that aim to capture and transmit knowledge, skills and competence from those who generate them to those who will transform them into socio-economic outcomes. In the context of the March 2000 Lisbon strategy and its aim to make the European Union the "world's most dynamic and competitive knowledge economy", knowledge transfer is considered to play an important role in helping to overcome obstacles such as a weak environment to stimulate high quality research and exploit research results. The introduction of new funding schemes and policies aimed at increasing knowledge flow between countries and sectors in Europe has increased the demand for studies of the impact of such policies and funding mechanisms and the development of relevant and accurate indicators related to them. The aim of this thesis was to study the dynamics of knowledge transfer in Europe and to examine how knowledge transfer can be measured and analysed through different indicators. This was done by studying co-authorships and collaborations within Europe as indicators of geographical knowledge transfer and patent citations as an indicator of sectoral knowledge flow.

The results showed that researchers from smaller countries co-authored more with other EU countries than those from bigger countries, while the co-authorship rate with extra-EU partners was not dependent on a country's size. Co-authorship patterns were also found to depend on the scientific field. The analysis also indicated that multilateral collaborations funded through the EU Framework Programmes are more exclusively European in nature. In contrast, co-publication patterns in multilateral collaborations suggested that European researchers tend to co-author more with global, rather than exclusively European partners and that this global multilateral orientation in co-publications continues to rise.

When using co-publications as an indicator for geographical knowledge flow, the results demonstrated that European research policy most likely has had an impact on research collaboration patterns. However, the results also strongly suggested that any direct impact was limited and did not over-ride self-selected collaboration patterns that continue to drive a more global, rather than exclusively European, research collaboration orientation. A more disaggregated scrutiny of publication patterns also underscored very clearly that collaboration strategies show considerable diversity across scientific fields, as well as countries. Further, the results suggest that some policies, to support innovation in regions with a low absorptive capacity (weak innovation activities and a low tech profile) e.g. supporting regional R&D through subsidies, may be less successful than the incorporation of qualified personnel at firms or the increase of local university-industry links.

The thesis also made several contributions to the discussion of research methods in this field by investigating the utility of some central indicators and approaches. The results showed that the corresponding author is most likely to appear first and thereafter most-likely to appear last in the byline. However, the analysis also indicated that these results are dependent on the number of authors in a paper and that national differences also exist, thus arguing for a fine-tuning of bibliometric tools, in order to more effectively capture the relative importance of author contributions. Similarly, the analysis examined the use of patent citations as an indicator for science-industry links and geographical localization at a regional level. It found that there are reasons to question the use of this indicator, specifically in a region with low absorptive capacity. The related results also highlighted that there is a need to differentiate between applicant and examiner citations when examining the knowledge base since examiners and applicants add different types of knowledge.

Keywords: Knowledge transfer; indicators; author position; bibliometrics; corresponding author; co-authorship; EU; Europe; funding; international collaborations; networks; publications; research policy; research assessment;

LIST OF PUBLICATIONS

I. Mattsson, P., Laget P., Nilsson, A., Sundberg, C.J., (2008). Intra-European vs. extra-European scientific co-publication patterns in Europe. *Scientometrics* 75(3), 555-574.

II. Mattsson, P., Laget P., Nilsson, A., Sundberg, C.J., (2010). What do European research collaboration networks in Life Sciences look like? *Research Evaluation* 19(5), 373-384.

III. Mattsson, P., Laget P., Sundberg, C.J., (2011). Is Correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. *Scientometrics* 87, 99-105.

IV. Azagra Caro, J. M., Mattsson, P., Perruchas, F., (2010). Smoothing the lies: Do patent examiners take notice if applicants include citations? Submitted to: *Journal of the American Society for Information science and Technology* (under review).

V. Azagra Caro, J. M., Fernandez de Lucio, I., Perruchas, F., Mattsson, P., (2009). What do patent examiner inserted citations indicate for a region with low absorptive capacity? *Scientometrics* 80(2), 441-455.

Publications outside thesis

- Mattsson, P., Molas-Gallart, J., Dagher, G., Meijer, I., (2010). Biobanking in transition: towards professional research infrastructures. *Science*. Invited.
- Compano, R., Bock, A.-K., Burgelman, J.C., Cabrera, M., Da Costa, O., Mattsson, P., Malanowski, N., (2006). Converging applications for active ageing policy. *Foresight* 8(2), 30-42.

CONTENTS

1 Introduction 1
1.1 The role of science 1
1.2 Knowledge transfer 2
1.3 What is research quality? 3
1.4 Science and technology measurement indicators 4
1.4.1 Bibliometric indicators 5
1.4.2 Patentometric indicators 6
1.5 Research and innovation policy 7
1.6 Geographic knowledge transfer – international collaborations 8
1.6.1 Benefits of and motivations for collaborating 8
1.6.2 Increase in international collaborations 10
1.6.3 Factors related to the patterns of research collaboration 11
1.7 Sectoral knowledge transfer – patent citations 11
1.7.1 Different patent systems 11
1.7.2 Who includes citations 12
1.7.3 Patent characteristics 13
2 Study rationale and aims 15
2.1 Study rationale 15
2.2 The main aim 17
2.3 Specific aims 17
3 Material and Methods 18
3.1 Bibliometrics as a method 18
3.2 Patentometrics as a method 18
3.3 Definitions 19
3.4 Study delimitations 19
3.5 Data collection and processing 20
3.6 Statistics 22
3.7 Methodological considerations 23
4 Results and comments 25
4.1 The increase of international collaborations 25
4.2 Scientific field and country differences 26
4.3 Countries have different roles 29
4.4 Researchers tend to collaborate in global multilateral settings 30
4.5 Author position related to contribution in co-authored papers 31
4.6 Indication of inventor's knowledge base 32
4.7 Factors influencing citation patterns 34
5 Discussion 37
5.1 Research implications 37
5.1.1 Dynamics of research collaborations 37
5.1.2 Country differences in co-publication 37
5.1.3 Field differences in co-publication 38
5.1.4 Byline positions in internationally co-authored papers 39
5.1.5 Differences between examiner- and applicant-added citations 40
5.1.6 Citations as an indication of science-technology links 40
5.2 Policy implications 41

5.3 Further work 43
6 Summary and conclusions..........45
7 Populärvetenskaplig sammanfattning..........48
8 Acknowledgements..........49
References..........50

LIST OF ABBREVIATIONS

BRIC	The countries of Brazil, Russia, India, and China
FP	Framework Programme
R&D	Research and Development
ERA	European Research Area
EC	European Commission
EU	European Union
IF	Impact Factor
NGO	Non-governmental organisation
S&T	Science and Technology
S&E	Science and Engineering
RTD	Research and Technology Development

1 INTRODUCTION

The introduction of new funding schemes and policies aimed at increasing the knowledge flow between countries and sectors in Europe has increased the demand for studies of the impact of policies and the development of relevant and accurate indicators. I believe that to justify research funding and to develop efficient policies there is a need to better understand the sources of and mechanisms by which new discoveries and innovations are developed and how this can be measured.

1.1 THE ROLE OF SCIENCE

Science derives from the Latin word scientia meaning knowledge. Even though many of the results from the work done by the Greek philosopher Aristotle have later been shown to be wrong he established the basis of a system explaining the world based on deductive reasoning, which can be further communicated and taught (Serres, 1995). One central difference between Aristotle's and today's use of the notion of science is that he did not treat the gathering of experience and raw data as part of science itself. Modern science was developed during the seventeenth century and introduced a new understanding of the natural world, what is known as the Scientific Revolution (Barrett, 2004). During this time new ways of studying the natural world, using methodological experimentation were developed (Serres, 1995).

In the eighteenth and nineteenth centuries, along with the industrial revolution, the majority of research was carried out by individual researchers at universities financing the research with their own funds or through private sponsorship. It was not until the twentieth century that scientific research became systematised, as corporations developed and as they discovered that continuous investment in research and development could be a key element of success in a competitive strategy (Serres, 1995). Since then the benefits of science have been recognised by the wider community as a means to address socio-economic challenges and strengthen competitiveness. Today, scientific research is funded by public authorities, by charitable organizations and by private groups, including many companies.

In an attempt to distinguish between science and research I would argue that science is related to the systematic production and validation of knowledge under adherence to strict rules of scientific integrity and epistemological principles. Research is a broader concept of inquiry, knowledge synthesis, acquisition that many stakeholders, not necessarily only scientists as such, perform to satisfy their information and knowledge needs, e.g. journalists, policy-makers, marketing and sales managers etc.

Two major types of research may be distinguished. Basic research aims at enhancing the understanding of fundamental principles driven by curiosity and without direct or immediate practical and useful implication (Calvert and Martin, 2001). Applied research, in contrast, either uses the findings of basic research or determines new methods or ways of achieving specific and predetermined objectives or practical aims.

The results of applied research are intended for a single or limited number of products, operations, methods or systems (OECD, 2002).

1.2 KNOWLEDGE TRANSFER

Knowledge can be divided into two different types: explicit and tacit knowledge. Explicit knowledge is codified knowledge and can be transmitted with formal language. Tacit knowledge can be defined as non-codified knowledge and is therefore more difficult to communicate. Tacit knowledge has in many cases been identified as the most valuable knowledge in science (Polanyi, 1999). It is common to distinguish between two forms of codified knowledge, *scientific publications* that are reflecting scientific ideas and *patents*, reflecting a technological invention.

Knowledge transfer is the process of transferring knowledge from one actor to another. This can take place at many different levels and between any actors that carry the knowledge and are interested in benefitting from the knowledge. In the literature knowledge transfer has specifically been researched in the context of firms. According to Argot and Ingram (2000) the creation and transfer of knowledge is a basis for competitive advantage in firms. Knowledge transfer can also take place between different sectors (public-private) such as when universities collaborate with industry (Cohen, Nelson and Walsh, 2002; Narin, Hamilton and Olivastro, 1997). In this case public research organizations have mainly been the knowledge producers while the private sector has been the user. Knowledge transfer can take place at different geographical levels such as between organizations within a region, country or internationally (Jaffe, Trajtenberg and Henderson, 1993). One common factor for all actors involved in knowledge transfer is that it seeks to organize, create, capture or distribute knowledge and ensure its availability for future users.

The benefits of knowledge transfer can either be observed immediately or it can take several years, which is the case for many socio-economic benefits. The processes are often very complex and the role of the involved actors and the relevance of the original piece of knowledge can vary immensely. It can therefore be difficult to calculate the socio-economic impact of a particular piece of knowledge, and its degree of success may be very difficult to determine (European Commission, 2009). In Figure 1 below different actors and channels for knowledge transfer resulting in socio-economic benefits are illustrated. The figure has its limitations since knowledge transfer does not only take place in a linear manner, but can also occur between all different actors in the figure (Cullen, 2008).

In this thesis I have studied geographic knowledge transfer taking place between different countries via research collaborations and knowledge transfer taking place between different sectors (such as between public and private sectors). It should be highlighted that these are not separable and can take place at the same time.

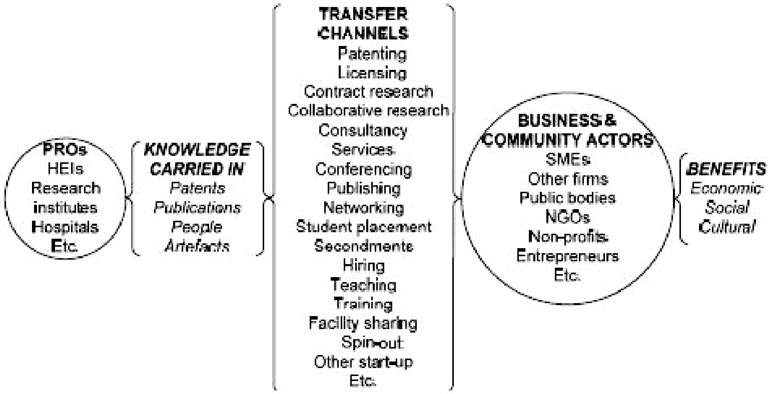

Figure 1: Knowledge transfer from public research organisation. Source: Cullen, 2008.

1.3 WHAT IS RESEARCH QUALITY?

Research quality assessment refers to a scientific process that takes into account all aspects of study design, in particular the judgment of whether the method used is appropriate to the research questions asked, selection of population and study object, measurement of outcomes, and robustness against systematic bias, non-systematic bias, and related errors (NCDDR, 2005).

Ever since the benefits of science were recognised by the wider public research quality has received increasing attention. For funders it is important to justify public spending by ensuring that the research funded is of the highest quality and benefit to society. This is typically done by establishing assessment and evaluation exercises using different measuring and validating methods. Research quality is also important for the scientific community, both as a means to improve research performance but also as a way of ensuring self-regulation by the academic community. This intra-professional process of evaluating research goes under the term "peer review" (Wikipedia, 2010).

Research projects are often funded through a competitive process, in which submitted research proposals undergo an evaluation procedure where only the ones considered to have the highest potential receive funding. In these processes differences between good research and bad research are often made. According to Litman (2010):

Good research...

> *.....reflects a sincere desire to determine what is overall true, based on available information*
> *....requires judgment and honesty*
> *.... is cautious about drawing conclusions, careful to identify uncertainties and avoids exaggerated claims*

....demands multiple types of evidence to reach a conclusion

in contrast, bad research...
....often uses accurate data, but manipulates and misrepresents the information

Research quality is not only influenced by the way researchers are carrying out their research, but also by a number of other factors, such as the environment were the research is carried out, culture, history, economic situation, and level of collaborations. In addition, the usefulness of research has come to play an increasing important role. In this context knowledge transfer processes from research institutions to other users is a mean to achieve technological development which might lead to social, cultural and personal benefits.

All these factors need to be taken into account when evaluating and assessing research quality. As was mentioned above a number of different indicators can be used to measure the research quality of an individual researcher or of other actors involved in the knowledge creation and diffusion of research. These include, among many others, different bibliometric and patentometric indicators, amount of funding, investment in R&D activities, number of PhD students and other research related staff, and number of research collaborations.

1.4 SCIENCE AND TECHNOLOGY MEASUREMENT INDICATORS

Actors investing in science and technology want to understand the sources and mechanisms by which new discoveries and innovations are developed and what the socio-economic consequences are. Different types of indicators can be used for this task. Indicators in this context can be defined as measurements that enable decision-makers, but also other actors involved in science and technology to assess progress towards the realization of intended outputs, outcomes, goals and objectives (UNESCO, 2010). More specifically, indicators can provide evidence whether a specific result has or has not been achieved. Indicators can exist on a number of different levels from inputs, outputs, to results (outcomes) and impacts. They can also provide information about the context or the assumptions under which these interventions operate.

Once the different objectives of, for example, a research project, research programme or general research policies have been identified, implementation can be viewed as a process whereby "inputs" (human, financial and organisational resources) are transformed into "outputs" (projects, bursaries etc.) through a set of activities. Through processes of interaction with potential direct users and beneficiaries the outputs are envisaged to be translated into beneficial results ("outcomes") and, eventually, these outcomes are meant to translate into beneficial "impacts" on society at large (UNESCO, 2010). The transformation of inputs into outputs can be referred to as an implementation process; results/outcomes (publications and patents, spin-off companies and products) and possible socio-economic impacts (improved health, reduced poverty, increased employments/tax revenues) are called effects, which are often long-term. The relation between policy objectives, inputs, outputs, results and impacts are schematically depicted in the Figure 2.

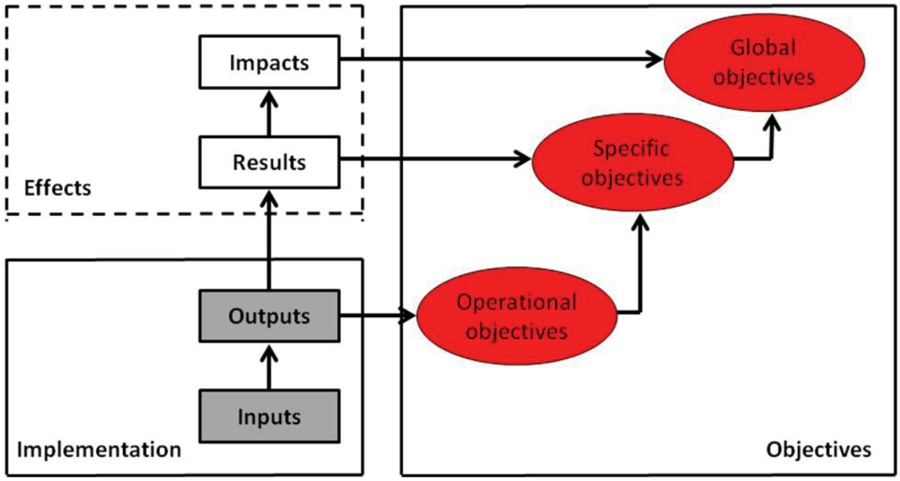

Figure 2: Indicator development at different levels. Source: Technopolis, 2010 based on EC evaluation guidelines.

The objectives illustrated in Figure 2 can be found at three different levels. Operational objectives are the immediate objectives of the implementation of the intervention and should be directly linked to the inputs e.g. the number of PhDs that should be funded by a programme. Specific objectives may include immediate advantages of the intervention e.g. that a PhD graduate has found a job with one of the companies that was involved in funding the programme. Global objectives are farther reaching objectives to which the intervention is contributing, sometimes further in the future, but which it cannot fulfil on its own. These objectives can be found at regional, national and international levels.

For this thesis a special emphasis is put on input in the form of funding and outcomes, including publications, collaborations and patents. More specifically, in research and innovation policy and science management issues related to research evaluation and assessment of input and outcomes have gained increasing importance and use (Moed, 2009).

1.4.1 Bibliometric indicators

Bibliometric indicators, defined as *"the application of mathematical and statistical methods to books and other media of communication"* play an important role in this context (Pritchard, 1969). They offer one way to differentiate between what is often labelled as high quality or excellent research and research that is of lesser quality (Evidence Ltd, 2007). Different bibliometric indicators have first and foremost been used as a response to research policy questions and as a tool for distribution of resources (Adam, 2002).

There are three types of bibliometric indicators: quantity indicators measuring the productivity of a particular research actor; quality indicators measuring the quality (also called performance) of an actor's output; and structural indicators measuring connections between publications, authors and areas of research (Rehn and Kronman, 2008). Quantity indicators generally involve counting and can include the number of papers and number of publications in top-ranked journals. There exist many different quality indicators both on journal and individual researcher level. The best known journal quality indicators include the Impact Factor (IF), measuring the citations to a journal over a certain time period with the intention to measure the importance of this journal in its field (Garfield, 1972). The journal-to-field impact score is an alternative to the IF and takes into account variations between different research fields (van Leeuwen and Moed, 2002). A common research quality indicator includes the Crown Indicator, which is calculated by dividing the average number of received citations by the average number that could be expected for publications of the same type and year and published in journals within the same field (Lundberg, 2007). The H-Index has also been widely used. An H-Index of a researcher is h if h among his N articles have at least h citations each and the other N-h articles have fewer than h citations each (Hirsch, 2005). All of these indicators do entail some disadvantages and it is difficult to find one indicator taking all important factors, e.g. field differences, time, or size, into account.

Developing co-publication indicators comes with its own challenges. Earlier works suggest that collaborative publications are more cited than articles with only one author (Lewison and Cunningham, 1991; Rigby and Edler, 2005). Glänzel and Schubert (2001) and Persson et al. (2004) studied the subject in more detail and found that internationally co-authored publications have a higher citation rate than domestic collaborations. Others have looked at the "core" journals (most cited) of a specific field and found that the number of co-authored papers is higher than in other journals (Beaver and Rosen, 1979; Gordon, 1980). Research activity, as expressed by the number of publications of the individual author, has also been shown to increase, when collaborations are carried out (Persson et al., 2004). In addition, one also has to bear in mind that these measurements vary by the specific scientific fields studied.

1.4.2 Patentometric indicators

According to the European Patent Office (EPO) a patent is a legal title granting its holder the right to prevent third parties from commercially exploiting an invention without authorization during a limited time period.

When a researcher or a company applies for a patent, it has to provide a description of the invention, proving its novelty, non-obviousness and utility. Many applications include 'prior art' in the form of previous inventions or other relevant scientific information describing the differences from existing patents. A patent that includes reference to prior art, referred to as a citation, within a specific document can be said to build on the knowledge in the document(s) it cites (Michel and Bettels, 2001). Citations in patents have been used as knowledge indicators and can be used to trace the information sources on which the invention is built. Further, they can illustrate the relations with other inventions e.g. geographic, sectoral and technological linkages. Many existing citation studies use patent citations to analyse knowledge flows from company to company, or from other sectors, e.g. research institutes and academia to companies (Meyer, 2002; Leydesdorff and Meyer, 2003).

Citations can be classified into “backward citations” or “forward citations”. A backward citation is the term used for a traditional citation and is the document that was published earlier than the document citing it. In turn, the newer document is called forward citation or citing document. The use of forward citations is customary to express the technological impact of the patented invention (Noma & Olivastro, 1984), often as a function of the characteristics of the patent (Allison & Sager, 2007). The use of backward citations is customary to express the knowledge base of the patented invention.

In addition one has to differentiate between two different types of references, or citations: patent references (PRs) and non-patent references (NPRs). The latter concept was introduced by Narin and Noma (1985) in an investigation of the science-technology links and it has been argued by the same authors to reveal a direct influence of science on technology (Narin et al., 1997). In contrast, Meyer (2000) does not claim such a strong link. NPRs include, but are not confined to books, articles in scientific journals or newspaper articles. Callaert et al. (2006) and Harhoff et al. (1999) found that in most cases NPRs are references to scientific journal articles. In general, PRs are considered to be indicators of the technological knowledge embodied in patents while NPRs are mainly seen as indicating scientific proximity and a measure of the knowledge flow between science and technology.

1.5 RESEARCH AND INNOVATION POLICY

Research and innovation policy is the area of public policy that pertains to policies affecting research and development, including related funding policies. Researchers have been increasingly called on by governments, business and the broader public to assist with solving many of the complex problems that societies face (Bammer, 2008). As a result, research policy can also be related to issues of applying scientific knowledge to the development of policies in other fields and to the broader question of how science and technology can best serve society. This becomes evident when studying the way research policy has been used to address other policy goals, for example in health care, foreign policy (Wagner, 2002), defence (Molas-Gallart, 2001), environment (Wagner, 2006; Engel and Ruschenburg, 2008) or with regard to technology innovations (Katz and Martin, 1997; Metcalf and Georghiou, 1998).

A great variety of different actors are involved in the research and innovation policy process, including both public and private actors. On the recipient or target side policies can be designed for and involve actors at different levels from the micro-level, including individual researchers and research groups, to the meso-level where research organisations, industry, NGOs, hospitals and other organisation can be found and to the macro-level that involves governments and agencies at national or international level. Similarly, research and innovation policies are being decided at different levels depending on the governance structure of the country or group of countries. In general, these levels include international (both EU and global level), national and regional policy-making. Still, research and innovation policy can also be made at a meso-level and then include different research organisations and companies. Policy-making at this level is often less formal and more open to changes.

Research funding is very much interlinked with and an integral part of research and innovation policy and again involves a diversity of actors. The major funding sources are governments (primarily carried out through universities and specialised government agencies) and industry. In the life sciences, scientific research is also carried out and funded by charitable foundations, especially in relation to developing cures for diseases and medical conditions. According to the OECD, around 2/3 of all R&D is carried out by industry, 10% by universities and 20 % by governmental research organisations (OECD, 2002). In developing countries the industry share is generally smaller.

1.6 GEOGRAPHIC KNOWLEDGE TRANSFER – INTERNATIONAL COLLABORATIONS

1.6.1 Benefits of and motivations for collaborating

The need for scientific teamwork has been recognised ever since the professionalization of science took place in France during Napoleon's time and later on in England and Germany (Beaver and Rosen, 1978).

The benefits of research collaborations are many and range from scientific to economic and political factors. Georghiou (1998), for example, distinguishes between direct and indirect benefits. Direct benefits accrue when accesses to complementary expertise, knowledge or skills help enhance scientific or technological excellence. In comparison, indirect benefits are targeted, when collaborations are driven by external goals of prestige, economic, political or cultural nature.

Beaver (2001) took a motivation-centred approach to explore the benefits of collaboration and listed eighteen purposes for which collaborations are taking place.

Table 1. Purposes for which researchers collaborate. Source: Beaver (2001).

1	Access to expertise.
2	Access to equipment, resources, or "stuff" one doesn't have.
3	Improve access to funds.
4	To obtain prestige or visibility; for professional advancement.
5	Efficiency: multiplies hands and minds; easier to learn the tacit knowledge that goes with a technique.
6	To make progress more rapidly.
7	To tackle "bigger" problems (more important, more comprehensive, more difficult, global).
8	To enhance productivity.
9	To get to know people, to create a network, like an "invisible college".
10	To retool, learn new skills or techniques, usually to break into a new field, subfield, or problem.
11	To satisfy curiosity, intellectual interest.

12	To share the excitement of an area with other people.
13	To find flaws more efficiently, reduce errors and mistakes.
14	To keep one more focussed on research, because others are counting on one to do so.
15	To reduce isolation, and to recharge one's energy and excitement.
16	To educate (a student, graduate student, or, oneself).
17	To advance knowledge and learning.
18	For fun, amusement, and pleasure.

Katz and Martin (1997) offer a different approach for clustering factors that drive collaborations. These factors are divided into four categories that include 1) financial reasons such as better access to funding or sharing of core-facilities that cannot be purchased by the individual laboratory; 2) social factors such as acknowledgement from the scientific community, increased networking effect (getting to know more people in the scientific community), or human social factors (we prefer to work in a group rather than individually); 3) knowledge collaborations, including education of students, improved technical, analytical, and theoretical knowledge; and, 4) political factors, such as Framework Programmes and other policy-based initiatives.

The tendency to collaborate outside national borders has been present since the early days of modern science (Sörlin, 2004). Motivations to collaborate internationally include all the above mentioned factors and as Georghiou (1998) puts it:

"*…the wider geographical coverage of a program, the greater the chance becomes of finding exactly the right partner*".

In a survey of senior science and technology officials in EU member states (carried out on behalf of the CREST working group on internationalisation of research and development, 2007), improved competitiveness and market access were the primary motives stated for international cooperation. Increasing the quality of research was reported to be a primary focus for just under half of all respondents, while addressing global issues and international development was the third major objective. These results underline the more general importance of research to wider economic ambitions and to international relations.

Even though international collaborations mainly bring benefits there are also some drawbacks. Katz and Martin (1997) discuss the increased financial and administrative costs, as well as the increased travel time that are associated with collaborations in general, all likely to be higher in international settings. Georghiou (1998) also mentions the differences between research priorities and policy support as a possible geographical impediment. Even when research is based on comparable fundamental research questions, the outcomes can be difficult to compare due to differences in policy support and priorities in different countries. Similarly, differences can also result in uncertainty of institutional responsibility within a joint project involving many different researchers from different countries. In the end a research team has to weigh the potential benefits against the possible costs of international collaborations.

1.6.2 Increase in international collaborations

In the last decade, a number of studies have indicated that international research collaborations have increased markedly as measured by the number and share of co-publications (Glänzel, 2001; Adams, et al., 2005; National Science Foundation, 2007). One of the first studies confirming this trend was conducted by Smith (1958) who observed an increased number of co-authored publications and started to use multiple-author papers as an index for collaborations. According to Adams et al. (2005) national collaborations between universities or universities and firms in the USA have doubled between 1981 and 1999, while international collaborations over the same period have increased five-fold. In a period of 20 years (1988-2008), international co-authored papers grew from 8% of the world's total S&E articles to 22 % (National Science Foundation, 2008). Another indicator of increased collaboration is that the number of authors per paper has increased and the share of publications with multiple institutions grew from 40 % to 61 % between 1988 and 2005 (National Science Foundation, 2008), Figure 3. The pattern of increased international collaborations is more obvious in EU countries than in the USA. Japan and the average EU country have increased their international collaborations three-fold, while in the USA a two-fold increase could be observed (Archibugi and Coco, 2004).

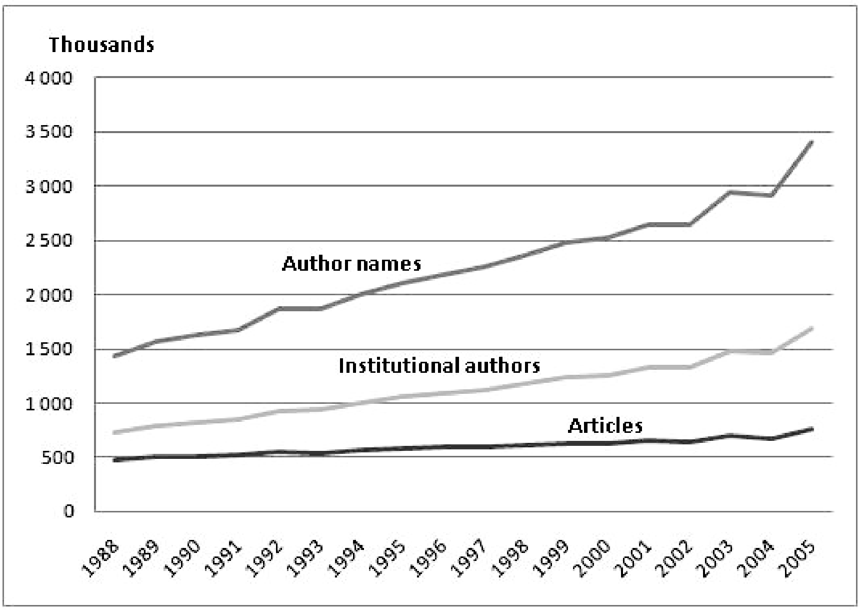

Figure 3: Worldwide Science & Engineering articles, institutional authors[1], and authors names: 1998-2005. Source: National Science Foundation.

The growth of collaborative research owes much to the provision of funding from governments intent to encourage interdisciplinarity and interactions between different sectors (Rigby and Edler, 2005). Bozeman and Boardman (2003) go as far as talking

[1] "Institution" refer to different departments or units within the same institution; multiple listings of the same department or unit are counted as one institutional author.

about a new "era of inter-institutional research collaborations" referring to US science and technology policy moving from decentralised support of small investigator-initiated research projects to large scale and oftentimes centralised, block grant-based, multidisciplinary research. In tandem, the scale, scope and diversity of funding programmes has increased and transnational, national and regional initiatives all play a larger role.

Other factors that can help explain the increasing collaboration rate is the growing use of expensive and large-scale instruments. When a research group does not have the required budget needed to purchase instruments it has to find someone to share the costs with. The interdisciplinarity of a field is also a factor which could affect a research group's tendency to collaborate. If the required knowledge cannot be found within the own research group field it has to be attained through external collaborations.

1.6.3 Factors related to the patterns of research collaboration

An actor that decides to engage in collaborative research has to choose with whom to collaborate. Understanding the factors that affect the selection of partners is crucial for understanding the dynamics of the resulting collaboration as they link micro-, meso- and macro-level collaborations.

The geographical proximity, as well as cultural, and language similarities between two countries have an impact on the way actors in different countries collaborate (Zitt et al., 2000). Katz (1994) found that distance has a negative correlation on the collaboration frequency.

In a study on university and industrial partnerships Geisler (1995) suggests that organisations are more likely to engage in collaborative R&D if they perceive their partner organisation to be resource interdependent. The report also concludes that cooperation is more likely to survive, if there is trust, good will, favourable prior beliefs, mutual psychological commitment and prior relations between parties. Finally, high degrees of institutionalisation and high intensity (active interaction) in the relationship increase the likelihood of its survival. Studying the EU Framework Programmes, Pohoryles (2002) and Okubo and Zitt (2004), have shown that the sustainability of research networks is dependent on good personal relations, institutional ties, availability of funds, joint publications, easy communication and the sharing of a research paradigm.

1.7 SECTORAL KNOWLEDGE TRANSFER – PATENT CITATIONS

1.7.1 Different patent systems

The rights conferred by a patent vary from country to country, even though a number of international agreements exist. Despite aims to harmonise legal proceedings to bring patent offices across the world closer together, the global landscape of patents remains multi-faceted. The three major patent offices are the European Patent Office (EPO), United States Patent and Trademark Office (USPTO), and the Japan Patent Office (JPO). The EPO examines and grants "European patents" which - subject to formal requirements - then acquire the same status and influence as national patents under national laws. A number of differences between the offices exist and are mostly rooted in diverging legal and cultural approaches. Both the EPO and the JPO are based on a first-to-file system, meaning that in the case of competing applications by independent inventors the legal right is granted on the basis of the filing or priority date of the application, regardless of the date of actual invention. The USPTO, however, operates on a first-to-invent system, meaning that a patent is granted to the person who first conceived and practiced the invention, rather than to the person who first filed the invention with authorities. Other differences also exist: the USPTO and the JPO have to tackle significantly more applications than the EPO because of different standards and proceedings. In the US the scope of patentable things is broader and includes business methods and software, unlike the EPO that requires patents to have a technical effect. Also, inventors in the United States are able to file successor applications to a pending application to update it with any improvements made by the inventor. The JPO has a higher rate of patent applications in comparison to the EPO mostly due to the fact that in Japan one and the same invention could constitute up to 10 different patents, with every technological aspect of the invention filed independently. This is not the case in Europe.

Caution is also needed when interpreting and comparing citation patterns among patents issued in Europe/Japan to the US. Since 2001, the applicant and/or its attorney for a US Patents and Trademark Office (USPTO) patent have had a ‘duty of candour’, which implies that they must, by law, include any prior art of an invention. This is not the case in Europe and Japan. Applicants to the USPTO, therefore, tend to not limit their references, but rather to include citations that are in any way relevant to their invention, in order to minimize the risk of their application being rejected. As a consequence, USPTO patents on average include more citations than Japanese or European ones (Alcácer and Gittelman, 2006). Meyer (2000) discusses other differences between the US and European systems, such as the generally lower education levels of US examiners, which results in the inclusion of citations that are not always directly relevant to the patent. In addition, the heavy workload of US examiners and the focus on English language documents could have a negative impact on the search for relevant documents. EPO patents’ citations are more focused towards a limited number of fields and therefore EPO applicant citations might be closer to real knowledge transfer than USPTO applications. This has further been supported by Michel and Bettels (2001) who argue that the general coverage of underlying technology in the US search report is much broader than in the EP. These differences should be taken account when analysing patent data from different countries.

1.7.2 Who includes citations

Citations can be included in a patent by 1) the inventor, 2) the legal expert, 3) the applicant, and 4) patent examiner. In studies based on patent citations, the assumption is made that all citations were included by the inventor, and thus they are an indicator of the knowledge that he/she possesses. However, in the course of a patent application, citations can be added first by the applicant, who is not necessarily the inventor, but might be the company that owns the property rights and/or second by the applicant's legal advisers. It is not possible to distinguish among these citations. These citations are reviewed by the patent examiner to determine the novelty of the invention and to decide whether to accept or reject the application. The examiner compares the invention with the prior art and can insert on the front page of the patent application, what in his/her opinion are other relevant references. Of course, an applicant can choose to leave out some of the knowledge base to make the invention appear more novel. It is the examiners' responsibility to identify whether this has been done and to add relevant information and remove non-relevant citations. Therefore, the references on the front pages of patent applications are not necessarily the same as in the original application. In a study carried out by Schmoch (1993) only 8.4% of all the citations included on the front page of EPO originated from the inventor. Another study conducted by Jaffe et al. (2000) also differentiated between applicant and examiner added citations and found that one-third of applications included a high level of the knowledge from cited works, one-third learned more about related knowledge during completion of the application and one-third had no prior knowledge about the cited information. Both these studies indicate that examiners add an important proportion of the citations included, and that aggregated citations can be a misleading measure of inventor knowledge transfer.

1.7.3 Patent characteristics

There are some other characteristics of patents that may condition the distribution of citations. Earlier investigations have mainly focused on geographic factors, time effects and technology classes.

Geographic characteristics

The geographic location of a citation can be used to trace the origin of a knowledge source and to determine how far awareness of an invention has diffused. Several authors have studied the differences between national and international citation patterns. Jaffe and Trajtenberg (1996) found that knowledge spillovers are geographically localized but that disparities fade over time. Almeida and Kogut (1999) concluded that local knowledge spillovers are more likely in high-tech regions such as the Silicon Valley, the Boston area and Austin in Texas. The idea is that local innovations are stimulated by nearby technology developments.

Technology classes characteristics

Thompson (2006) tested the correlation between technology classes and geography and found national citations were more common when both the patent and the cited patent belonged to the same technology class.

When differentiating between applicant and examiner added citations, Sampat (2004) provides some empirical evidence, using USPTO data, and suggests that applicants are more likely to include citations in fields where patenting is important such as chemicals and pharmaceuticals. Drawing on the previous literature, Sampat argues that in fast-moving, more technology intensive classes, the ability of examiners to access current information is limited, and applicants are better informed about the closely related prior art and therefore include more citations than examiners. In technologies that are less science intensive, citation rates between examiners and applicants will generally be more similar. In the case of EPO patents, Criscuolo and Verspagen (2008) find that the share of applicant citations is higher in chemistry and materials and lower in semiconductors, telecommunication, audiovisual and information technology.

Also the type of citation has to be taken into account when comparing technology classes. Callaert et al. (2006) find a larger number of EPO examiner citations in patent-to-patent citations in mechanical engineering and machinery and patent-to-paper citations in chemistry and pharmaceuticals. In the USPTO, process engineering and ocial equipment ranks first for patent-to-patent citations and Chemistry and Pharmaceuticals for patent-to-paper citations. Sampat (2004) finds that examiners in the US face particular challenges in identifying prior art in emerging technological fields, focusing on nanotechnology. In Spain, Acosta and Coronado (2003) show that 85% of patent applicants' patent-to-paper citations are concentrated in only three sectors (chemistry, pharmaceuticals and biotechnology). However, patent-to-patent citations predominate in high and medium-high technology sectors, such as electrical engineering and instruments, which is why Leydesdorff (2004) suggests that a sector such as biotechnology is not a valid model for how university-industry interactions occur in general.

Time effect characteristics

Time effects seem to be influential in the distribution of examiner citations. US examiner time constraints have become tighter, with the increase in the number of patent applications outpacing increases in the number of examiners (Merrill et al., 2004). Time and resource constraints necessarily limit the comprehensiveness of examiners' prior art search, producing 'citation inflation' in US Patent and Trademark Office (USPTO) patents (Hall et al., 2000). Callaert et al. (2006) confirm an increase in USPTO citations between 1991 and 2001, while for European Patent Office (EPO) patents the average number of examiner citations per patent in the same period has been constant (or even slightly decreasing).

2 STUDY RATIONALE AND AIMS

2.1 STUDY RATIONALE

In the context of the March 2000 Lisbon strategy to make the European Union the "world's most dynamic and competitive knowledge economy", knowledge transfer plays an important role. The idea of a European Research Area (ERA) grew partly out of the realisation that research in Europe suffered from several weaknesses. A weak environment for stimulating high quality research and exploit results was one of the most relevant identified weaknesses and is an important backdrop for this thesis. In order to increase the competitiveness of European R&D it was considered essential to increase research funding, stimulate and support development of new scientific fields and facilitate collaborations by, for example, removing existing barriers, both in a geographical and sectoral sense (European Commission, 2000). As a result, European research policy came to move beyond the Framework Programmes and increasingly addressed broader issues, such as the mobility of researchers, enhancement of research infrastructure, more standardized patent systems and better coordination activities (Banchoff, 2002).

The introduction of new funding schemes and policies aimed at increasing knowledge flow between countries and sectors in Europe has increased the demand for studies of the impact of existing policies, as well as the development of related indicators that are relevant and accurate. Underlying this important need to justify new research funding and to develop efficient and effective policies is a broader and even more important concern to better understand the sources of and mechanisms by which new discoveries and innovations are developed and what their socio-economic consequences are.

As a result, the use, meaning and implications of different indicators have come to play an increasingly important role in policy-making. More specifically, indicators can provide evidence whether a specific result has or has not been achieved through a specific policy mechanism. Indicators exist at a number of different levels from input, output, to result (outcomes) and impact. In addition, they can also provide information about the context or the assumptions under which these policy interventions operate.

In this thesis I put a special emphasis on input indicators in the form of funding and outcome indicators mainly in the forms of publications, patents and collaborations.

In the first two papers, co-authorship patterns are examined to study the relation between European research policy and transnational research collaborations, representing geographic knowledge flow. Earlier studies show that countries have an increasing collaboration rate with other countries and suggest a growing internationalisation of science (Smedby and Trondal, 2005; Zitt and Bassecoulard, 2005). However, it is not clear whether these changes have taken place equally among EU member states or if some countries behave differently. Therefore this thesis sought to capture a more differentiated picture. It focused on the way European researchers collaborate among each other in comparison to collaborating with partners outside Europe. The thesis also distinguished between bilateral and multilateral collaborations and it sought to contribute to a better understanding of factors that are important when developing policies supporting research collaborations.-In the third paper the relevance and use of bibliometric indicators as a means to indicate the knowledge contribution of researchers is investigated. Since the number of authors in a paper has increased over the years, it is difficult to determine the individual contribution of authors. Therefore it is of interest to study the relation between corresponding author and author position to determine the utility of these indicators for assessing and evaluating the contributions of individual authors.

Finally, in paper IV and V the use and utility of patent citations to justify research funding and to evaluate performance is studied. The policy relevance of this examination is rooted in the fact that patents and the citations they include are a frequently used measure of knowledge flow between sectors (Hall, 2004). In addition, they are widely discussed at the European policy level in the context of concerns about patent system that are fragmented and differ among countries. The majority of existing studies treat patent citations in a rather undifferentiated way. They pay limited attention to how differences in geographical location, techno-economic specialisation, patent systems and with regard to who is the knowledge contributor lead to different practices of citation use in patents, which in turn may influence the meaning of this indicator in assessing knowledge flows.

2.2 THE MAIN AIM

The aim of this thesis was to study the dynamics of knowledge transfer in Europe and to examine how knowledge transfer can be measured and analysed through different indicators.

2.3 SPECIFIC AIMS

The specific aims of the studies on which this thesis is based were to:

- study to what extent researchers in EU countries collaborate within EU and with outside partners and whether differences between countries and scientific fields can be discerned;
- examine the relation between funding mechanisms and dynamics of research collaborations;
- investigate the patterns and relevance of author position in internationally co-authored papers related to their degree of contribution;
- analyse the utility of citations in patents as an indicator to measure knowledge transfer in regions with low absorptive capacity (weak innovation activities and a low tech profile);
- analyse the distribution of examiner and applicant added patent citations according to patent characteristics;

The first two aims are mainly contributing to the understanding of knowledge transfer within Europe. The other three aims are mainly contributing to a better understanding of how indicators can be developed and used to measure and analyse knowledge transfer.

3 MATERIAL AND METHODS

3.1 BIBLIOMETRICS AS A METHOD

Knowledge transfer between researchers can take place at different levels and in a multitude of ways. This also implies that knowledge transfer should be studied using a variety of different methods. The most common way to communicate that a research collaboration has resulted in something is through a joint publication. The methodology to study articles published in peer-reviewed journals is called bibliometrics and has been developed into a research field in its own right. Modern ways of studying publications have largely been inspired by de Solla Price and the seminal work he carried out. In the book "Little Science-Big Science" (1963) he analysed the publication of research results and presented a number of quantitative evaluation techniques. De Solla Price was the first to examine the increasing trend of collaborations among chemistry researchers by using bibliometrics.

The argument for using co-authorship as an indicator of research collaborations is based on the fact that research results do not become acknowledged and recognised until the rest of the research community is aware of it. It is therefore essential for researchers to publish their results. If several researchers have contributed to the work that led to the results it is common practice, according to the Vancouver Group, that these should be included in the author list of the written work (ICMJE, 1997). Using co-authorship metrics as a method also makes it possible to examine a large set of data, which is of particular importance when studying international collaborations.

3.2 PATENTOMETRICS AS A METHOD

Quantitative studies of patent information can be called "patentometrics" or "patent bibliometrics" (Narin, 1994). Some of the main reasons for using patents as an indicator for knowledge transfer and technological change relate to the quantity of available data, its accessibility and the different types of details included in a patent document (Griliches, 1990). A patent file contains information such as the name of the inventor(s) and their addresses, as well as the name of the organisation to which the patent right may be assigned. It also lists one or more patent classes to which the patent right may have been assigned by the examiners, cites a number of previous patents and sometimes also scientific articles to which the particular invention may be related and it also comprises a description of the invention. This information has been used not only in simple aggregated form to document patent activity for a specific year or country, but also for more differentiated and comprehensive analyses of citation networks and patterns for a technology or a scientific field, or the degree of interdisciplinarity involved.

3.3 DEFINITIONS

A number of different definitions need to be clarified before presenting the methodology used in this thesis.

International collaborations have been defined in a number of earlier papers. Katz and Martin (1997) differentiated between different levels of collaborations and made a distinction between international and intranational collaboration forms. They defined international collaboration as collaboration between scientists with addresses in different countries while the attribute intranational refers to collaborations within one country. Collaborations can also be categorized into *internal collaborations*, defined as collaborations within the researchers' own institution/organization (intra-institutional), and *external collaborations* (inter-institutional), which can then be further divided into *national collaborations* and *international collaborations.*

For the purpose of this thesis *intra-EU collaborations* are those that take place between scientists from EU member States, candidate and other associated European countries. *Extra-EU collaborations* are those that take place between scientists from a EU Member State and scientist from other, non-EU countries. *Global collaborations* involve at least two countries from EU member States, candidate and other associated European countries and one or several non-European countries.

The size of a country was defined by the number of articles published in the country and is presented on a log scale, to illustrate the differences between country sizes. In this thesis I referred to small countries particularly with regard to Slovakia, Portugal, Ireland, Czech Republic, Hungary, Greece and to large countries with regard to UK, Germany, France, and Italy. It is difficult to make a clear, overall distinction between small and large countries, because the number of articles varies across scientific fields.

3.4 STUDY DELIMITATIONS

The scope and focus of the study was delineated in several ways. In Paper I, a comparison between different scientific fields was included and in Papers II-III the focus was then put on collaborations in Life Science. In Paper I, the classifications of the publication database Web of Science (WoS) were used to analyse co-authorship in different fields. The scientific fields studied included Agriculture, Biology & Environmental Sciences (ABES), Clinical Medicine (CM), Life Science (LS), Physical, Chemical & Earth Sciences (PCES), and Engineering and Computing & Technology (ECT). In Papers II and III, the classification of OST (*Observatoire* des sciences et des technologies) (Zitt and Teixeira, 1996) was used to identify articles published in the field of Life Science, which includes Fundamental Biology and Medicine. OST's classification overlaps with the WoS classification but also covers a number of fields that are included in the ABES.

Although the European Union consists since 2007 of 27 Member States and 8 Candidate and Associated Countries the first three studies of the thesis have confined themselves to focus on 18 countries, see table 2. These countries accounted for 99% of the total production of co-publications of the 25 member states before 2007 (Mattsson et al., 2008). As for funding, these 18 EU Member States accounted for 99 % of the total research funding in FP6.

Table 2: Classification of studied countries.

Abbreviation	Country
AT	Austria
BE	Belgium
CZ	Czech republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
NL	Netherlands
PL	Poland
PT	Portugal
SE	Sweden
SK	Slovakia
UK	United Kingdom

In Papers IV-V the focus is on the regional level, with the selection of three Spanish regions: Alicante, Castellon and Valencia. The reason for choosing these regions was that they are regions with low absorptive capacity, whereas most other studies focus on the use of patent citations in high tech regions. The main features of these region with low absorptive capacity are:

- The low-tech profile of its economic structure (predominance of microfirms in services and traditional manufactures).
- The weaknesses of its innovation activities (innovation occurs, but it is mostly incremental and through machinery and equipment acquisition, with little expenditure on R&D).
- The scarcity of qualified personnel at firms, even in knowledge-intensive sectors.
- Policy emphasis on enhancing technology transfer (similar to high-tech regions or countries).

Since the aim is to study knowledge transfer between sectors only industry patents were included in the studies. The years covered include 1999-2003.

3.5 DATA COLLECTION AND PROCESSING

The data used for carrying out the co-authorship analysis, in Papers I-III was retrieved from the Thomson Scientific (formerly Institute of Scientific Information) online database Web of Science (WoS). That database is the most commonly used data source for carrying out bibliometric analysis even though it has a number of limitations (Moed, 2002).

The WoS database includes a number of different types of scientific documents such as book reviews, communications, conference abstracts, and letters. For the purpose of studying “new science” - not reviews and other commentary or republication of already published materials- only articles and notes were selected for further analysis. The information found in the database includes title of the article, language, names of authors, addresses of authors, journal information, reprint address (address of the corresponding author), and references. For the purpose of this thesis, information regarding author name and address(es), as well as author position were further explored.

Information regarding the EU Framework Programmes, studied in Paper II, was collected from the Community Research and Development Information Service (CORDIS). Only projects related to programmes focusing on Life Science were included. CORDIS is an information resource dedicated to European Union research and development activities. It includes information about all Framework Programmes and is also the official website (www.cordis.lu) for the publication of all calls for proposals. Information such as participating organisation name, type of organisation, address of organisation, and the role of the organisation (coordinator or participant) was further analysed.

An article/project was assigned to a country based on the information in the address field. Each link between distinctive countries was rated at a maximum once per article/project. Only publications and projects in collaboration with researchers from at least two countries were included in the studies. Bilateral collaborations were analysed by selecting articles or projects in which authors from only two countries participated while multilateral collaborations, also referred to as networks, were analysed selecting articles and projects, in which more than two countries participated.

Table 3. Definitions of different types of collaborations

	Nr. of countries
International collaboration	≥2
(International) bilateral collaboration	=2
(International) multilateral collaboration	>2

The majority of journals require that each article should identify one of the authors as the corresponding author. The intention is that readers should be able to comment and ask questions about the published article. The role of the corresponding author is therefore to respond to these questions but also to be responsible for correspondence with the journal before acceptance and publication. The corresponding author should also be able to declare any competing or conflicting interest and to explain the presence and order of co-authors (ICMJE 1997).

In the address field of WoS the corresponding author is labelled as reprint author, but in this thesis (Paper III) referred to as corresponding author. Before 1998, less than 60% of the publications had this tag while from 1998 onwards on average 98% include the reprint label.

As for Papers IV-V the source of patent application information was collected from the Spanish Patent and Trademark Office (OEPM) covering the years from 1999 to 2003. The OEPM database includes a field for the name(s) of the patent applicant. Identifying how this relates to specific firms was not straightforward and involved checking each patent to classify and standardize it. In total 1382 patents of the Valencian Community registered between 1999 and 2003 were collected. The number of citations in the full text of the patent application form for each patent was counted. Also the number of citations in the prior art report (included by the examiner) was counted. The citations were thereafter classified according to

- scientific or technologic nature, i.e. NPRs or PRs, respectively.
- geographic location of PRs and Thomson Scientific's Science Citation Index (SCI)-NPRs.
- year of application
- route of protection (national vs. international)
- ownership or cooperation (with a firm, with a research centre, with an innovation/technology centre, and with an individual)
- technology class: each of the eight sections at the first level of the International Patent Classification (IPC)
- economic sector: two-digit Statistical Classification of Economic Activities (NACE) of applicant firm

3.6 STATISTICS

When comparing the five scientific fields, in Paper I, the percentage of national production and international collaborations (intra-EU and extra-EU) out of the total numbers of publications in the field was calculated.

The concept of preferential partnership, Paper II, was calculated using the Probabilistic Activity Index (PAI) (Okubo and Zitt, 2004). This index is correcting for the size of the countries and calculates the observed co-authorship between two countries to the value that would be expected from their respective size. If size is not corrected for, large countries such as the UK, France and Germany would be expected to be the first partner for the majority of countries studied.

$PAI = (Cx,y * T)/(Cx*Cy)$

where Cx,y is the number of co-authored papers between country x and y
T is the total number of international co-authored papers by the eighteen countries
Cx is the total number of internationally co-authored papers by country x
Cy is the total number of internationally co-authored papers by country y

A preferential partnership can be considered to exist when the PAI-index between two countries is more than one and thereby stronger than what would be expected from the publication output. The PAI index was also used to identify preferential collaborations in the FPs.

In Paper III, the first author's institutional origin was manually compared to the last author's origin. A probability sampling method (a = 0.05), more specifically systematic random sampling, was used to select a representative sample of 1000 articles proportional to the size of the selected countries. To examine, if there were any differences between countries I used the Wilcoxon Matched Pair Test

In Papers IV and V differences between variables were calculated using ANOVA and the non-parametric test Kruskal-Wallis ANOVA.

The data in this thesis was analysed by using the software programme Statistica (Tulsa, USA).

3.7 METHODOLOGICAL CONSIDERATIONS

A number of methodological considerations need to be taken into account that qualify and delineate the findings reached. As this thesis has shown, bibliometrics and patentometrics can provide important empirical facts to support evidence-based research policies targeting knowledge transfers. It should be remembered, however that the techniques are not perfect and a number of different shortcomings can be discerned both for the databases and the analytical tools used.

When comparing the publication rates of different countries it is important to remember that not all articles published are listed in the WoS database, which only includes the largest and leading journals. Smaller, more local and nationally oriented journals are not included. Also, language is an issue, since WoS tends to have an Anglo-Saxon and American bias. This benefits researchers that have English as a mother tongue or countries where the level of English as a second language is well advanced e.g. the Nordic countries and the Netherlands.

Different countries also have different publication cultures. In some countries publishing in peer-reviewed journals is the norm among PhD candidates, while in others it is more common to publish monographs. Also, the way of structuring authorship order and corresponding authorship position in a paper can differ between countries, as shown in Paper III.

Bibliometrically-supported comparisons between fields also need to be approached with caution. Differences here may be due to different publication strategies. Mathematicians and other theorists tend to publish less than researchers in experimentally intensive fields such as in most of the sub-fields of Life Science (Moed et al., 2004). Publications in the field of physics do on average involve more authors than articles in the field of chemistry (Archibugi and Coco, 2004). Since only Paper I includes several fields and Papers II-III focus on Life Science, where publishing research results in a journal article is the dominant way, this problem is limited even though publication cultures between different sub-fields can vary. Bibliometricians recognise these challenges and a number of normalisation methods exist with the aim to minimise such differences (Wallin, 2005).

With regard to mapping collaborations it is important to keep in mind that some forms of collaborations do not result in co-published articles but may rather involve the sharing of research infrastructure, exchange of material or samples, intellectual property or some kind of informal collaboration which involve the stimulation of knowledge creation. These could be additional outcomes of collaborations taking place in FP projects for example.

Measuring international collaboration as an outcome of funding instruments also faces challenges of attribution since nearly all research today is performed with multiple funding and very few research groups have funding from only one source. To actually determine the effect of an instrument or programme one must study both the collaborations that existed before a specific funding mechanism was introduced and those collaborations that actually were initiated as a result of the funding.

For patents limitations similar to those encountered for publications exist regarding field differences and country differences. In some economic sectors, such as ICT (information and communication technology) open-source approaches have evolved into a significant way of knowledge sharing (Blind et al., 2005). In other fields such as in the pharmaceutical sector patenting is the established norm. It is therefore important to keep in mind these economic sector differences when using patents as an indicator.

In addition, as was mentioned in section 1.7.1 different patent systems exist in different countries and these differences need to be considered as well when using patent as indicators. More specifically, patent citations have been criticized by several scholars as being ‘noisy’ when used to measure knowledge transfer, with the examiner added citations producing most of the ‘noise’. These critiques can be addressed in three different ways. The first is to adopt more careful wording to refer to the relation between patents and their citations: ‘interactions’, ‘links’ or ‘linkages’, ‘vicinity’, etc. (Tijssen, 2001; Callaert et al., 2006). A second is to use patent data in alternative ways to visualise relations within the knowledge base, e.g. through co-classification in technology classes to show perhaps that countries are not an appropriate unit of observation (Leydesdorff, 2008). A third is to promote a quantitative approach to the difference between examiner and applicant citations.

In this thesis I have studied the usefulness of patent citations when measuring knowledge transfer in different context further. The limitations and recommendations will therefore be further discussed in the following chapter, Results and Comments.

4 RESULTS AND COMMENTS

In this section, results from Papers I-V and other unpublished results are presented. In relation to the key results some comments are provided. The first section concerns the results from Papers I and II from which it can be concluded that international collaborations have increased. Further, differences between fields and countries are presented (Papers I and II and unpublished results). In the third section, the project management role and authorship position for scientists from different countries are discussed (unpublished material). The results demonstrate that researchers tend to collaborate in global multilateral collaborations rather than European ones (Paper II). Next, the results show that the corresponding author is most likely to appear first and second most likely to appear last in the byline of internationally co-authored papers (Paper III). Furthermore the results from Paper IV-V are presented. The main results from these two studies is that a number of different factors need to be taken into account when using patent citations as a measure of knowledge transfer and related indicators need to be further developed.

4.1 THE INCREASE OF INTERNATIONAL COLLABORATIONS

In accordance with earlier studies, an increase in international co-authorship frequency, over the period 1998-2003, could be discerned among the 18 studied countries, Poland being the exception (Figure 4).

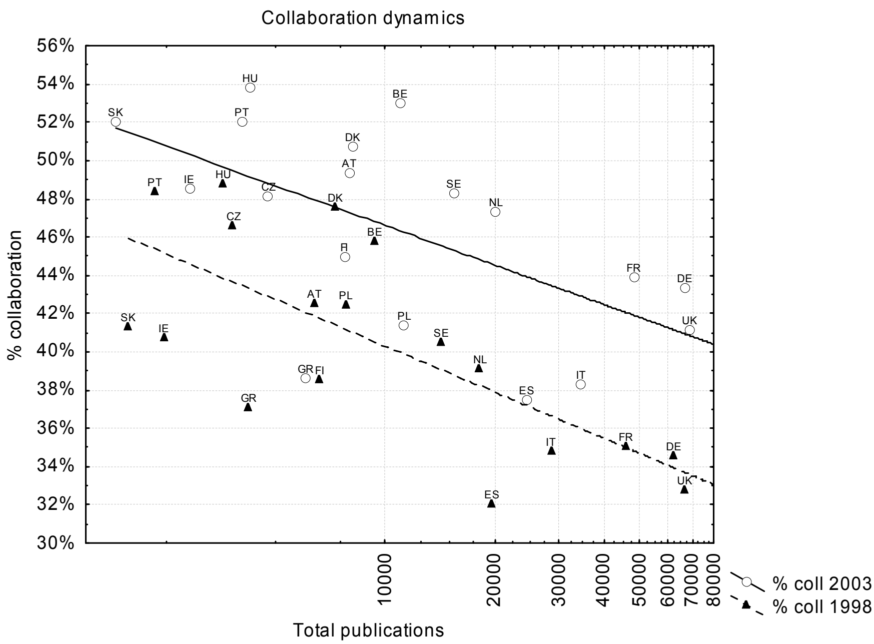

Figure 4: Collaborations (share of total publications) and total number of publications for EU-18 countries 1998 and 2003.

As can be seen in Figure 4, a significant difference in co-authorship rates between small and big countries both in year 1998 (p=0.0014) and 2003 (p=0.0290) could be observed. Smaller countries did collaborate more than bigger ones and the total publication rate for smaller countries has increased more than for bigger countries (p=0.0429).

Multilateral (involving more than two countries) co-authorships, during 2003-2005 accounting for 36% of all co-authorships, have increased more than bilateral (involving only two countries) international co-authorships in all studied countries (Figure 5). Still the majority, 64%, of international co-authored articles involve only two countries. No significant differences between the size of countries could be discerned.

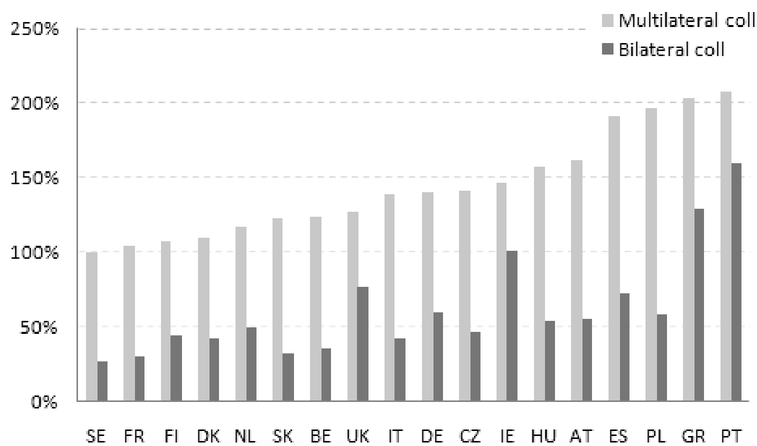

Figure 5. Relative increase of multilateral and bilateral collaborations between the periods 1995-1997 and 2003-2005.

4.2 SCIENTIFIC FIELD AND COUNTRY DIFFERENCES

Collaboration patterns differed between scientific fields. In Clinical Medicine (CM) extra-EU collaborations (collaborations with non-European countries) did not differ significantly between countries of different sizes while in Life Sciences (LS) bigger countries (UK, DE, and FR) had more extra-EU collaborations than smaller countries.

In Figure 6 the share of total multilateral collaborations, is shown for the different countries/fields. There was a significant difference between countries of different sizes with regard to their share of multilateral collaborations in the total number of publications in all scientific fields except Agriculture, Biology & Environmental Sciences (ABES), with small countries participate more in multilateral collaborations.

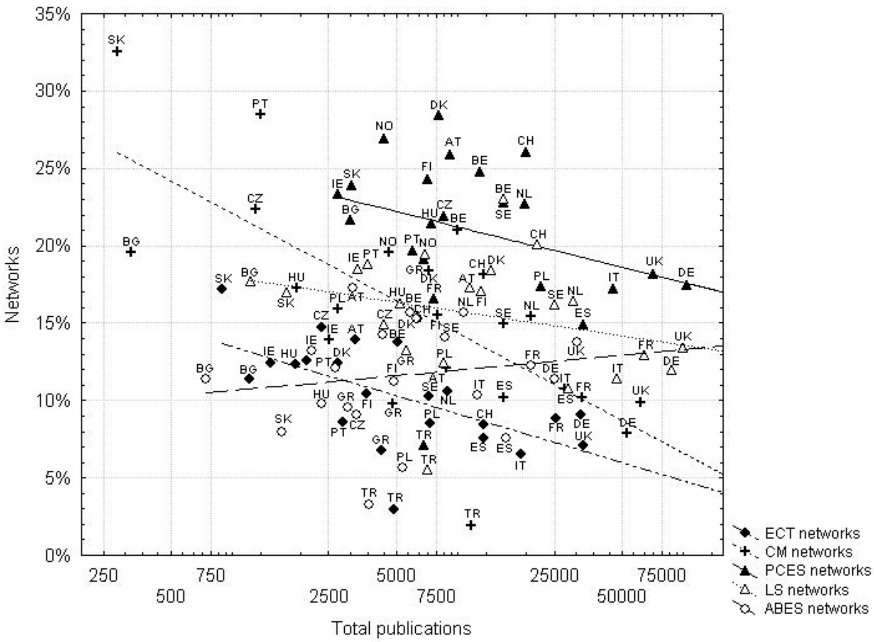

Figure 6. Share of multilateral collaborations in the five scientific fields.

In addition, there was a significant ($p<0.05$) difference between the intra- and extra-EU multilateral collaborations in all fields, with intra-EU multilateral collaborations including the largest number of countries. CM was the field with most countries involved (3.67 countries/article). Intra-EU multilateral collaborations were also most present in CM, where multilateral collaborations accounted for on average 26% of all the EU collaboration publications. In LS, Physical, Chemical & Earth Sciences (PCES), Agriculture, Biology & Environmental Sciences (ABES), and Engineering and Computing & Technology (ECT) intra-EU multilateral collaborated papers accounted for between 17-21% of the total EU collaborations. Multilateral collaborations were less common in extra-EU collaborations in all fields.

Since differences between countries and fields can be due to field specialisation, the rate of publications in a field in relation to the total production was studied (Figure 7).

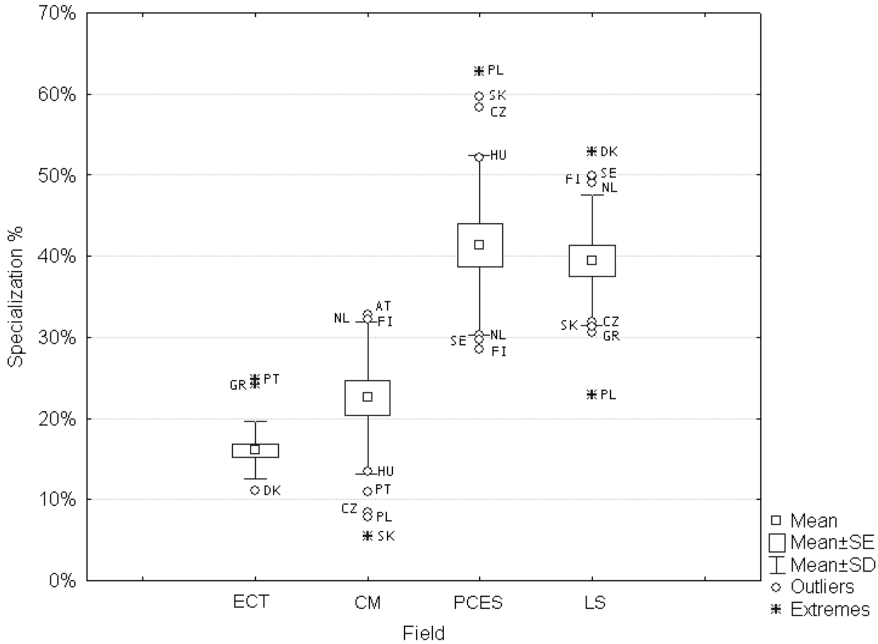

Figure 7. Country field specialisation.

In ECT both Greece and Portugal appear as extreme values with a relatively high share of publications. In CM none of the EU18 countries had a significant specialisation, no country emerges as extreme value. It is noteworthy that the Czech Republic, Poland, and Slovakia had less than 10 % of its publication volume in CM. In PCES only Poland had a specialisation. In LS, Poland appears as an extreme value due to its low publication rate while Denmark was found to have a specialisation. When adding-up the specialisation rate in the four fields it exceeds 100%, due to the fact that some journals (e.g. Nature and Science) belong to more than one scientific field. In table 4, the correlation between the above mentioned specialisation and the rate of intra/extra-European collaboration is presented.

Table 4. Relation, p-value, between specialisation and collaboration mode.*Excluding SK

Field	Intra-European	Extra-European
ECT	0.6129	0.0197
CM	0.1986*	0.5631
PCES	0.4718	0.0915
LS	0.4771	0.1761

The table shows that specialisation in a field, measured by the rate of publications in a field, does not impact the collaboration rate neither in intra- nor in extra-European co-authorship. The exception includes extra-European collaborations in ECT. One explanation could be that ECT involves more companies than other fields and a specialisation would therefore also involve more collaboration with global companies.

4.3 COUNTRIES HAVE DIFFERENT ROLES

The role of countries in multilateral collaborations was further investigated at the European level by studying the share of corresponding authors with the help of co-publication data.

The results revealed that authors from Austria, Belgium, France, Italy, Netherlands and UK are more likely to appear as corresponding authors than authors from other EU17 countries. In the case of Czech Republic, Hungary, Poland, Portugal, and Slovakia the opposite was found. Multilateral collaborations with these countries rather tended to have other authors from EU17 countries as corresponding author.

The result also suggests that large countries such as UK, France, Italy, and Netherlands act more frequently as corresponding author in EU17 networks. New member states and especially Eastern European countries rather act as contributors.

In bilateral co-publications Germany, Italy and the UK stood out. In the case of DE and UK, the participating national authors appear less often as corresponding author compared to other collaborating countries. In Italy the opposite was found with Italian authors appearing as corresponding authors significantly more often than other authors from EU17. In country/country pairs where preferred partnerships exist, the share of corresponding authorship was often balanced and none of the countries appeared more often than the other e.g. Ireland with UK, Portugal with Spain, Sweden with Denmark and Finland.

Next, the rate of coordination in the Framework Programmes 5 and 6 was studied. Figure 8 illustrates the rate of countries taking on a coordination role in relation to the total number of project participation. The results are similar to the co-publication data with the large countries being more likely to take on a coordination role while smaller countries rather have a participating role.

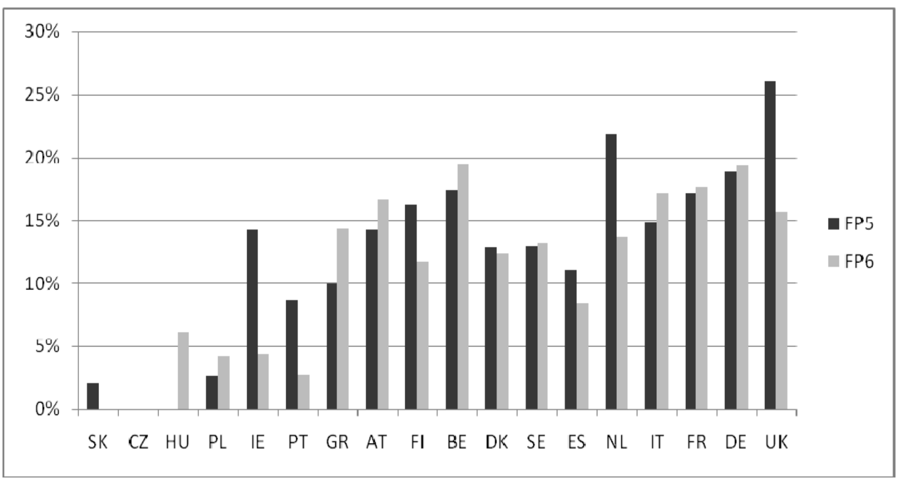

Figure 8. Share of projects where the researchers from a country have a coordination role in FP5 and FP6 in relation to the overall project participation rate.

4.4 RESEARCHERS TEND TO COLLABORATE IN GLOBAL MULTILATERAL SETTINGS

The participation rate in Framework Programmes indicates that the number of involved non-European countries has increased over the different Framework Programmes. In FP4, only 4% of projects involved non-European countries and Israel accounted for 86 % of these, probably due to the fact that it was the most scientifically active of the associated countries. In FP5, the participation rate of non-European countries increased to 8 %, with Israel accounting for the majority of these (69 %). In FP6, the non-European participation was pushed up to 31 %, a rise that is attributable to the inclusion under FP6 of a cooperation programme targeting developing countries. If only projects were considered that involved non-European countries, the US participated in 1% of such projects under FP4, in 15% under FP5, and in 7% under FP6. Participation rates for other non-European industrial countries are similarly low.

In accordance with the FP data, the co-authorship data showed that global multilateral co-authorship had increased significantly ($p<0.05$) between 2005-2003 and 1995-1997 and was also relatively more common than European multilateral co-authorship (Figure 9). Multilateral collaborations were more common in global co-authored articles where they made up on average 45% compared to 27% in only EU co-authored papers. Figure 8 also shows that global multilateral co-authorship has increased more.

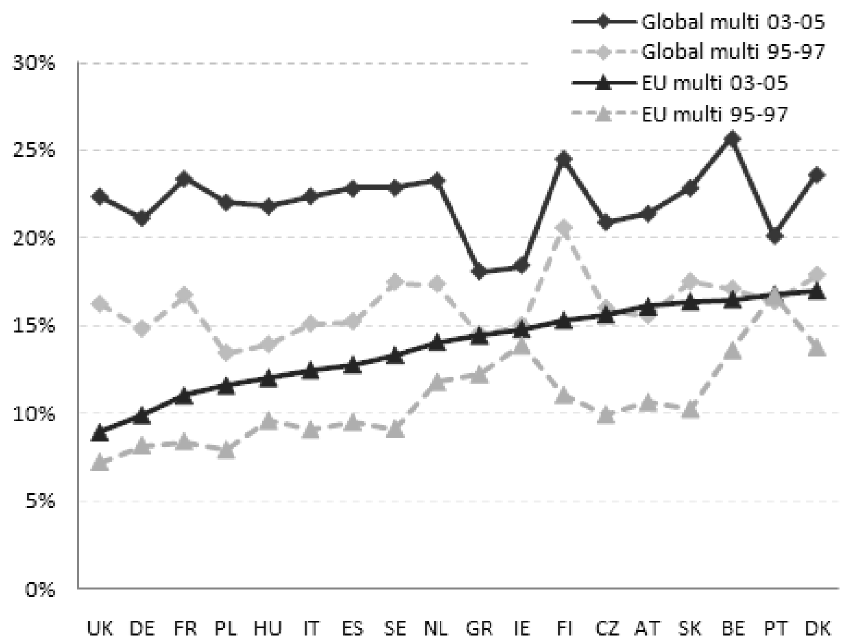

Figure 9. Share of global and EU multilateral co-authored papers of total numbers of co-authored papers.

All countries had at least one preferred collaboration partner, defined by PAI indexes[2] above one, in both co-publication data and the Framework Programme data. This can partly be explained by geographic, linguistic, or political proximity. Smaller countries and new member states are more likely to have preferential co-authorship links with other EU18 countries. The majority of preferential co-authorships had a higher degree of bilateral co-authorship in comparison to non-preferential collaborations.

4.5 AUTHOR POSITION RELATED TO CONTRIBUTION IN CO-AUTHORED PAPERS

When the number of authors in an internationally co-authored article was more than two, the first author accounted for the majority (52 %) of corresponding authors of all such papers (Figure 10). The last author appeared as corresponding author in 39 %. Only 9 % of internationally co-authored papers had a corresponding author positioned elsewhere in the byline. Corresponding author appearing as first author in the byline was more common when the number of authors was less than seven.

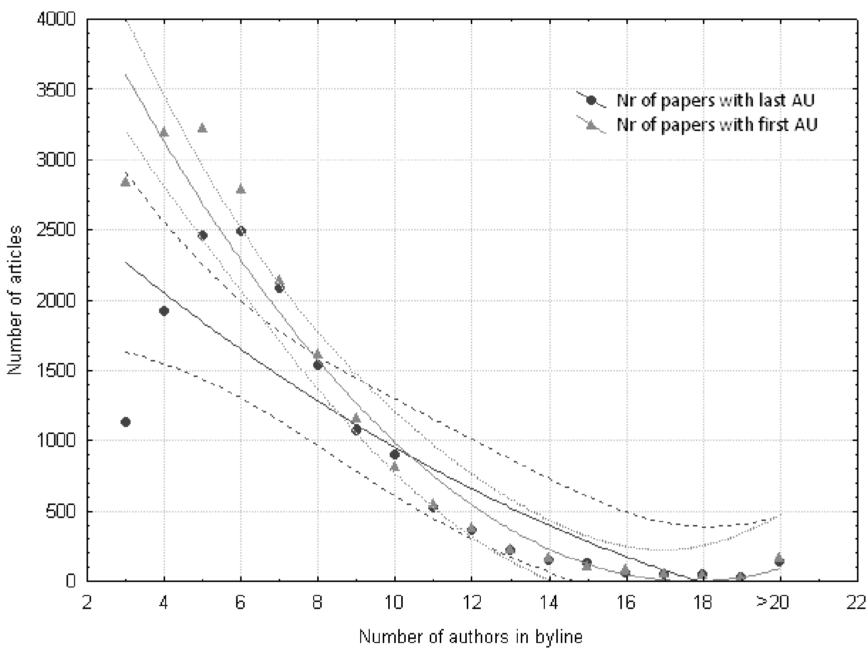

Figure 10. Distribution of papers by position of corresponding author.

There was a significant difference between countries with regard to the byline position of corresponding author (Figure 11). The explanation is due to different scientific publication cultures. When researchers based in Scandinavia, the Netherlands, and

[2] Probabilistic Activity Index (PAI) (Okubo and Zitt, 2004). This index is correcting for the size of the countries and calculates the observed co-authorship between two countries to the value that would be expected from their respective size.

former Eastern countries were corresponding authors it was more common that the corresponding author appears first in the byline.

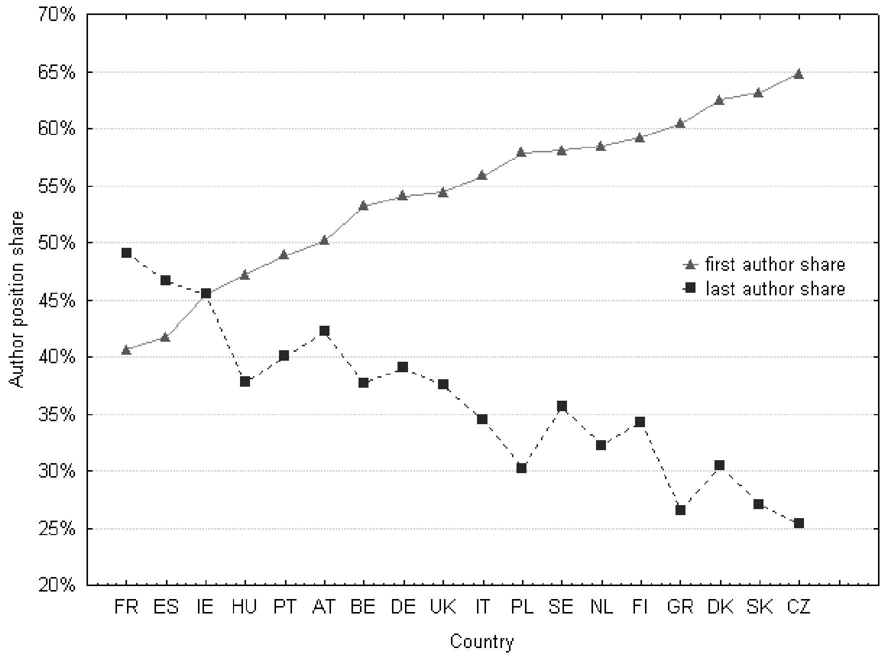

Figure 11. Byline position differences between countries.

4.6 INDICATION OF INVENTOR'S KNOWLEDGE BASE

In table 5 the number of citations per patent examiners is presented. The average number of examiner added citations is 4.83. Neither the average nor the mode is changing significantly when excluding patents with applicant added citations. The results suggest that most examiner citations originate with the examiners themselves and that examiner citations are not very representative of applicant knowledge flows in a region with low absorptive capacity such as the Valencian Community.

Table 5. Examiner added citations with and without applicant citations.

	Nº of patents	Nº of examiner citations	Average	Mode	Frequency
All patents	571	2758	4.83	4	23%
Patents without applicant citations	409	1987	4.86	4	22%
Patents with applicant citations	162	771	4.76	4	24%

Citations as an indicator to trace science-industry links

As can be seen from table 6 only 2% of examiners citations are non-patent references (NPRs) while 18% of applicant added citations are NPRs. This suggests that the scientific knowledge base among companies is low and rather based on technological knowledge.

Table 6. Type of citations added by examiners vs. applicants.

Type of cited reference	Nº of examiner citations	Nº of applicant citations
Patent References	2707 (98%)	547 (83%)
Non-Patent References	51 (2%)	122 (18%)
Total	2758	669

Comparing the rate of examiner and applicant citations between different technologies and between different economic sectors shows that science based technologies and sectors do not present many more citations per patent than more non-scientific technologies and sectors. Differences between applicant and examiner added citations can mainly be seen in non-scientific based technologies where the number of applicant citations is much smaller compared to examiners. In the supplier-dominated and production-intensive sectors, for example, examiners cite more when applicants cite less.

Table 7. Patent citations added by examiner and applicants in different technology classes.

Technology class (IPC)	Avg. Nº of examiner citations (mode; frequency of mode)	Avg. Nº of applicant citations (mode; frequency of mode)	Sig.
A. Human Necessities	4.48 (3; 17%)	1.18 (0; 72%)	*
B. Performing Operations; Transporting	4.06 (3; 18%)	0.55 (0; 83%)	*
C. Chemistry; Metallurgy	3.16 (3; 24%)	2.41 (0; 54%)	*
D. Textiles; Paper	4.24 (4; 21%)	0.43 (0; 75%)	*
E. Fixed Constructions	4.02 (4; 19%)	1.26 b(0; 67%)	*
F. Mechanical Engineering; Lighting; Heating; Weapons; Blasting	3.27 (2 and 3; 18%)	0.19 (0; 87%)	*
G. Physics	3.20 (3; 24%)	0.17 (0; 82%)	*
H. Electricity	3.48 (4; 22%)	0.82 (0; 61%)	*
Significance of intragroup variation	*	*	
Economic sector			
Supplier-dominated sectors			
Manufacture of wood, paper, publishing, media (NACE 20, 21, 22)	4.80 (4; 29%)	0.16 (0; 86%)	*
Manufacture of rubber and plastic products (NACE 25)	5.06 (5; 21%)	0.69 (0; 71%)	*
Trade, maintenance and repair (NACE 50, 51, 52)	4.33 (4; 27%)	0.38 (0; 78%)	*
Other business activities (NACE 70, 71, 74)	4.23 (4; 23%)	1.24 (0; 59%)	*
Other supplier dominated	5.40 (4; 28%)	0.28 (0; 79%)	*

Sectors (NACE 01, 17, 18, 19, 36, 45, 63, 85)			
Production-intensive sectors			*
Manufacture of other non-metallic mineral products (NACE 26)	4.22 (3 and 4; 21%)	1.29 (0; 75%)	*
Manufacture of fabricated metal products, except machinery and equipment (NACE 28)	4.60 (4; 23%)	0.53 (0; 80%)	*
Manufacture of machinery and equipment n.e.c. (NACE 29)	5.05 (5;22%)	0.35 (0; 81%)	*
Other production intensive sectors (NACE 15, 33, 34)	4.62 (4; 19%)	1.88 (0; 62%)	*
Science-based sectors			
Manufacture of chemicals and chemical products (NACE 24)	4.69 (5 and 3; 22%)	4.47 (0; 38%)	n.s.
Manufacture of electrical machinery and equipment n.e.c. (NACE 31)	4.50 (4; 30%)	0.20 (0; 90%)	*
R&D and computer activities (NACE 72, 73)	5.47 (5; 26%)	2.00 (0; 74%)	*
Other science-based sectors (NACE 14, 23, 32, 40)	4.57 (4; 29%)	1.74 (0; 65%)	*
Significance of intragroup variation	*	*	

In an analysis focusing on patents where applicants include at least one citation, differences between examiner and applicant added citations can be discerned. In for example C. Chemistry; Metallurgy, examiners introduce fewer citations than applicants. In H. Electricity and in B. Performing Operations; Transporting, the reverse was true. Therefore, there are indications that examiners try to complement applicants' knowledge disclosure. The findings are similar for economic sectors.

4.7 FACTORS INFLUENCING CITATION PATTERNS

Geography of citations

In table 8 the origin of patent references (PRs) is presented. Of the PRs added by examiners 84% correspond to knowledge generated abroad, with the majority being based in USA, 42%. Patents for firms from Valencia that have been issued by international patent offices have a higher degree of foreign citations, i.e. 92%, with the USA accounting for 53% of these. As for national citations the majority is from the Valencia region (60%).

In terms of the geographical origin of the PRs or patents cited by applicants, table 8 shows that the proportion of foreign ones is high (71%), but not as high as for citations added by examiners. Similar to the case of examiners, the USA accounts for the largest proportion, but the distance to other regions is smaller, e.g. to regions in the European Union. National citations are more common among applicants than examiners, 27%, and in accordance with examiners citations from the region are more common than citations from other Spanish regions.

Table 8. Patent references added by examiners vs. applicants according to their geographic origin.

Geographic origin of cited patents	Nº of examiners citations of other national patents	Nº of examiners citations of other international patents	Total
Foreign	**1 811 (83%)**	**475 (92%)**	**2 286 (84%)**
European Union countries	568 (31%)	252 (53%)	830 (36%)
United States	856 (47%)	113 (24%)	969 (42%)
Japan	164 (9%)	47 (10%)	211 (9%)
China	16 (1%)	1 (0%)	17 (1%)
Rest of the world	31 (2%)	62 (13%)	93 (4%)
Spanish, with foreign applicant	176 (10%)	-	176 (8%)
Spanish	**382 (17%)**	**21 (4%)**	**403 (15%)**
From Valencian region	230 (60%)	10 (48%)	240 (60%)
Outside Valencian region	149 (39%)	11 (52%)	160 (40%)
Undetermined	**0**	**18 (4%)**	**18 (1%)**
Total	**2 193**	**514**	**2707**

Geographic origin of cited patents	Nº of applicant citations of other national patents	Nº of applicant citations of other international patents	Total
Foreign	**258 (64%)**	**131 (90%)**	**389 (71%)**
European Union countries	47 (18%)	66 (50%)	113 (29%)
United States	157 (61%)	33 (25%)	190 (49%)
Japan	7 (3%)	13 (10%)	20 (5%)
China	3 (1%)	0 (0%)	3 (1%)
Rest of the world	0 (0%)	19 (15%)	19 (5%)
Spanish, with foreign applicant	44 (17%)	-	44 (11%)
Spanish	**144 (36%)**	**6 (4%)**	**150 (27%)**
From Valencian region	74 (51%)	5 (83%)	79 (53%)
Outside Valencian region	47 (33%)	1 (17%)	48 (32%)
Undetermined	**0**	**8 (6%)**	**8 (1%)**
Total	**402**	**145**	**547**

Examiner added citations complement applicant citations since they include most citations in regions where applicants include the smallest. Comparing the three regions included in the analysis indicate that there are variations among how citation are included. Alicante, having the lowest per capita income, has more examiner citations compared to the other two regions. The explanation for this is that the technologies and sectors that include more examiner citations are more present in Alicante than other regions.

Time

Table 9 shows that the average number of examiner patent citations decreases over time. So does the frequency of mode. When looking at applicant added citations significant differences over time cannot be discerned. This may be due to harmonisation with European standards, which are characterised by fewer but more relevant citations. This has been encouraged since 1991 when the EPO gave the OEPM the responsibility for providing search reports for international patent applications.

Table 9. Patent references added by examiners vs. applicants according to their geographic origin.

Year of application	**Avg. Nº examiner citations (mode; frequency of mode)**	**Avg. Nº applicant citations (mode; frequency of mode)**	**Sig.**
1999	5.87 (5; 23%)	1.26 (0; 81%)	*
2000	4.82 (4; 25%)	1.82 (0; 72%)	*
2001	4.64 (4; 24%)	0.95 (0; 74%)	*
2002	4.71 (3 and 4; 21%)	1.05 (0: 65%)	*
2003	4.40 (4; 24%)	0.83 (0; 69%)	*

5 DISCUSSION

5.1 RESEARCH IMPLICATIONS

5.1.1 Dynamics of research collaborations

In accordance with earlier studies (Persson, Glänzel and Danell, 2004; Adams, et al., 2005; National Science Foundation, 2007) the results of my thesis point to an increase in co-authored papers. In this thesis, however, I investigated this issue further by differentiating between collaborations taking place only in Europe (intra-European) and collaborations between a European country and one or more partners from outside Europe (extra-European). This was found to apply both for intra-EU, extra-EU and global co-authorship. A comparison between intra-EU and extra-EU collaborations showed that co-authorship within Europe is more common and has also increased more than extra-EU, yet not as much as global collaborations.

European co-authored papers were more likely to be the product of collaborations at a bilateral level, involving only two countries. At the same time, multilateral co-authorships, involving more than two countries, have grown more strongly, albeit from a smaller base, than bilateral co-authorships. One possible explanation for this pattern is that the bulk of funding for international co-operation comes from national sources, which are more likely to support bilateral agreements rather than large multilateral international collaborations.

5.1.2 Country differences in co-publication

The co-publication behaviour differed between countries and was found to be associated with the size of the country, measured by the total number of publications. Similar results have been observed by earlier studies (Schubert and Braun, 1990; Narin et al., 1991; Melin and Persson, 1996; Glänzel and Schubert, 2001).

The most likely explanation for these differences is that in a big country the number of total scientists is likely to be higher than in a smaller country, which also means that the probability for a scientist to find a suitable national collaborator is higher in a large country. This is further supported by Narin et al. (1991) who concluded that the United States and United Kingdom are more reluctant to or see less need to collaborate with international partners than countries such as Italy, because of the differences in scientific "size" and number of scientists. The results also suggest that smaller countries are more likely than large countries to join bigger collaboration networks involving many countries.

I could only find a relation between co-publication behaviour and country size when collaborations took place within Europe. In extra-European collaborations, size was not found to be an influencing factor. Smaller countries collaborate more with other EU countries than bigger countries, while this difference cannot be observed in collaborations with extra-EU countries. Exceptions include UK, France and Germany, the three countries with the highest number of publications. The explanation is probably because these countries have a large national market with researchers where necessary competences and skills can be found. For possible explanations of this difference please see the policy discussion (section 5.2).

The results also show that geographic proximity plays an important role in the selection of collaboration partners, (Wagner, 2006). The results confirm that proximity has a persistent, influential role in both FPs and co-authorship. Clusters and preferential partners related to geographic proximity do still exist, both in co-authorship, as well as in FPs. Earlier studies had investigated this issue and arrived at similar conclusions. (Jaffe, 1989; Frenken, 2002; Liang et al., 2006; Frenken et al., 2007). My results with newer data sources confirm the continuing validity of these insights, despite the arrival of a new generation of information and communication technologies related to the Internet that afford unprecedented long-distance communication and collaboration tools (Wagner, 2006).

Linguistic factors can help explain some of the persistence in preferential partnerships. This was obvious in country pairs such as Germany-Austria, France-Belgium, and Netherlands-Belgium. In the latter case, a difference between the two main parts of Belgium, Wallonia (French speaking) and Flanders (Flemish speaking), could be distinguished.

Bilateral collaborations between preferential partnership countries also show some interesting characteristics with regard to corresponding author share. While an overall analysis of corresponding author shares in collaborations clearly showed differences between countries (see 5.2), these difference mostly disappear for preferential partnerships in which the two participating countries tend to split corresponding authorships fairly evenly.

One plausible explanation is that it can be assumed that these preferential collaborations are not based on new initiatives, such as the FPs, and have most likely existed for several years or decades which have resulted in long-term sustainable collaborations. In these kinds of stable collaborations it is not uncommon to develop projects jointly. The resulting publications will then alternate between mentioning the two addresses as corresponding authors reflecting the shared contribution.

5.1.3 Field differences in co-publication

Differences in co-publication patterns could also be observed across different scientific fields. In Physical, Chemical & Earth Sciences (PCES), for example, the extra-EU co-authorship rate was higher than in the other fields studied. This could be related to the expensive and large-scale equipment e.g. synchrotrons and telescopes required for research in this field. The equipment can rarely be found at the national level and researchers are therefore entirely dependent on international initiatives. Similarly, Earth Sciences can be classified as a field where international collaborations are motivated by access to geographically-specific natural resources.

Life Science (LS), including immunology, cell biology and pharmacology is also found to be very international-collaboration oriented, yet exhibits lower international collaboration rates than PCES. Life Science often requires access to and processing of large data quantities. However, according to Wagner (2005), data driven collaborations are less international-collaboration-intensive, which is in accordance with my results. Today, huge online databases, including sequences, gene expression patterns, markers and mapping data, and protein structures, are available free of charge on the Internet where researchers can upload and download information, thus in many cases reducing the need for joint, formalised collaborations to gain access to these resources.

The Engineering and Computing & Technology (ECT) field was found to be somewhat less extra-EU collaboration-intensive than both PCES and LS. Motivations for collaboration in this field can be expected to relate to the sharing of equipment and data.

My analysis also finds that Clinical Medicine (CM) was the least internationally collaborative field studied. At first sight this is surprising, since applied sciences, such as CM, are often more experimentally oriented and therefore come with a stronger need to collaborate (Price, 1963; Hagstrom, 1965; Gordon, 1980; Frame and Carpenter, 1979, Luukkonen et al., 1992). In addition, collaborations in this field can be motivated by the objective to share data related to for example statistics and clinical tests.

However, the types of collaboration most suitable and practiced for CM may rather unfold at the national than at the international level, since clinical trials, for example, can benefit from proximity and local or national collaborations between universities and hospital are often common. This is further supported by the finding that multiple affiliations, which are more likely to involve institutions in one and the same country, are most common among authors in CM related publications.

5.1.4 Byline positions in internationally co-authored papers

The continuing practice of structuring authorship for a paper in the traditional way, i.e. the first author being the junior researcher while the final author is the mentor/supervisor (Burman, 1982; Reisenberg and Lundberg, 1990; Bhopal et al., 1997; Rennie et al., 1997) could be partly supported by my analysis. The results indicated that the majority of an article's first and last authors have the same address. Moreover, in the case of the last author being the corresponding author the likelihood of first and last authors having the same address was significantly higher than when the first author appeared as corresponding. A qualitative study investigating the seniority of authors would be needed to examine this issue further.

These results combined with earlier research suggest that candidates for promotion or tenure would be well advised to highlight publications for which they acted as corresponding author, especially if they were not the first or last author. The observation that the first and last author account for the majority of correspondence suggest that some of the recent efforts to develop bibliometric formulae to measure author impact (Ball, 2005; Gomez-Alonso, 2004) would need to take into account both the number of authors on a paper and their position in the byline, in order to be more accurate.

5.1.5 Differences between examiner- and applicant-added citations

In many studies based on patent citations, the assumption is usually made that all citations were included by the inventor, and thus they are an indicator of the knowledge that he/she possesses. However, citations can be added first by the applicant, who is not necessarily the inventor, but might be the company that owns the property rights and/or second by the applicant's legal advisers or later by the patent examiners. In this thesis citations added by the applicant and by the examiner have been distinguished, with the aim of reflecting possible differences in knowledge contribution.

Only very few patents could be found with applicant added citations. This could suggest that analyses of examiner citations overestimate the weight of the explicit knowledge base of patents in regions with low absorptive capacity (weak innovation activities and a low tech profile).

Moreover, the results show that examiner citations are related to the techno-economic structure of the region. If the local industry has very few leading and patenting sectors, there will be fewer examiner citations. For other sectors, even though the knowledge base exists, examiner patent citations will not capture it. Therefore, studies on examiner citations should consider the techno-economic structure. Examiner citations may be representative of the knowledge base where there is strong industrial specialisation in highly patenting or leading sectors but require complementary analysis in every other case.

When the fact that examiners also assign IPC codes to patents (Kang et al., 2007) is taken into account, the 'noise' in citations grows exponentially. However, we also show that using applicant citations creates problems, since applicants may add large numbers of (hardly justified) citations in some patents.

Therefore the results suggest that there is a need to separate between examiner and applicant citations in research studies, since examiner added citations overestimates the importance of knowledge flows.

5.1.6 Citations as an indication of science-technology links

The result of a low number of scientific citations (NPR) indicates that using citations as an indicator of science-technology links in a region with low absorptive capacity could be questioned. The fact that patents belonging to firms quote only a few scientific publications may suggest that these links may rather be technological in nature. It may also be that universities contribute in a useful, but short-term oriented way, with low-level knowledge. The results are coherent with the fact that the Valencia region's companies are acting as technology followers rather than leaders and that patents have a low degree of novelty with the predominance of incremental over radical innovation.

5.2 POLICY IMPLICATIONS

When studying geographical knowledge flow using co-publications as an indicator it is tempting to draw the conclusion that the increased networking within Europe and the increasing number of countries participating in networks result from policy induced collaborations e.g. ERAnets and the FPs. At the same time, European researchers tend to co-author more in global (involving both European and non-European based researchers), than exclusively European multilateral networks.

A more specific comparison of these collaboration types with regard to the involvement of non-European countries provided insights into the possible influence of funding and the differences between modes of collaboration networks. In Europe, a number of "top-down" funding initiatives have been put in place in support of the European Research Area (ERA), with the aim of creating an "integrated knowledge creation network". The existing funding (but not participation in general) is restricted to European member states, accessing- and candidate countries. FPs often require involvement from several countries, sometimes preferably with new member state or candidate countries. These funding requirements are found to exert a stronger influence on resulting partnerships than the principal openness of FP supported networks to non-European partners. When examining the research partnerships in Life Science projects funded under the FPs, my analysis showed that the involvement of non-European industrialised countries with the exception of Israel has been rather limited.

In contrast, the analysis of co-authorship data that provides a broader account of the overall collaboration practices of researchers showed the opposite. Here researchers are acting in favour of globalized, rather than purely European multilateral co-authorship. These findings resonate with the argument by Wagner and Leydesdorff (2005) that growth in international collaborations is driven by the self interest of individual researchers rather than by structural, institutional or policy-related factors.

In an increasingly globalized world, with especially the BRIC countries investing heavily in research, it is important to remain open to new opportunities and not to focus all the attention on intra-European issues. Even though the EC highlighted the importance of a greater concerted S&T cooperation with the rest of the world in its vision for ERA (European Commission, 2000) and although it proposed a more proactive geographical policy to broaden access to world-wide knowledge the results in this thesis indicate that this global outlook is not very visible in the existing FP supported collaboration networks using co-publication as an indicator. Maybe it is time for existing European programmes to open-up more proactively to non-European partners and also provide more financial support for global collaborations.

Differences between countries, with smaller countries collaborating more within Europe and participating to a larger extent in networks involving many countries, could be explained by a number of factors. One could argue that Europe is used as a "national" playing field for small countries that cannot find sufficient knowledge collaboration at home, while bigger EU countries in themselves are sufficient in size. Another factor could be that it is easier for larger countries to attract foreign researchers (non-national), since these countries can offer more opportunities for foreign researchers. On the other hand, extra-EU collaborations, for which differences in collaboration patterns between countries of different sizes largely disappear, may be rather the result of size-independent factors for which linguistic, historical (e.g. former colonies) or cultural ties in form of so called preferential collaborations play a relatively more important role.

The role of countries in networks was further investigated at the European level by studying the share of corresponding authors in co-publication data. The results suggest that there is a difference in the rate of corresponding authorship between bigger and smaller countries. Smaller countries are found to have a relatively higher share of corresponding authorship in bilateral collaborations as compared to multilateral ones. This situation is reversed for larger countries.

A possible explanation is the following: in bigger, S&T-advanced countries the required competences and experience to manage large, multi-partner collaborations is more likely to exist than in smaller countries. These large projects also come with a considerable administrative burden, which in itself is costly and requires experience that can only be acquired from similar projects. It is not "unfair" to claim that bigger more S&T-advanced countries have a competitive advantage and that this would explain why larger countries more often appear as corresponding author or coordinator in FP projects.

Different national scientific cultures also seemed to influence the corresponding author position. A call for a shift in national publication culture towards common international standards could make research contributions better visible, attributable and comparable. In addition, better harmonization would also support the development of datasets and methods that can be used for tracking and analysing scientific publication practices.

The differences between collaboration patterns and drivers across scientific fields suggest that policies promoting collaborations need to be tailored to the drivers for collaboration that apply to the specific field. Where research requires expensive and large-scale equipment, such as in PCES, policies should prioritise support for equipment/resource-sharing. Collaborations motivated by data-sharing may not require the same level of policy attention, since they are nowadays mostly facilitated by the Internet and the situation is similar for collaborations that revolve around the sharing of ideas. However, policy initiatives aimed at facilitating geographical knowledge flow through the mobility of researchers do still play an important role to support all types of collaborations.

The traditional use of patent citations to trace knowledge flows and science-technology links has many flaws in regions with low absorptive capacity, as this thesis has highlighted. Therefore it is important to take these factors into account when using citations for justifying research funding and evaluating performance.

The finding that the majority of citations are to international patents suggests that some policies to support innovation in regions with a low absorptive capacity are less successful, e.g. supporting regional R&D through subsidies.

Likewise the relatively low frequency of NPRs suggest that the knowledge base within firms is low and that therefore policies aimed at strengthening local university-industry should be more prominently considered.

Increasing resource allocation to patent examiners would allow for more detailed search reports, which would most likely result in increased incentives for applicants to disclose more information in patent applications. More resources for patent examinations could also translate into better standards for storing information and thus make it possible to analyse full-text applicant citations a significant benefit for examining knowledge transfer.

Overall, many of my findings suggest a very differentiated, context-specific picture of knowledge transfer and the factors that drive them. Combined with the fact that different countries and regions have a specialisation in different fields and technologies this may suggest that instead of trying to make the European research area fully integrated by design, related programmes could be more accommodating to self-selected collaborations, so called “bottom-up” initiatives, and the strategic leverage of existing differences, experiences, synergies, complementarities and competitive advantages that they afford.

5.3 FURTHER WORK

This thesis contributed to the understanding of knowledge transfer mainly by studying quantitative parameters such as input data in the form of funding networks and output data in the form of co-publication patterns and patent citations. To be able to further understand the impact of funding instruments and the evolution of European research networks several other indicators need to be taken into account such as the role of European funding in the national context and expected and added value of European initiatives e.g. FPs and other instruments supporting collaborations. This could be done by more qualitative studies which also take the national context and situation into account. All along the way it is important to keep in mind that European research policy has had different impacts on different countries.

This thesis has mainly contributed to the macro-level analysis of European research collaborations by studying European countries. At the end of the day, collaborations are made up of individual researchers with different behaviours that cannot always be aggregated at country level or generalised for policy purposes. It would therefore be interesting to carry out more case studies and stakeholder interviews to study the impact of policies on individual researchers, their collaboration decisions and output.

Another interesting area for future research relates to issues around the concept of “star scientists”. Previous research has suggested that better renowned researchers tend to get more visibility and are in a better position to access research funding than less renowned colleagues for new research they conduct. This has been referred to the so called “Matthew effect”. It is not clear to what degree easier access to cumulative funding actually continues to enhance scientific output and quality or whether specific thresholds can be discerned. A better understanding of these issues might provide critical input for designing future research funding policies.

In parallel to the concept of “star scientists” the same research question could be asked for “star regions” or “star clusters”. The underlying policy question would then be whether to fund in these “star regions” rather than less developed regions if the aim is to increase the innovation performance.

6 SUMMARY AND CONCLUSIONS

The aim of this thesis was to study knowledge transfer in Europe and how these can be measured and analysed through indicators focusing on geographical and sectoral knowledge flow. Knowledge transfer plays an important role in making Europe the "world's most dynamic and competitive knowledge economy" (European Commission, 2000) and to overcome weaknesses such a weak environment to stimulate high quality research and exploit results.

Overall, it can be concluded that European research collaborations have become more international over the last decades, drawing on competences from different national sources. Looking at patterns of co-authorship, it is clear, that researchers for various reasons seek access to global knowledge and global networks rather than exclusively European partnerships and that this global multilateral orientation in co-publications continues to increase.

Smaller countries in Europe are more prone to collaborating internationally than bigger ones, as expressed by their share of international co-publications in relation to their total country publication output.

However, at closer inspection differences related to the size of countries could only be found for collaborations that take place within Europe, but were not an influencing factor when collaborations were carried out with countries other than the EU member states.

A plausible explanation for these differences relates to the fact that when the required competence cannot be found at home, research groups in small countries seem to use Europe as a "national" playing field, while bigger EU countries in themselves are often large enough to provide sufficient domestic opportunities for collaboration. This explanation was also supported by the finding that bigger countries overall have a higher share of exclusively national publications than smaller countries.

The situation is different for extra-EU collaborations however, which seem to address different needs. Linguistic, historical or cultural ties in form of so called preferential collaborations seem to be more influential factors in this context.

The understanding of the relationship between funding mechanisms and dynamics of research collaborations was further advanced by the results showing that "top-down" funding disbursed through the EU Framework Programmes were much more exclusively European in nature. This pattern exists despite a general (yet not financial) encouragement from the European Commission to collaborate with non-European partners.

In marked contrast, evidence from research outputs as measured by co-publication patterns in multilateral co-authorships suggests that European researchers tend to co-author more with global (involving both European and non-European countries), rather than exclusively European partners and that this global multilateral orientation in co-publications is further on the rise. These results shed some important light on the actual influence (and the limits thereof) of different funding policies that seek to nurture more integrated, diverse and inclusive research partnerships.

To investigate the patterns and relevance of author position in internationally co-authored papers with regard to their relative contribution to the research effort, the corresponding author position in the article was studied. The results show that the first and last author account for the majority of correspondence and therefore can be expected to have assumed a more prominent role in the research effort.

The analysis also partly supports the claim that the traditional way of structuring author positions for a paper – junior researcher as first author, his or her supervisor as final author - continues to be practiced. First, it was found that the majority of an article's first and last authors have the same address. Second, in the case of last author being the corresponding author, the likelihood of first and last authors having the same address was significantly higher than when the first author appeared as the corresponding author.

In addition, the analysis indicates that different national scientific cultures seemed to influence the corresponding author position. This means that a call for a shift in national publication culture and author position practices towards common international standards could make research contributions more visible, attributable and comparable.

All these insights suggest that some of the recent efforts to develop bibliometric formulae to measure author impact would need to take into account not only the number of authors on a paper, but also their position in the byline in order to be more accurate.

A closer look at the use of patentometric indicators has yielded a similar picture. In this thesis patent citations were considered in more detail and distinguished between citations added by the applicant vs. the ones added by the examiner with the aim of reflecting possible differences in knowledge contribution.

A relatively low number of patents with applicant added citations could be identified. This could suggest analyses of total patent citations overestimate the weight of the explicit knowledge base of patents in regions with low absorptive capacity (weak innovation activities and a low tech profile).

Similarly, the finding of a low number of scientific citations in patents for regions with low absorptive capacity also suggests that patent citations as an indicator of science-technology links needs to be approached with extra caution in these contexts.

It is important to take these qualifications into account when using citations for justifying research funding and evaluating performance.

In summary, this thesis has provided several pieces of evidence that advance the understanding of knowledge transfer and how related indicators can be used and developed. It strongly demonstrates that European research policy has had an impact on research collaboration patterns. However, the evidence has also shown that this impact has its limits and does not over-ride self-selected collaboration patterns that continue to drive a more global rather than exclusively European orientation of international research collaborations by European countries. A more disaggregated scrutiny of publication patterns has also underscored very clearly that collaboration strategies show considerable diversity across scientific fields as well as countries, thus emphasising the importance of tailoring and being context-specific when designing supporting research policies. Finally, the analysis has also made a contribution at methodological level, arguing for the need of fine-tuning bibliometric and patentometric tools, in order to more effectively capture the relative importance of author contributions. As for the use of patent citations, the results suggest that it would be premature to defend the use of patent citations as indicators of technological impact and knowledge flows especially in regions with low absorptive capacity (weak innovation activities and a low tech profile).

This thesis highlight the importance that policymakers should be better informed about how indicators can be used as a tool when designing policies. Many policymakers today have a limited insight into the process of knowledge flows and its socio-economic consequences and consider it as a linear, automatic and geographically localized process.

Knowledge flows are essential and are even likely to play a more important role in the future for tackling the global challenges that the world faces in the 21st century. Understanding the dynamics of knowledge transfers and how these can be measured and analysed will be crucial for developing policy avenues for tackling future socio-economic challenges. My hope is that this thesis has made a contribution to this effort.

7 POPULÄRVETENSKAPLIG SAMMANFATTNING

Kunskapsöverföring består av processer som syftar till att fånga och förmedla kunskap och kompetens som kan omvandlas till nytta och ekonomisk tillväxt. Inom ramen för Lissabonstrategin 2000, med målsättningen att göra EU till ”världens mest dynamiska och konkurrenskraftiga kunskapsbaserade ekonomi” spelar kunskapsöverföring en viktig roll. Införandet av nya finansieringssystem och åtgärder för att öka kunskapsflödet mellan länder och sektorer i Europa har ökat efterfrågan på effektstudier av policies och utveckling av relevanta och exakta indikatorer. Syftet med denna avhandling var att undersöka hur indikatorer kan användas i beslutsfattande genom att studera kunskapsöverföring inom Europa med fokus på geografiska och sektoriella kunskapsflöden. Detta gjordes genom att studera sampublikationer och samarbeten inom EU:s ramprogram. Sektoriella kunskapsflöden analyserades med hjälp av patentciteringar.

Resultaten visade att mindre länder samförfattade i större utsträckning med andra EU-länder än vad större länder gjorde. Samförfattande med länder utanför EU var inte beroende av landets storlek. Resultaten visade även att samförfattande skiljer sig åt mellan olika vetenskapliga områden. När det gäller samarbeten i de projekt som finansierats av EU:s ramprogram var dessa av en mer exklusiv europeisk karaktär medans samarbeten illustrerade av sampubliceringar visade att europeiska forskare tenderar att sampublicera mer globalt, snarare än med enbart europeiskt baserade forskare och att denna globalisering fortsätter att öka. Resultat gällande en artikels korresponderande författare (corresponding author) visade att denna oftast är placerad först i författarordningen och näst ofta som siste författare. Resultaten var beroende på antalet författare som förekom i artikeln och nationella skillnader förekom. Slutligen visar resultaten att patentciteringar som en indikator för samverkan mellan näringsliv och forskning och geografisk lokalisering inte är lika relevant i en region med låg absorptionsförmåga som kännetecknas av låg innovationsförmågan och lågteknologisk industri som i en högteknologisk region.

Att använda sig av sampublikationer som en indikator för geografiska kunskapsflöden visar att den europeiska forskningspolitiken har påverkat Europeiska forskningssamarbeten. Dessa effekter har dock troligen sina begränsningar och råder inte över de självvalda samarbeten som fortsätter att utvecklas mot ett mer globalt snarare än enbart ett europeiskt samarbetsmönster. Resultaten visar även att publikationsmönster skiljer sig åt mellan olika vetenskapsområden och länder. Vidare tyder resultaten på att vissa policies, som ämnar stödja innovation i regioner med låg absorptionsförmåga t.ex. genom subventioner, kan vara mindre effektiva än t.ex. anställning av kvalificerad personal vid företag eller ökningen av samarbete mellan akademi och näringsliv. Slutligen bidrar resultaten till en metodologisk diskussion, där finjustering av bibliometriska verktyg behövs för att på ett mer effektivt sätt ta hänsyn till den relativa betydelsen av författares bidrag i en artikel. När det gäller användningen av patentciteringar, tyder resultaten på att patentciteringar som indikatorer för teknologisk utveckling och kunskapsflöden särskilt i regioner med låg absorptionsförmåga bör användas med försiktighet.

8 ACKNOWLEDGEMENTS

This thesis and the more than five years of work that have gone into it would not have been possible without the invaluable support of my supervisors, colleagues, family and friends. I would in particular like to thank:

My supervisors:
Carl Johan Sundberg, my main supervisor, for all the support you have given me over the years. For your energy, your engagement and those questions that make me think once more.

Patrice Laget, my co-supervisor, for your enthusiasm and curiosity and for inspiring me to get involved in policy issues.

Anna Nilsson Vindefjärd, my co-supervisor, for the support and belief in me.

Colleagues
My former colleagues at IPTS with a special thank you to Joaquin Azagra-Caro.

My colleagues at Faugert & Co and all my other colleagues at Technopolis.

My UBE colleagues at Karolinska Institutet with at very special thanks to Charlotta Dahlborg and Danielle Lewensohn.

Friends and family
All my friends over the world, thanks for always being there and showing me that life is more than work.

Mamma tack för att du alltid finns där och för ditt intresse att alltid lära dig mer.

Pappa nu är jag äntligen färdig.

Jeanette och Josefine mina älskade systrar som bidragit med skratt, mat och husrum.

Farmor och Farfar tack för att ni och Raggarön finns.

Dieter, without you these words would never have been written! You are the best supporter and friend. I love you!

REFERENCES

Adam, D., (2002). The counting house. *Nature,* 415:726-9.

Adams, J. D., Black, G. C., Clemmons, J. R., Stephan, P. E., (2005). Scientific teams and institutional collaborations: Evidence from US universities, 1981-1999. *Research Policy,* 34: 259-285.

Alcácer, J., Gittelman, M., (2006). Patent citations as a measurement of knowledge flows: The influence of examiner citations. *Review of Economics and Statistics*, 88: 774–779.

Allison, J.R., Sager, T.W., (2007). Valuable Patents Redux: On the Enduring Merit of Using Patent Characteristics to Identify Valuable Patents. *Texas Law Review*, 85:1769-1797.

Archibugi, D., Coco, A., (2004). International partnerships for knowledge in business and academia - A comparison between Europe and the USA. *Technovation,* 24: 517-528.

Argote, L., Ingram, P., (2000). Knowledge Transfer in Organizations: Learning from the Experience of Others. *Organizational Behavior and Human Decision Processes*, 82: 1–8.

Azagra-Caro, J.M., Fernández de Lucio, I., Perruchas, F., Mattsson, P., (2009). What do patent examiner inserted citations indicate for a region with low absorptive capacity? *Scientometrics*, 80: 441–455.

Ball, P., (2005). Index aims for fair ranking of scientists. *Nature*, 436: 900.

Bammer, G., (2008). Enhancing research collaborations: Three key management challenges. *Research Policy,* 37: 875-887.

Banchoff, T., (2002). Institutions, Inertia and European Union Research Policy. *Journal of Common Market Studies*, 40:1-21.

Barrett, P. 2004. *Science and Theology Since Copernicus: The Search for Understanding.* Continuum International Publishing Group. Pretoria.

Beaver, D.D., Rosen, R., (1978). Studies in Scientific Collaboration. Part I. The professional origins of scientific co-authorship, *Scientometrics*, 1:65-84.

Beaver, D. D., (2001). Reflections on scientific collaboration, (and its study): past, present, and future. *Scientometrics,* 52: 365-377.

Bhopal, R., Rankin, J., McColl, E., Thomas, L., Kaner, E., Stacy, R., et al., (1997). The vexed question of authorship: views of researchers in a British medical faculty. *British Medical Journal*, 314:1009-12.

Blind, K., Edler, J., Friedewald, M., (2005). *Software Patents: Empirical Evidence and Policy Implications*. Edward Elgar, Cheltenham/UK, and Northampton/USA.

Bozeman, B., and Boardman, C., (2003). *Managing the New Multipurpose, Multidiscipline University Research Centers: Institutional Innovation in the Academic Community*. IBM Center for the Business of Government. Washington, D.C.

Breschi, S., Cusmano, L., (2004). Unveiling the texture of a European research area: Emergence of oligarchic networks under the EU Framework Programmes. *International Journal of Technology Management*, 27:747–772.

Burman, K.D., (1982). Hanging from the masterhead – reflection on authorship. *Annals of Internal Medicine,* 97:602–5.

Callaert, J., van Looy, B., Verbeek, A., Debackere, K., Thus, B., (2006). Traces of Prior Art: An analysis of non-patent references found in patent documents. *Scientometrics*, 69: 3-20.

Calvert, J., Martin, B.R., (2001). *Changing Conceptions of Basic Research?* Background Report prepared for the OECD Workshop on Policy Relevance and Measurement of Basic Research, Oslo 29-30 October 2001.

Cohen, W.M, Nelson, R.R, Walsh, J.P., (2002). Links and impacts: The influence of public research on industrial R&D. *Management Science*, 48: 1-23.

CREST Working Group., (2007). *Analytical Report Policy Approaches towards S&T Cooperation with Third Countries*. Brussels.

Criscuolo, P., Verspagen, B., (2008). Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. *Research Policy*, 37: 1892-1908.

Cullen K., (2008). Scottish universities in the marketplace. Pp. 89-101 in Engwall, L. and Weaire, D. (eds*.): The university in the market.* London: Portland Press.

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

Engel, A., and Ruschenburg, T., (2008). The uneven spread of global science: patterns of international collaboration in global environmental change research. *Science and Public Policy*, 35: 347-360.

European Commission, (2009). *Metrics for Knowledge Transfer from Public Research Organisations in Europe.* EUR 23894.

European Commission, (2000). *Towards a European Research Area, Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions*. Brussels, COM (2000) 6.

Evidence Ltd., (2007). *The use of bibliometrics to measure research quality in UK higher education institutions*. Universities UK.

Frame, J.D., Carpenter, M.P., (1979). International research collaboration. *Social Studies of* Science, 9:481–497.

Frenken, K., (2002). A new indicator of European integration and an application to collaboration in scientific research. *Economic Systems Research*, 14:345–361.

Frenken, K., Hoekman, J., van Oort, F., (2007). *Towards a European Research Area*. NAi Publishers, Rotterdam.

Garfield, E., (1972). Citation as a tool in journal evaluation. *Science*, 178:471-479.

Geisler, E., (1995). Industry-University technology Cooperation: A Theory of Inter-Organizationsal Relationships. *Technology Analysis & Strategic Management*, 7:217-229.

Georghiou, L., (1998). Global cooperation in research. *Research Policy*, 27:611-626.

Glänzel, W., (2001). National characteristics in international scientific co-authorship relations. *Scientometrics* 51:69-115.

Glänzel, W. Schubert, A., (2001). Double effort = Double impact? A critical view at international co-authorship in chemistry. *Scientometrics,* 50:199-214.

Gomez-Alonso, J., (2004). Author! Author! *Journal of the American Medical Association*,, 292, 1815.

Gordon, M.D., (1980). A critical reassessment of inferred relations between multiple authorship, scientific collaboration, the production of papers and their acceptance for publication. Scientometrics, 2:193–201.

Griliches, Z., (1990). Patent statistics as economic indicators. *Journal of Economic Literature*, 28: 1661-1707.

Hall, B., (2004). http://www.oecd.org/dataoecd/45/23/33835392.pdf.
Harhoff, D., Narin, F., Scherer, F. M., Vopel, K., (1999). Citation frequency and the value of patented inventions. *Review of Economics and Statistics*, 81: 511–515.
Hirsch, J.E., (2005). An index to quantify an individual's scientific research output. *Proceeding of National Academy of Science USA*, 102:16569-16572.
ICMJE., (1997). Uniform Requirements for Manuscripts Submitted to Biomedical Journals. Journal of the American Medical Association, 277:927–34.
Jaffe, A., Trajtenberg, M., Fogarty, M.S., (2000). The meaning of patent citation: Report on the NBER/Case-Western reserve survey of patentees. NBER Working Paper, 7631.
Jaffe A, Trajtenberg M, Henderson R., (1993). Geographic localization of knowledge spillovers as evidenced by patents citations. *Quarterly Journal of Economics,* 63: 577-98.
Jaffe, A.B., (1989). Real effects of academic research. *American Economic Review,* 79:957-970.
Kang, I.S, Na, S.H., Kim, J., Lee, J.H., (2007). Cluster-based patent retrieval. *Information Processing & Management*, 43: 1173-1182.
Katz, J. S., Martin, B.R., (1997). What is research collaboration? *Research Policy*, 26:1–18.
Katz, J. S., (1994). Geographical Proximity and Scientific Collaboration. *Scientometrics,* 31: 31-43.
Hagstrom, W.O., (1965). *The Scientific Community*. Basic Books. New York.
Lewison, G., Cunningham, P., (1991), Bibliometric studies for the evaluation of trans-national research. Scientometrics, 21 : 223–244.
Leydesdorff, L., (2008). Patent Classifications as Indicators of Intellectual Organization. *Journal of the American Society for Information Science and Technology*, 59: 1582-1597.
Leydesdorff, L., Meyer, M., (2003), The Triple Helix of university-industry-government relations. Scientometrics, 58: 191–203.
Leydesdorff, L., (2000). Is the European Union becoming a single publication system? Scientometrics, 47:265-280.
Liang, L.M., Zhang, L., Kretschmer, H., Scharnhorst, A., (2006). *Geographical and lingual preferences in scientific collaboration of the European Union (1994–2003)*. In Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting Nancy (France).
Luukkonen, T., (2002). Technology and market orientation in company participation in the EU Framework programme. *Research Policy,* 31:437-455.
Luukkonen, T., Tijssen, R.J.W., Persson, O.,(1993). The measurement of international scientific collaboration. *Scientometrics,* 28:15-36.
Luukkonen, T., Persson, O., Sivertsen, G., (1992). Understanding patterns of international scientific collaboration. *Science Technology & Human Values*, 17:*101–126.*
Lundberg, J., (2007). Lifting the crown-citation z-score*, Journal of Informetrics*, 1:145-154.
Mattsson, P, Laget, P., Nilsson, A., Sundberg, C.J., (2008). Intra-EU vs. extra-EU scientific co-publication patterns in EU. *Scientometrics,* 75:555-574.
Melin, G., Persson, O., (1996). Studying research collaboration using co-authorships. *Scientometrics,* 36:363-377.

Metcalf, J.S., and Georghiou, L., (1998). *Equilibrium and Evolutionary Foundations of Technology Policy*. OECD STI Review 22, 75-100.
Meyer, M. (2002), Tracing knowledge flows in innovation systems. *Scientometrics*, 54: 193–212.
Meyer, M., (2000). What is special about patent citations? Differences between scientific and patent citations. Scientometrics, 49: 93.
Michel, J., Bettels, B., (2001). Patent citations analysis: A closer look at the basic input data from patent search reports. *Scientometrics*, 51: 185.
Moed, H.F., (2002). The impact-factors debate: the ISI's uses and limits. Nature, 415:731-732.
Moed, H.F., Glänzel, W., Schmoch, U., (2004). Handbook of quantitative science and technology research: The Use of Publication and Patent Statistics in Studies of S&T Systems. Kluwer Academic Publishers. Dordrecht.
Moed, H.F., (2009). UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics,* 74:153-61.
Molas Gallart, J., (2001). Government defence research establishments: the uncertain outcome of institutional change. *Defence and Peace Economics* 12: 417 – 437.
Narin, F., Hamilton, K. S., Olivastro, D., (1997), The increasing linkage between U.S. technology and public science. *Research Policy*, 26: 317–30.
Narin, F., (1994). Patent bibliometrics. *Scientometrics*, 30: 147-155.
Narin, F., Stevens, K., Whitlow, E.S., (1991). Scientific Cooperation in Europe and the Citation of Multinationally Authored Papers. *Scientometrics,* 21:313-323.
Narin, F., Noma, E., (1985). Is technology becoming science? *Scientometrics*, 7: 369-381.
National Science Foundation, (2007). *Asia's Rising Science and Technology Strength: Comparative Indicators for Asia, the European Union, and the United States*. NSF 07-319. Arlington, VA.
NCDDR, (2005). *What are the Standards for Quality Research?* Technical Brief 9. Southwest Educational Development Laboratory.
Noma, E., Olivastro, D., (1984). Are there enduring patents? *Journal of the American Society for Information Science and Technology*, 36: 297 – 301.
OECD, (2002). The measurements of Scientific and Technological Activities Frescati Manual 2002. Paris Cedex.
Okubo, Y., Zitt, M., (2004). Searching for research integration across Europe: a closer look at international and inter-regional collaboration in France. *Science and Public Policy,* 31:213-226.
Persson, O., Glänzel, W., Danell, R., (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60:421–432.
Pohoryles, R., (2002). The Making of the European Research Area – a View from Research Networks. *Innovation: The European Journal of Social Science Research,* 15:325-340.
Polanyi, M., (1999). *The Tacit dimension*. London:Routledge & Kegan Paul.
Price, D.J.D., (1963), *Little Science, Big Science*. Columbia University Press, New York.
Pritchard, A., (1969). Statistical bibliography or bibliometrics? *Journal of Documentation,* 25:348-349.

Rehn, C., Kronman, U., (2008). *Bibliometric handbook for Karolinska Institutet.* Karolinska Institutet University Library. Stockholm.

Reisenberg, D., Lundberg G.D., (1990). The Order of Authorship: Who's on First? *Journal of the American Medical Association*, 264:1857.

Rennie, D., Yank, V., Emanuel, L., (1997). When Authorship Fails – A Proposal to Make Contributors Accountable. *Journal of the American Medical Association,* 278:579-85.

Rigby, J. Edler, J., (2005). Peering inside research networks: Some observations on the effect of the intensity of collaboration on the variability of research quality. *Research Policy,* 34:784-794.

Roediger-Schulga, T., Barber, M.J., (2007). *R&D collaboration networks in the European Framework Programmes: Data processing, network construction and selected results*. United Nation University. Maastricht.

Scherngell, T., Barber, M.J., (2009). Spatial interaction medelling of cross-region R&D collaborations: empirical evidence from the 5th EU Framework programme. *Papers* in Regional Science, 88:531-546.

Schmoch, U., (1993). Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics*, 26: 193–211.

Schubert, A., Braun, T., (1990), International Collaboration in the Sciences, 1981-1985. *Scientometrics,* 19:3-10.

Serres, M., Latour, B., (1995). *Conversations on science, culture and time*. The University of Michigan Press.

Smedby, J.C., Trondal, J., (2005). Globalisation or europeanisation? International contact among university staff. *Higher Education*, 49:449-466.

Smith, M., (1958). The Trend toward Multiple Authorship in Psychology. *American Psychologist,* 13:596-599.

Sörlin, S., (2004). *Europas idehistoria. 1492-1918*. Natur och Kultur. Stockholm.

Technopolis, (2010). *BBMRI: an evaluation strategy for socio-economic impact assessment*.

Tijssen, R.J.W., (2001). Global and domestic utilization of industrial relevant science: patent citation analysis of science–technology interactions and knowledge flows. *Research Policy*, 30: 35–54.

Litman, T., (2010). Evaluating Research Quality. Victoria Transport Policy Institute.

UNESCO, (2010). Result-Based Programming, Management and Monitoring (RBM) *approach as applied at UNESCO*.

van Leeuwen, T.N., Moed, H.F., (2002). Development and application of journal impact measures in the Dutch science system. *Scientometrics,* 53: 249-266.

Wagner, C.S, Leydesdorff, L., (2005). Network Structure, Self-Organization and the Growth of International Collaboration in Science. *Research Policy,* 34:1608-1618.

Wagner, C.S., (2006). International collaboration in science and technology: promises and pitfalls, in: Box, L., Engelhard, R. (Eds.), *Science and Technology Policy for Development*, Dialogues at the Interface. Anthem Press, London, pp. 165-177.

Wagner, C.S., (2002). The elusive partnership: science and foreign policy. *Science and Public Policy*, 29: 409-417.

Wallin, J.A., (2005). Bibliometric Methods: Pitfalls and Possibilities. *Basic & Clinical Pharmacology & Toxicology*, 97:261-275.

Wikepedia. http://en.wikipedia.org/wiki/Peer_review. Accessed 2011-01-20.

Zitt, M., Teixeira, N., (1996). Science macro-indicators: Some aspects of OST experience. *Scientometrics,* 35:209-222.

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.

Zitt, M., Bassecoulard, E., (2005). Internationalisation in science in the prism of bibliometric indicators, in: Moed, H., Glänzel, W., Schmoch, U., (Eds.), *Handbook of Quantitative science and technology research*. Kluwer Academic Publishers. Dodrecht.

I

Jointly published by Akadémiai Kiadó, Budapest
and Springer, Dordrecht

Scientometrics, Vol. 75, No. 3 (2008) 555–574
DOI: 10.1007/s11192-007-1793-x

Intra-EU vs. extra-EU scientific co-publication patterns in EU

PAULINE MATTSSON,[a,b] PATRICE LAGET,[a] ANNA NILSSON,[b,c]
CARL-JOHAN SUNDBERG[b]

[a] *IPTS-JRC, Edificio Expo, C/Inca Garcilaso, Sevilla (Spain)*
[b] *Unit for Bioentrepreneurship, LIME, Karolinska Institutet, Stockholm (Sweden)*
[c] *ITPS, Embassy of Sweden, Washington DC (USA)*

The increase of co-authored papers is a recognized fact. At the same time the factors influencing this change is not well known. This article aims at studying the patterns of EU science co-authorships. We analyzed articles published in 18 EU countries and their intra-EU (within EU) and extra-EU (with partners outside EU) co-publication pattern in five scientific fields. The results point to a Europeanization of shared co-authorship rather than an internationalization outside Europe. Smaller countries co-authored more with other EU countries than bigger countries while the co-authorship rate with extra-EU partners was not dependent of the country's size. The co-authorship patterns were also found to depend on the scientific field. Engineering and Computing & Technology was the field with the highest level of national publications and Physical, Chemical & Earth Sciences the field with the highest level of both intra-EU and extra-EU collaborations. Furthermore, our data shows that specialization in a field does not impact the co-authorship rate neither in intra- nor in extra-EU co-authorship. These results support the view that a single market for research is developing within the EU with a seamless extension of national systems into other Member States ones.

Introduction

International collaborations among researchers have increased significantly during the last decade and are assumed to have an increased impact on the quality of the

Received July 30, 2007

Address for correspondence:
PAULINE MATTSSON
Unit for Bioentrepreneurship, LIME, Karolinska Institutet
Berzelius väg 3, 171 77 Stockholm, Sweden
E-mail: pauline.mattsson@cmi.ki.se

0138–9130/US $ 20.00

outcome [LUUKKONEN & AL., 1992; GEORGHIOU, 1998; GLÄNZEL, 2001]. One of the first studies confirming this trend was conducted by SMITH [1958] who observed an increased number of co-authored publications and started to use multiple-author papers as an index for collaborations. According to ADAMS & AL. [2005] national collaborations between universities or universities and firms in USA have doubled between 1981 and 1999, while international collaborations over the same period have increased five-fold.[1] A regression analysis, from the same study, explains the increase with the rise of public R&D investment, private control of universities, and increased mobility of PhDs. The work also reveals that the number of authors involved in each publication has doubled and in some countries even tripled in the same period. ARCHIBUGI & COCO [2004] observed the same trend over a ten-year period. The article concludes that the increased international collaboration pattern is more obvious in EU countries than in the USA. Japan and the average EU country have increased its international collaborations three-fold while in USA a two-fold increase could be observed. The study also reveals that of all EU internationally co-authored papers, the share of intra-EU collaborations have increased (1986–1988 to 1995–1997) while the share of USA collaborations have decreased, both for individual member states as for the entire EU.

Even though the development of easier communication and travel have made international collaborations simpler, the main factor underlying the growth of international collaborations is the increase of international funding programs [LUUKKONEN & AL., 1993]. This could suggest that the last few decades of EU efforts to increase collaborations within EU have succeeded. Examples of such programs include the different framework programs, initiated by the EU Commission. Theses programmes have had a focus on strengthen the socio-economic competitiveness of Europe with the main objectives to involve and integrate the least developed countries in EU and to increase mobility between countries and sectors. Our hypothesis is that the collaboration pattern among EU member states has changed over the last years. Earlier studies show that countries have an increasing collaboration rate with other countries and suggest a growing internationalisation of science. It is not clear whether these changes have taken place equally among member states or if some countries behave differently. In contrast to earlier studies this article first treats EU as a homogenous group in order to make global comparisons. Secondly we study each of the countries collaborations separately differentiating between EU-member states, candidate countries and other European countries (Norway and Switzerland).

This study investigates the collaboration pattern of EU member states using articles published by researchers in 18 EU countries, accounting for 99% of the total EU25 production, candidate countries and other European countries (see Table 1), and their co-authorship, within EU (intra-EU), and with global partners outside the EU (extra-

[1] Taking into account that national collaborations are more frequent than international.

EU). Further, we investigated more specifically whether the networking differed between scientific fields, countries and modes of collaborations. The research questions asked in this study included:

1. To what extent do researchers in EU countries collaborate within EU and outside?
2. Do specific country characteristics determine a country's collaboration pattern?
3. Does the collaboration pattern differ between scientific fields?

Table 1: Classification of studied countries

Intra-EU countries	Abbreviation
Austria	AT
Belgium	BE
Czech Republic	CZ
Denmark	DK
Finland	FI
France	FR
Germany	DE
Greece	GR
Hungary	HU
Ireland	IE
Italy	IT
Netherlands	NE
Poland	PL
Portugal	PT
Slovakia	SK
Spain	SP
Sweden	SE
UK	UK
Candidate countries	
Bulgaria	BG
Turkey	TR
Other European countries	
Norway	NO
Switzerland	CH

Literature review

Why do researchers choose to collaborate internationally?

Motivations to collaborate internationally are often the same as the overall incentives to collaborate. BEAVER [2001] lists eighteen purposes for which collaborations are taking place. Also KATZ & MARTIN [1997] include a list of proposed factors contributing to collaborations. The factors can be divided into four categories such as 1) financial reasons including better access to funding, sharing of core-facilities that can not be purchased by the individual laboratory, 2) social factors such as acknowledgement from the scientific community, increased networking effect (getting to know more people in the scientific community), the human social factors (we prefer to work in a group rather than individually), 3) knowledge collaborations including education of students, improved technical, analytical, and theoretical knowledge and 4) political factors, including Framework Programmes and other policy based initiatives. Motivations to collaborate internationally include all the above mentioned factors and as GEORGHIOU [1998] puts it "…the wider geographical coverage of a program, the greater the chance becomes of finding exactly the right partner". ANDERSSON & PERSSON [1993] highlight the potentials of synergy of ideas and reduced experimental costs due to shared core facilities as the two driving factors for increased international collaboration.

The benefits of collaborations have been observed in a vast amount of literature. An often used quantitative indicator measuring the qualitative importance of a publication is the citation rate. Earlier works suggest that collaborative publications are more cited than articles with only one author [LEWISON & CUNNINGHAM, 1991; RIGBY & EDLER, 2005]. GLÄNZEL & SCHUBERT [2001] and PERSSON & AL. [2004] studied the subject more specifically and found that internationally co-authored publications have a higher citation rate than domestic collaborations. Others have looked at the core journals, of a specific field and found that the number of co-authored papers is higher than in other journals [BEAVER & ROSEN, 1979; GORDON, 1980]. The activity, in relation to the number of publications, of the individual author has also been shown to increase when collaborations are carried out [PERSSON & AL., 2004]. One has to remember that these measurements are correlated to the scientific fields studied and statistics differ between fields.

International collaborations bring several benefits but there are also some drawbacks. KATZ & MARTIN [1997] discuss the increased financial costs, increased administration costs, and increased travel time being consequences from collaborations. GEORGHIOU [1998] also mentions the differences between research priorities and policy support as a geographical impediment. Different countries have different policy supports and priorities; this can lead to confusion in the exploitation of research results regardless if the research is based on comparable fundamental research questions. Similar differences can also result in uncertainty of institutional responsibility within a joint project involving many different researchers from different countries. In the end a research team has to weigh the potential benefits against the possible costs of international collaborations.

Scientific field differences

The level of collaboration may also differ between scientific fields. Differences between basic and applied research have already been studied [FRAME & CARPENTER, 1979; LUUKKONEN & AL., 1992]. Applied sciences is often more experimentally oriented and therefore the need to collaborate is also stronger [PRICE, 1963; HAGSTROM, 1965; GORDON, 1980]. The growing use of expensive and large-scale instruments could be a cause of increased collaboration. When a research group does not have the required budget needed to purchase instruments it has to find someone to share the costs with. The interdisciplinarity of a field is also a factor which could affect a research group's tendency to collaborate. If the required knowledge cannot be found within the own research group this has to be attained from external collaborations.

WAGNER [2005] categorized the scientific research fields according to four motivation factors influencing international collaborations; 1) data driven, 2) resource driven, 3) equipment driven, and 4) idea/theory driven. 1) Data driven collaborations are defined as the incentives to share data with relevant partners, 2) resource driven collaborations are often associated with unique and rare resources, 3) equipment driven collaborations imply cost-sharing, maintenance, and access to, not necessarily sharing of, large-scale equipment and 4) idea/theory driven collaborations imply knowledge transfer. Wagner formulated the hypothesis that international collaborations are to a higher degree motivated by the need to share data and theories.

ARCHIBUGI & COCO [2004] found that earth- and space sciences followed by physics and mathematics are the sectors that collaborate most internationally. Biology and biomedical research are other sectors where international collaborations are important. When both national and international collaborations were taken into consideration, clinical medicine followed by biomedical research and earth- and space sciences were the sectors with highest share of co-authorship. Concerning USA's international collaborations, astronomy, mathematics and statistics, and physics are the most international fields while agriculture, biology, and medicine are the least internationally. Life science is the field were international collaborations have increased fastest [ADAMS & AL., 2005]. A questionnaire carried out by BOZEMAN & CORLEY [2004] showed the same pattern with zoology, mathematics, materials engineering, and psychology being the most international fields while industrial engineering, mechanical engineering, health professions, biochemistry, and other biological and life sciences being the least collaborative.

KATZ & MARTIN [1997] found that 40–50% of all publications related to Clinical Medicine have more departments mentioned in the publications than number of authors. Similar patterns can be seen in other fields but to a less degree, biomedical research and physics (10–15%), biology, earth and space science (5–10%), and chemistry, engineering and mathematics (<5%). These results suggest that many researchers in clinical medicine tend to hold joint posts with several departments or universities.

Geographical collaborations

The geographical proximity, cultural-, and language similarities between two countries have an impact on the way countries collaborate. In a study carried out by ZITT & AL. [2000] these factors are shown to have a major influence on the collaboration pattern. The aim in this article is not to investigate individual countries but rather to focus on the macro level (Europe) and leave the underlying factors describing preferable collaboration patterns for later studies. KATZ [1994] found that distance has a negative correlation on the collaboration frequency. Countries studied included Australia, Canada, and UK which diminish the linguistic dimension as the three of them are Anglo-Saxon and have English as an official language.

Large countries collaborate less internationally than smaller countries, measured by the absolute number of researchers [SCHUBERT & BRAUN, 1990; NARIN & AL., 1991; MELIN & PERSSON, 1996; GLÄNZEL & SCHUBERT, 2001]. The reason for this is that in a big country the number of total scientists is likely to be higher than in a smaller country. The probability for a scientist to find a potential national collaborator is higher in a country with many scientists than in a country with fewer scientists. NARIN & AL. [1991] concluded that United States and United Kingdom are more reluctant to collaborate with international partners than countries such as Italy, because of the differences in scientific size, number of scientists. In a study carried out by MELIN [1999], differences within the country was studied. The most international oriented universities having a ten percentage unit higher international collaboration rate than the least internationally oriented. Interestingly, when the largest, USA, and smallest, Iceland, countries were excluded from the sample, no correlation between country size and international collaborations were revealed, which does not correspond to out hypothesis. In the same study it was concluded that neither the size of a university nor the geographical location have any impact on national and international co-authorship.

In the same study [MELIN, 1999], EU and USA collaborations were compared, it was shown that there is a significant difference between the number of international co-authored papers, EU universities being more international than American universities. The study does not mention to what extend these collaborations take place within the EU or outside. The same difference between Europe and USA was obtained in a study carried out by ARCHIBUGI & COCO [2004]. Over a 13-year period between 1986 and 1999, USA lost both in world share of international collaborations and world share of total number of scientific publications. When considering the percentage of international co-authored papers (fractional basis), USA has had a higher annual growth rate than most of the EU countries (whole-count basis). The interpretation of the results suggests that USA has decreased its role as an international collaboration partner while the share of the entire scientific production has increased more than in the average EU country. EU countries, according to the study, that tend to collaborate less internationally include Spain, Finland, Ireland, and Greece. Finally, they concluded that intra-EU collaborations have increased while EU collaborations with USA, both for EU as whole as well as individual members, have decreased. Still, USA is the most frequent collaboration partner for EU countries in absolute terms [GEORGHIOU, 1998].

Networks

The collaboration pattern is a result of a complex dynamic process that changes constantly. In this study a network can have national, intra-EU, or extra-EU dimensions. An individual researcher can collaborate both nationally and internationally independent and/or dependent on others. Therefore there is no evidence that national networks necessarily act as intermediary links to international collaborations [SCHOTT, 1993]. Within Europe, the wish to increase interactions between researchers from different EU countries has led to the creation of programs aiming to involve researchers from several countries e.g. EraNet. If these initiatives have resulted in broader co-publication networks is to be further investigated. While different forms of collaborations and their nature have been studied in numerous initiatives the actual size of a network and its actors are less known. ADAMS & AL. [2005] measured the network size and found that all scientific fields have increased from an average of 2.5 authors/paper in 1981 (mathematics being the smallest (1.531) and medicine the largest (3.259)) to an average of 3.9 in 1999, (mathematics being the smallest (1.907) and physics the biggest (7.264)). They also showed that physics is the field that has grown fastest in terms of number of authors/publication over the years between 1981 and 1999. The results suggest that the number of co-authors is growing.

Analytical approach

Data collection

In this study, the intensity of collaboration[2] between researchers in different European countries was measured using data retrieved from ISI online database, Current contents and Dialog published in 1998 and the period 2000–2003. Only articles where considered though this study aims at reflecting "new science" compared to for e.g. book reviews, letters etc. The scientific fields studied include Agriculture, Biology & Environmental Sciences (ABES), Clinical Medicine (CM), Life Science (LS), Physical, Chemical & Earth Sciences (PCES), and Engineering and Computing &

[2] In this article when we refer to collaborations we only consider co-authorship, keeping in mind that collaboration patterns can not only be explained by co-authored articles. Examples of collaborations that are not visible with bibliometrics include sharing of equipment and resources, funding, transmission of know how, and manpower.

Technology (ECT), all based upon the classification done by ISI. The chosen fields are all from the natural sciences where peer-review articles play a more important role than in the social sciences.

Collaborations are often categorized into *internal collaborations*, defined as collaborations within the researchers own institution/organization (intra-institutional), *external collaborations* (inter-institutional) that can be divided into *national collaborations* and *international collaborations.*[3] In this study only international collaborations were considered. The main objective with this study was to analyse the collaboration pattern between individual EU states, candidate and other European countries and their collaboration pattern 1) within EU (intra-EU collaborations) and 2) the rest of the world (extra-EU collaborations). Out of the 25 EU member states, 18 countries were selected for further investigation (they vouch for 99% of the total 25 member states production), also candidate countries and other European countries were investigated, see Table 1.

A table of countries national publication output, their intra-EU collaborations, extra-EU collaborations, and collaborations with USA, was created for each of the scientific fields.

Data processing

The evolution of intra-EU collaborations and EU countries collaborations with USA was studied, data from 1998 and 2003 was used. The rate of collaborations was calculated as a percentage of the total number of articles published in a country. The size of the scientific community in USA is similar to the size of EU and can therefore be considered as two comparable attractors for individual EU countries. We compare if the attraction power of the two communities has changed over the years with the underlying hypothesis that EU is becoming an increasingly important collaboration space. For the rest of the analysis, extra-EU countries, including USA, candidate and other European countries if not specified differently, were considered.

When comparing the five scientific fields, the percentage of national production and collaborations (intra-EU and extra-EU) out of the total numbers of publications in the field was calculated.

The size of a country was identified by the number of articles published in the country and is presented on a log scale, to illustrate the differences between country sizes. In the article we will refer to small countries including in particularly Slovakia, Portugal, Ireland, Czech Republic, Hungary, Greece and big countries including UK, Germany, France, and Italy. It is difficult to make an exact division of countries though the number of articles differs between scientific fields.

[3] Outside the author's own country.

Networks were defined as publications including more than two countries. This study focuses on the EU collaboration pattern where a network can include either only intra-EU collaborators or extra-EU collaborators. First, the total networking, including all possible networks, was computed for the five scientific fields. Second, the rate of networking, calculated as the percentage of the total number of publications, in each collaboration form, and finally the network size, defined as the number of countries involved in each network, were calculated.

Results

Evolution of collaborations

First the evolution of co-authorship was studied. The data in Figure 1 confirms earlier research that co-authorship have increased during the last years ($p<0.0001$), the exception include Poland. This was true for all EU countries. The figure also indicates that there was a significant difference in collaboration rate between small and big countries both in year 1998 ($p=0.0014$) and 2003 ($p=0.0290$), smaller countries collaborating more than bigger ones. Furthermore, the total publication rate for smaller countries has increased more than for bigger countries ($p=0.0429$), except Slovakia which total publication output decreased with 7%. No differences between new vs. old member states could be observed.

Figure 2a illustrates that EU countries collaborations with USA is not dependent of the country sizes, neither in 1998 ($p=0.3943$) nor in 2003 ($p=0.9032$). Within EU, smaller countries collaborate more with other EU countries than bigger countries ($p<0.0001$ in 1998, $p=0.0002$ in 2003). The intra-EU collaborations have increased between 1998 and 2003, which is in accordance with earlier results; exceptions include Portugal, the Czech Republic, and Poland. Intra-EU collaborations have increased more than EU collaborations with the USA. For Greece the opposite was true. The figure also revels that the overall number of published articles has increased. In Figure 2b we normalize the data over this increase. The same results were obtained, – i.e. EU collaborations have increased more than collaborations with US –, when collaborations with candidate and other European countries are included. The collaboration rate increases but the pattern is still the same (Figure 2b[4]).

[4] BG and SK are excluded from the figure. Both countries had a decrease in publications but the EU collaboration rate is higher than with US.

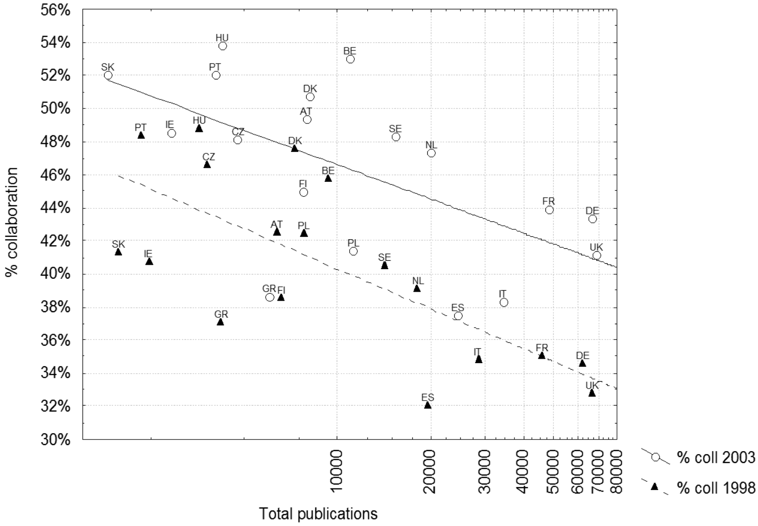

Figure 1. Collaborations (share of total publications) and total number of publications for EU-18 countries 1998 and 2003

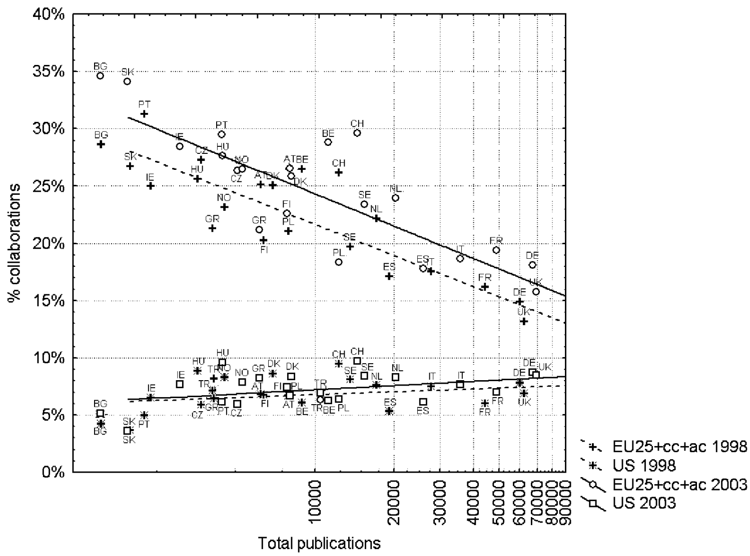

Figure 2a. Collaboration differences between US and EU
Collaborations (share of total publications) with USA and EU and total number of publications for EU-18 countries 1998 and 2003;

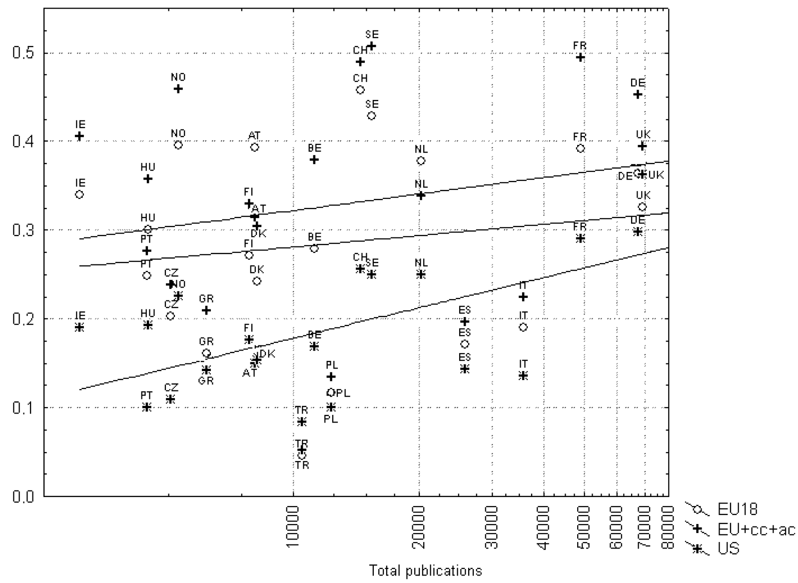

Figure 2b. Collaboration differences between US and EU
normalized collaboration rate, (coll. in 2003 – coll. in 1998) / (tot. pub. 2003– tot. pub. 1998)

Scientific field differences

Collaboration patterns might differ between scientific fields which are further investigated in this section. The same collaboration pattern was found in all scientific fields for articles published in 2000–2003. Smaller countries tended to collaborate internationally significantly more than bigger countries and it was primarily intra-EU collaborations that accounted for the differences. PCES was the field where intra-EU collaborations were most common, followed by LS.

The situation for extra-EU collaborations was more complex, Figure 3. In both CM and ECT extra-EU collaborations did not differ significantly between countries of different sizes. In LS and PCES the big countries (UK, DE, and FR) accounted for the difference of extra-EU collaborations, bigger countries having more extra-EU collaborations. ABES is the only field where there is a significant difference between small and large countries in, big countries collaborating more with extra-EU. PCES was the scientific field where most extra-EU collaborations took place followed by LS, ECT, and CM.

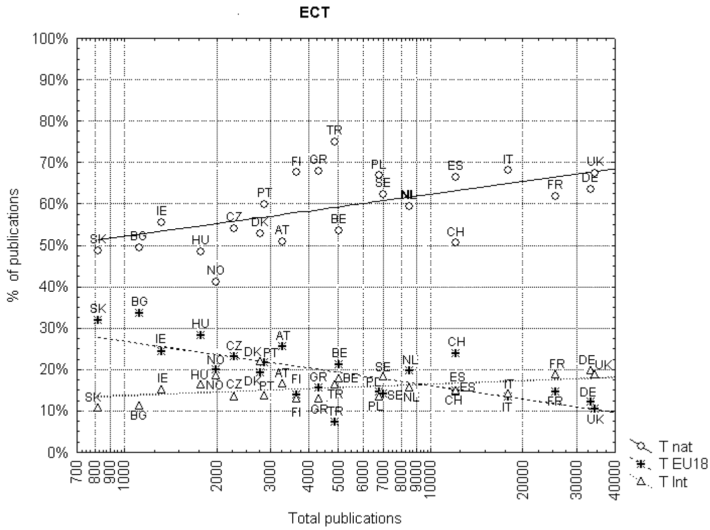

Figure 3a. Collaboration (share of total publications) differences between scientific fields (ECT) and modes of collaborations (intra-EU, extra-EU)

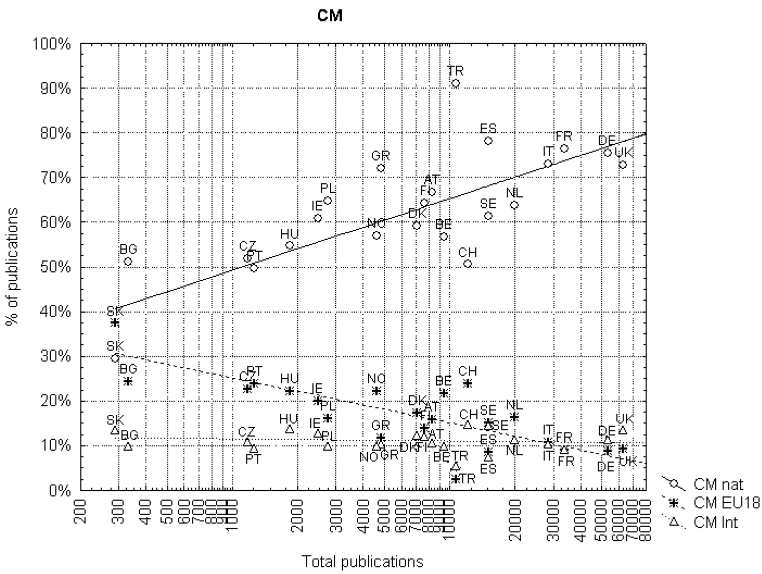

Figure 3b. Collaboration (share of total publications) differences between scientific fields (CM) and modes of collaborations (intra-EU, extra-EU)

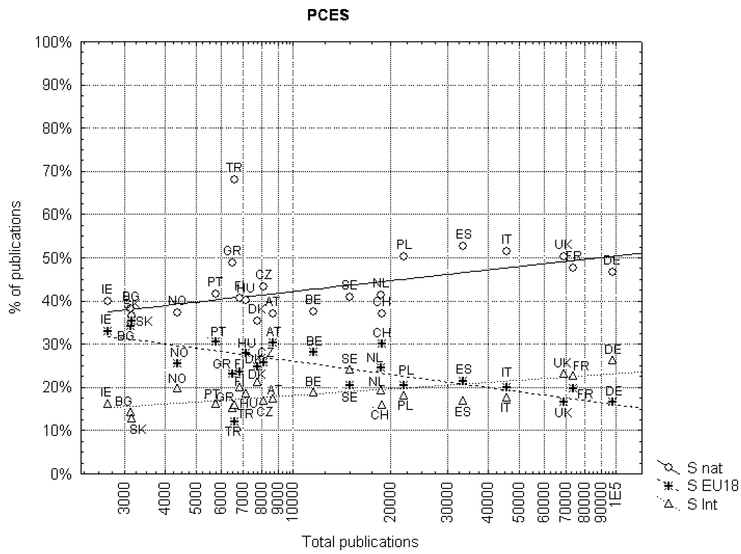

Figure 3c. Collaboration (share of total publications) differences between scientific fields (PCES) and modes of collaborations (intra-EU, extra-EU)

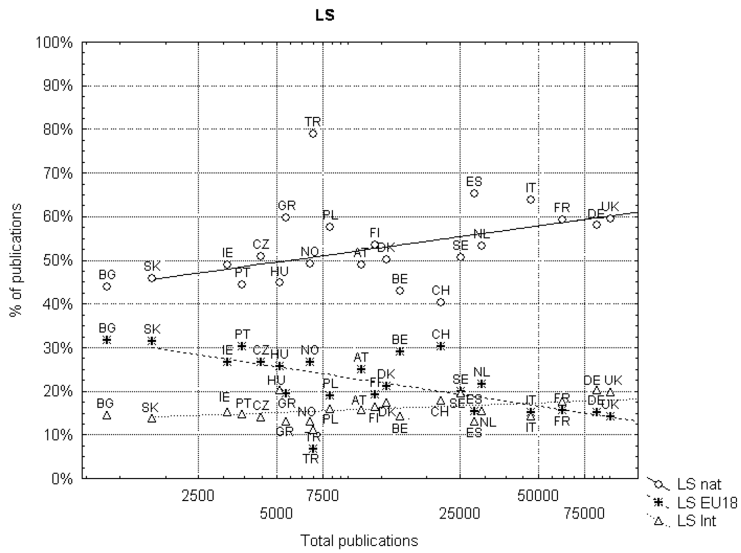

Figure 3d. Collaboration (share of total publications) differences between scientific fields (LS) and modes of collaborations (intra-EU, extra-EU)

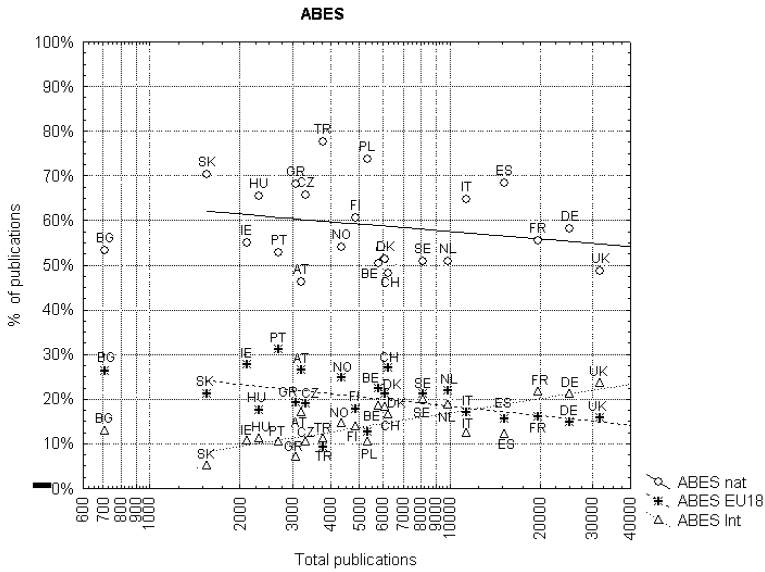

Figure 3e. Collaboration (share of total publications) differences between scientific fields (ABES) and modes of collaborations (intra-EU, extra-EU)

A country's national share increases significantly ($p<0.05$) with the country size in all scientific fields, except ABES, indicating that the level of international collaborations differ between scientific fields. In ABES smaller countries tend to have a higher national production than bigger, not significant. ABES is also the field where smaller countries have the highest national share. For the bigger countries the highest national production can be found in CM. PCES was the field with the lowest national share.

Collaborations are also more or less present in different countries and also dependent on the scientific field e.g. Finland had a low collaboration rate in ECT in comparison to its collaboration rate in other fields.

EU vs. non-EU differences

We tested the importance of candidate countries and found that when included as Intra-EU collaborations the collaboration rate does not change (in some countries only with a 1% increase). When other European countries (Norway and Switzerland) were included the collaboration rates change more in some countries than others.

It is especially for the Nordic countries, Germany and Austria that the inclusion of other European countries in the EU collaborations increases the collaboration rate. For the Nordic countries it is the addition of Norway and for Austria and Germany it is the addition of Switzerland that increases the collaboration rate.

Networks

In this section the networking in the different scientific fields and collaboration forms was further analyzed, with the objective to study EU interactions. In Figure 4 the share of total networking, is shown for the different countries/fields. There was a significant difference between the size of the country and the share of networks in all scientific fields except ABES, indicating that small countries participate more in networking. Networking was most common in PCES while it was least common in ECT.

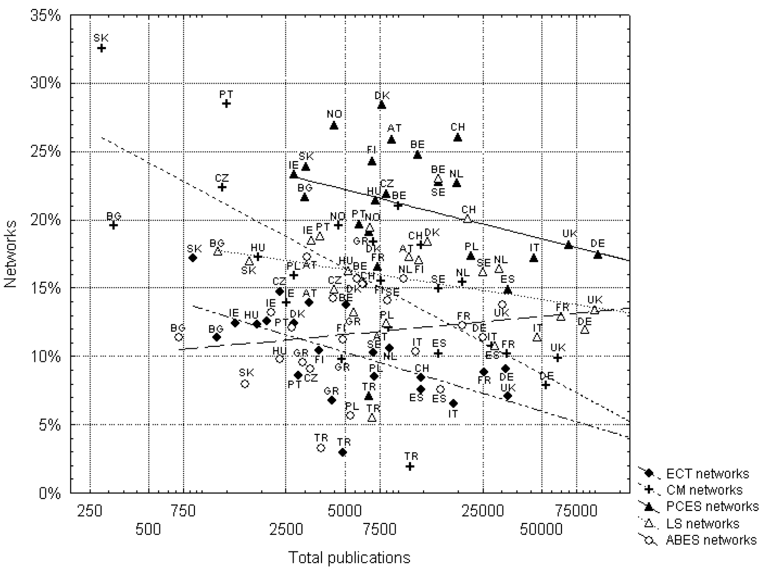

Figure 4. Share of networks in the five scientific fields

The network size was defined as the number of countries co-authoring an article. In all fields there was a significant ($p<0.05$) difference between the intra/extra EU network sizes. The same pattern was found in all fields, intra-EU including the largest number of countries. CM was the field with most countries involved (3.67 countries/article).

In Figure 5 the relative importance of networking is illustrated in both intra-EU and extra-EU collaborations.

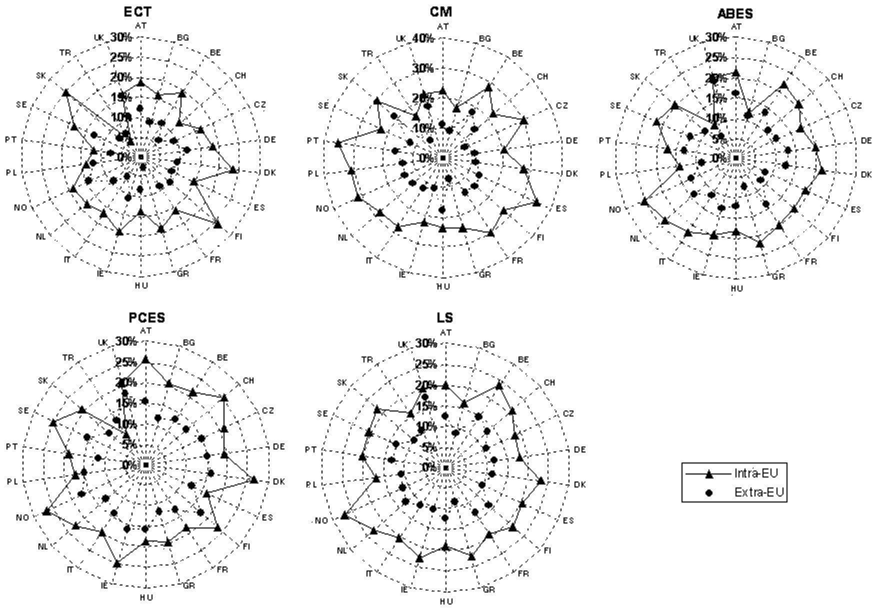

Figure 5. Network rates in intra-EU and extra-EU collaborations

Networks are most common in intra-EU collaborations for all fields. Intra-EU networks were most presented in CM, where networking accounted for on average 26% of all the EU collaboration publications. In LS, PCES, ABES, and ECT intra-EU networking papers account for between 17–21%. Networking is less common in extra-EU collaborations in all fields. Extra-EU networks were most present in PCES where they make up 14% of all extra-EU collaborations. ECT was the field where both intra- and extra-EU networking was least present.

Discussion

The study was carried out with the objective to shed light on EU collaborations, the factors influencing a country's collaboration pattern, and differences between scientific fields. Three specific questions were addressed:

1. To what extent do researchers in EU countries collaborate within Europe and outside?

In accordance with earlier research it can be concluded that co-authorships in general have increased and especially so intra-EU co-authorship. Co-authorship within Europe is more common than publications with extra-EU based authors. Exceptions include UK, France and Germany, the three countries with the highest number of publications. One explanation could be preferred co-authorship with former colonisations and countries with linguistic and cultural similarities. Another explanation could be the ability to attract foreign researchers which could be easier in a big country with more exposure. Finally, networks in intra-EU collaborations play a more important role than in extra-EU where collaborations are carried out on a more bilateral level. From a policy perspective it is tempting to believe that the increased networking and the increasing number of countries participating in networks are results from policy induced collaborations e.g. ERAnets.

2. What are the factors influencing a country's collaboration pattern?

Both collaboration differences related to country size, measured by total number of publications, and scientific fields could be observed. Earlier studies have already argued for differences between small and big countries. In this study the finding was investigated further and the results show that differences related to size can only be found when collaborations take place within Europe but was not an influencing factor when collaborations were carried out with countries other than the EU member states. Smaller countries collaborate more with other EU countries than bigger countries while the collaboration rate with extra-EU partners was not dependent of the country's size, the exception includes ABES. The results were robust and even if the smallest and biggest countries were removed there was still a significant difference between country sizes. This is not in accordance with the study carried out by MELIN [1999] where there was no correlation between size and international collaborations after exclusion of Iceland (smallest country) and the USA (biggest country).

When the required competence can not be found in the own country, research groups in small countries seem to primarily look for geographical proximity. This was supported with the fact that bigger countries have a higher share of pure national publications than smaller countries. One could argue that Europe is used as a "national" playing field for small countries while bigger EU countries in themselves are sufficient in size. On the other hand, extra-EU collaborations seem to address different needs where linguistic, geographical, or cultural ties in form of so called preferential collaborations seam to be more of an influential factor.

Smaller countries were more involved in networking, publications including more than two countries, than big countries but there was no correlation between the size of a country and the size of the networks, number of countries co-authoring a paper.

3. Does the collaboration pattern differ between scientific fields?

Our results indicate that PCES was the scientific field where most international collaborations took place followed by LS, ECT, ABES and CM, which is in line with the study carried out by ARCHIBUGI & COCO [2004]. The same pattern was true for both intra-EU and extra-EU collaborations, with the exception of ABES. The results also illustrate that PCES has high extra-EU co-authorship rate. PCES is a field which often requires expensive and large-scale equipment e.g. synchrotrons and telescopes. The equipment can rarely be found on the national level and researchers are therefore entirely dependent on international based initiatives. Earth Sciences can be classified as a field where international collaborations are motivated by access to specific resources e.g. water, soil or other natural resource. LS include subfields such as e.g. immunology, cell biology, and pharmacology, which include the processing of large data. According to WAGNER [2005] data driven collaborations should be less international collaboration intensive which is in accordance with our results. Today, huge online databases including e.g. sequences, markers and mapping data, and protein structures, are available free of charge on Internet where researchers can upload and download information. Articles in ECT, according to the results, were less extra-EU collaborative intensive than in PCES and LS. The collaborations that take place in this field could be initiated by the motivation to share equipment and data. Finally, it could be concluded that CM was the least international collaborative field where occurring collaborations mainly are driven by the objective to share data such as e.g. statistics and clinical tests. In this study, national collaborations were not investigated which could make up for the international collaborations carried out in CM. One could expect frequent collaborations between universities and closely located hospitals in this field.

It was only in ABES where there was a significant difference in extra-EU collaboration rate between small and big countries. This could possible be explained by preferable collaborations with emerging scientific countries e.g. former colonisations.

The results could suggest that the EU Framework Programs (FP's) for R&D have succeeded in the attempt to increase collaborations between member states. 'Data-sharing collaborations' are nowadays mostly facilitated by Internet, similarly as for 'ideas-sharing collaborations' where also personal encounters at conferences may play an important role. Therefore, the efficiency of policies aiming at increasing collaborations can be questioned in the first two cases where as in the case of 'equipment/resource-sharing collaboration', further international initiatives might be fruitful. Also, initiatives aiming at facilitating the mobility of researchers could support collaborations even further.

Conclusion

The differences between intra- and extra-EU collaboration patterns clearly point to a Europeanization of shared co-authorship rather than Internationalization. Earlier work has argued that this Europeanization is over-estimated [MOED, 1991; LEYDESDORFF, 2000; ZITT, 2000]. These studies have not included the majority of the EU member states, or not differentiated between intra- vs. extra-EU co-authorship, or are too old to reflect the current state in Europe.

It is not clear whether the FP's or/and individual countries have a more specific role in the attraction or initiation of collaborations and whether the size of a country, new- or established member states makes a difference. Also, the importance of preferential collaborations, due to geographical, linguistically, and political links have to be further investigated. The authors are testing whether these preferential collaborations is the explanatory factor for the so-called Europeanization. The first results suggest that this is not the case and even if collaborations with a preferred partner are removed the average EU country still collaborate with intra- EU than extra-EU, showing that the collaborations are not completely scattered among regional clusters.

References

ADAMS, J. D., BLACK, G. C., CLEMMONS, J. R., STEPHAN, P. E. (2005), Scientific teams and institutional collaborations: Evidence from US universities, 1981–1999. *Research Policy*, 34 : 259–285.

ANDERSSON, A. E., PERSSON, O. (1993), Networking Scientists. *Annals of Regional Science*, 27 : 11–21.

ARCHIBUGI, D., COCO, A. (2004), International partnerships for knowledge in business and academia – A comparison between Europe and the USA. *Technovation*, 24 : 517–528.

BEAVER, D. D. (2001), Reflections on scientific collaboration, (and its study): past, present, and future. *Scientometrics*, 52 : 365–377.

BOZEMAN, B., CORLEY, E. (2004), Scientists' collaboration strategies: implications for scientific and technical human capital. *Research Policy*, 33 : 599–616.

FRAME, J. D., CARPENTER, M. P. (1979), International research collaboration. *Social Studies of Science*, 9 : 481–497.

GEORGHIOU, L. (1998), Global cooperation in research. *Research Policy*, 27 : 611–626.

GLÄNZEL, W. (2001), National characteristics in international scientific co-authorship relations. *Scientometrics*, 51 : 69–115.

GLÄNZEL, W., SCHUBERT, A. (2001), Double effort = Double impact? A critical view at international co-authorship in chemistry. *Scientometrics*, 50 : 199–214.

GORDON, M. D. (1980), A critical reassessment of inferred relations between multiple authorship, scientific collaboration, the production of papers and their acceptance for publication. *Scientometrics*, 2 : 193–201.

HAGSTROM, W. O. (1965), *The Scientific Community*, Basic Books, New York.

KATZ, J. S. (1994), Geographical proximity and scientific collaboration. *Scientometrics*, 31 : 31–43.

KATZ, J. S., MARTIN, B. R. (1997), What is research collaboration? *Research Policy*, 26 : 1–18.

LEWISON, G., CUNNINGHAM, P. (1991), Bibliometric studies for the evaluation of trans-national research. *Scientometrics*, 21 : 223–244.

LUUKKONEN, T., PERSSON, O., SIVERTSEN, G. (1992), Understanding patterns of international scientific collaboration. *Science Technology & Human Values*, 17 : 101–126.

LUUKKONEN, T., TIJSSEN, R. J. W., PERSSON, O., SIVERTSEN, G. (1993), The measurement of international scientific collaboration. *Scientometrics*, 28 : 15–36.

MELIN, G. (1999), Impact of national size on research collaboration – A comparison between Northern EU and American universities. *Scientometrics*, 46 : 161–170.

MELIN, G., PERSSON, O. (1996), Studying research collaboration using co-authorships. *Scientometrics*, 36 : 363–377.

NARIN, F., STEVENS, K., WHITLOW, E. S. (1991), Scientific cooperation in Europe and the citation of multinationally authored papers. *Scientometrics*, 21 : 313–323.

PERSSON, O., GLÄNZEL, W., DANELL, R. (2004), Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60 : 421–432.

PRICE, D. J. D. (1963), *Little Science, Big Science*, Columbia University Press, New York.

RIGBY, J., EDLER, J. (2005), Peering inside research networks: Some observations on the effect of the intensity of collaboration on the variability of research quality. *Research Policy*, 34 : 784–794.

SCHOTT, T. (1993), World science – globalization of institutions and participation. *Science Technology & Human Values*, 18 : 196–208.

SCHUBERT, A., BRAUN, T. (1990), International collaboration in the sciences, 1981–1985. *Scientometrics*, 19 : 3–10.

SMITH, M. (1958), The trend toward multiple authorship in psychology. *American Psychologist*, 13 : 596–599.

WAGNER, C. S. (2005), Six case studies of international collaboration in science. *Scientometrics*, 62 : 3–26.

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.

II

Research Evaluation, 19(5), December 2010, pages 373–384
DOI: 10.3152/095820210X12809191250924; http://www.ingentaconnect.com/content/beech/rev

What do European research collaboration networks in life sciences look like?

Pauline Mattsson, Patrice Laget, Anna Nilsson Vindefjärd and Carl Johan Sundberg

For scientists and policy-makers it is important to understand the value of networks and collaborations for scientific quality and commercialisation of research results. The interplay between funding mechanisms and research collaborations is of particular importance in this context. We explore this interplay with an empirical analysis of international research collaborations involving EU countries as manifested by co-publication patterns and participation in life science projects funded by EU Framework Programmes. Our data confirms the importance of geographical proximity, however, it also indicates that 'top-down' funding disbursed through FPs is related to collaborations that are more European despite encouragement of collaborations with non-European partners. In contrast, co-publication patterns suggest that European researchers tend to co-author more with global, rather than exclusively European partners and that this tendency is on the rise. These findings shed light on the influence of different funding policies that seek to nurture more diverse and integrated research partnerships.

THE TENDENCY to collaborate outside national borders has been present since the early days of modern science (Sörlin, 2004) but, in the last decade, a number of studies have indicated a marked increase of international research collaborations measured by co-publications (NSF, 2007; Glänzel, 2001; Adams *et al*, 2005). Also the attention from both the scientific world as well as among policy-makers has increased and resulted in a number of policy initiatives encouraging collaborations, such as the majority of Framework Programme activities. The benefits of research collaborations are many and range from scientific to economic and political factors. Georghiou (1998), for example, distinguishes between *direct* benefits that accrue when accesses to complementary expertise, knowledge or skills to enhance scientific or technological excellence are the principal motivations and *indirect* benefits when collaborations are driven by external goals of an economic, political or cultural nature.

Researcher-initiated collaborations are most likely based on the researchers' ideas and resource needs within specific scientific thematic areas and may be arranged with or without targeted support. If the initiators are politicians or policy-makers the aims are expected to have more of a political or economical character. These 'top-down' initiatives are more likely to have a specific budget for a limited number of years and calls for proposal typically contain a number of non-scientific criteria the applicants have to fulfil. Research collaborations can also be initiated by companies and private foundations. This study is limited to study only the first two alternatives without indicating that these collaborations are more important.

Pauline Mattsson (corresponding author) is at the Unit for Bioentrepreneurship, LIME, Karolinska Institutet, 171 77 Stockholm, Sweden; and the Technopolis Group, Grevgatan 15, 114 53 Stockholm; Sweden; Email: pauline.mattsson@ki.se; Tel. +49 1773809848. Patrice Laget is at the Inserm Unit 646, 10, rue Andre Boquel, 49100 Angers, France; Email: patrice.laget@univ-angers.fr. Anna Nilsson Vindefjärd is at the Unit for Bioentrepreneurship, Stockholm, Sweden and at the Swedish Association of the Pharmaceutical Industry, 118 92, Stockholm, Sweden; Email: anna.nilsson.vindefjard@lif.se. Carl Johan Sundberg is at the Unit for Bioentrepreneurship, Stockholm, Sweden; Email: Carl.J.Sundberg@ki.se.

The authors would like to thank Joaquin Azagra-Caro, John Skår and Dieter Zinnbauer for their valuable comments on earlier versions of this paper.

In Europe much effort has been put into increasing networking between European countries using a number of 'top-down' initiatives. The Framework Programmes (FP) for research, technological development and demonstration (RTD) are the central instruments to foster and strengthen RTD at the European level. Since the first FP was introduced 1984, its importance has been enhanced, as indicated by the increased overall budget for FP and the increased share of R&D funding coming from the European Commission in all/most EU countries (Eurostat, 2009; Widhalm *et al*, 2001). Since the first proposal in the year 2000 to establish a European Research Area (ERA), research collaborations have been regarded as a means to ascertain better integrated research activities within Europe (EC, 2000). This is further highlighted by the fact that the vast majority of FP projects also have non-scientific requirements, two of which are of particular importance for this study:

1. Funding (but not participation) is restricted to EU member states, candidate and accessing countries; and
2. Projects require participants from several member states (multilateral collaborations).

Therefore, this study addresses the issue of Europeanisation of collaboration networks by exploring the structure and pattern of research collaboration networks from two different angles, one reflecting collaboration networks in the funding phase and one type reflecting the outcome of research.

More specifically, we carried out an empirical analysis of international research collaborations involving EU countries in the field of life sciences as manifested by:

- The composition of research partnerships in projects funded under European Commission (EC) Framework Programmes 4–6; and
- Co-publication networks over a time period of 10 years (1994–2006).

The aim was to examine whether the structure of EU-funded activities is reflected in the output while at the same time not to create a dichotomy between European research policy-initiated collaborations and collaborations reflected on in co-publications.

International research collaboration and research funding — a diverse landscape with national and international actors

Policy-initiated research collaborations

Researchers are increasingly called on by governments, business, community groups and others to assist with solving many of the complex problems that societies face (Bammer, 2008). The growth of collaborative research owes much to the provision of funding from governments concerned to encourage interdisciplinarity and interactions between different sectors (Rigby and Edler, 2005). Bozeman and Boardman (2003) goes as far as talking about a new 'era of inter-institutional research collaborations' referring to US science and technology policy moving from decentralised support of small investigator-initiated research projects to large-scale and oftentimes centralised, block grant-based, multidisciplinary research. In tandem, the scale, scope and diversity of funding programmes have increased and transnational, national and regional initiatives all play a role.

When national governments commit to international collaboration activities it is in the form of either shared infrastructure or funding. These initiatives are expected to have an overall policy focus and take place as international agreements with the objectives of creating new or improving existing relations between countries. National funding bodies rarely finance entire multilateral international collaboration networks. They rather focus their support on researchers based in the home country and on bilateral agreements for several reasons.

First, the ability for governments to oversee and to influence research directions is reduced when the number of countries that participate in agreements increases (Wagner, 2002). Second, the administrative burden of the involvement of several countries is often too big to be shouldered by an individual national funding organisation. Finally, national protectionism may be involved in the allocation of national sources requiring a focus on securing national return on investment. In international networks however, the home-country participants receive only a smaller share of benefits, which are dispersed throughout the entire network. As a consequence, multilateral collaborations are therefore more likely to be funded by and result from international or supranational funding

Since the main source supporting international, mostly intra-European, collaborations comes from funding initiatives established by the EC, it is expected that multilateral collaborations have become more prevalent, assuming that output is related to funding. As a matter of fact, the relative share of funding from the EC has increased, from 6% of the gross domestic expenditure on R&D in 1995 to 9% in 2005.

Hypothesis 1: Multilateral collaborations have increased.

Europeanisation of networking

Since 2000, much attention at European level has been focussed on the realisation of an ERA. Over the 10-year period since its 'launch' the objectives have expanded but, overall, the main aim is to promote European competitiveness, jobs and growth.

Over the 10-year period since the 'launch' of the European Research Area, its objectives have expanded but, overall, the main aim is to promote European competitiveness, jobs and growth

This is done through a multitude of activities where integration of research activities is one of several objectives. In the policy document, *Towards a European Research Area*, the European Council, the body of EU member state ministers, called for 'better integrated research activities at Union level'. The underlying vision for such an integrated system is that researchers will be able to choose freely to collaborate with other scholars purely based on research grounds in view of scientific excellence without cultural, geographical and linguistic proximity biasing such decisions (EC, 2000).

The bulk of efforts supporting the realisation of the ERA vision have been carried out through the already existing FPs. FP5 (1998–2002) was the first programme to mention explicitly the objective of an ERA. Later on in FP6 (2002–2006) the initiative Network of Excellence was designed to facilitate the networking of talented researchers independent of origin. The foremost aim is to shift the focus from small projects towards fewer and larger ones. One criterion for participation is that the minimum number of participants is set as three independent legal entities based in different countries, of which at least two should be member states or associated candidate countries. Another collaboration-focussed initiative under FP6 is ERA-NET, which is specifically geared towards providing support for the trans-national networking and coordination of national research programmes. These initiatives are of long-term nature and the pattern of researchers' collaboration levels has not yet been evaluated. This focus on collaboration continues. In the FP7 (2006–ongoing) more than 60% of the total budget is allocated to collaborative network activities.

Several recent studies have begun to investigate the level of European research integration, but the conclusions vary. In an early contribution to this debate Leydesdorff (2000) concluded that Europe in the early 1990s could not be considered as a single publication system, based on Markov property of the distribution. Leydesdorff also argued that the objective of further integration might be less important in the RTD policies developed by the EC than keeping the system in good shape and meeting competition from the outside world. The FPs and similar programmes should be considered as resource allocations for this purpose.

Similarly, Banchoff (2002) concluded that two decades of efforts towards an economic and monetary union have not succeeded in an integrated European S&T landscape with integrated national policies. The lack of integration is partly due to strong national interests but also to the institutionalisation of European institutions, such as the FPs, resulting in administrative and time-consuming management (Luukkonen, 2002).

In contrast, Stein (2004) arrived at a more positive conclusion from a study of the European knowledge system[1] along the lines of scientific co-operation; co-operation in technology development; and S&T policy. She came to the conclusion that the examined indicators — for example, number and variety of projects, programmes, networks, policy structures — support the existence of a European knowledge system. Adding to this empirical inconclusiveness are Okubo and Zitt (2004) who paint a mixed picture from their analysis of co-publications to study the evolution of Europeanisation, defined as the intensity and orientation of collaborative linkages. They conclude that Europeanisation, measured by European co-authorships as a proportion of all international co-authorship, has increased marginally during a 10-year span, 1987–1989 and 1997–1999. Only the least Europeanised countries showed an increase and the most Europeanised countries showed a slight down-trend.

The broad areas of support have been fairly stable though specific areas — for example, ICT, energy, environmental and life sciences research — but topic-specific questions have varied. The majority of instruments have mainly focused on European industry (Arnold *et al*, 2008). This is obvious when studying the level of companies participating in the different FPs. As for life science-related areas, this is less the case; the research areas targeted in FP4 and FP5 were rather basic science-oriented whereas FP6 has been more applied.[2] This is further highlighted through the participation of industry in life science-related projects having a participation rate of 5% in FP5 and 15% in FP6.

Since the main funding supporting international collaborations in Europe comes from the EC, more specifically the FPs, and the importance has increased, we expect to see an increase in European co-authorship, assuming policy instruments have an impact on the output.

Hypothesis 2: Europeanisation is reflected in co-publications.

Geographical proximity

The extent to which the efforts to create an integrated European research area have borne fruit since the year 2000 has, to our knowledge, not been studied. A growing literature has begun to develop and test hypotheses on the factors important in shaping international collaborations networks (Breschi and

Cusmano, 2004; Roediger-Schulga and Barber, 2007; Scherngell and Barber, 2008). The hypotheses include geographical proximity, language, shared history, and funding mechanisms.

The significance of geographical proximity has received increasing attention from a variety of perspectives (Pond *et al*, 2007; Greunz, 2003). Although improved transport and communication means and increases in R&D funding are all factors facilitating long-distance collaborations (Wagner, 2006), Frenken *et al* (2007) found that national borders still have an impeding impact on European research collaborations. A number of reasons have been advanced to explain this situation. According to the 'system perspective', national research and innovation policies are still shaped by borders and rooted in culture traditions. These national traditions and research systems shape the behaviour and regulation patterns in higher education, science, labour relations, and financial systems (Lundvall, 1992; Lundvall *et al*, 2002). From a somewhat different angle, Jaffe (1989) stressed that the transfer of tacit knowledge is sensitive to spatial proximity and that face-to-face interactions are, and will probably always be, important. A number of other studies have begun to empirically test and explain the importance of geographical proximity both at the national as well as the international level (Okubo and Zitt, 2004; Frenken, 2002; Zitt *et al*., 1999; Katz, 1994). Taking this information into account we formulate the following hypothesis:

Hypothesis 3: Preferential collaborations between countries are present in both co-publications and FP projects.

Analytical framework

Data collection

The publication data consisting of articles and notes were gathered from the ISI Web of Knowledge of Thomson Reuters. Eighteen EU member states, accounting for 99% of the total EU-25 publication production, were included (Table 1). Two separate datasets were studied, one covering the years 1995–1997 (before ERA), and one covering the years 2003–2005, when the ERA concept had already been introduced. This study is limited to the field of fields of fundamental biology and medicine life sciences according to the classification of the OST Office of Science and Technology (Zitt and Teixeira, 1996).

The bilateral and multilateral co-authorship pattern (of the selected 18 countries) with all EU countries including candidate and accessing countries and the global co-authorships pattern (both European and other countries appear in the address) were studied. Intra-national and single-authored papers are not analysed in this article since the aim was to study international collaborations.

Table 1. EU member states included in the study

EU countries	Abbreviation
Austria	AT
Belgium	BE
Czech Republic	CZ
Denmark	DK
Finland	FI
France	FR
Germany	DE
Greece	GR
Hungary	HU
Ireland	IE
Italy	IT
Netherlands	NL
Poland	PL
Portugal	PT
Slovakia	SK
Spain	ES
Sweden	SW
UK	UK

The information regarding the FPs (FP4, FP5, FP6) was collected from CORDIS. Only projects related to programmes focusing on life sciences were included.[3]

Definitions

An article was assigned to a country, based on the address field of authors. Each link between distinctive countries is rated one. International co-authorship (collaboration) is an article including at least two countries. An international bilateral co-authored (collaboration) paper includes two countries while multilateral co-authored articles include more than two countries. The terms co-authorship and collaboration are used interchangeably (Table 2.

European collaborations involve only European countries (EU member states + candidate + accessing countries). Global collaborations also involve non-European countries. The same nomenclature was used for the countries participating in the FPs.

Table 2. Levels of co-authorship

Co-authorship	No. of countries
International co-authorship	≥ 2
(International) bilateral co-authorship	= 2
(International) multilateral co-authorship	> 2

Statistics

The concept of preferential partnership was calculated using the probabilistic activity index (PAI) (Okubo and Zitt, 2004). This index corrects for the size of the countries and calculates the observed co-authorship between two countries to the value that would be expected from their respective size. If size is not corrected for, large countries such as the UK, France and Germany would be expected to be the first partner for the majority of countries studied.

$(C_{x,y}*T)/(C_x*C_y)$

where $C_{x,y}$ is the number of co-authored papers between countries x and y;

T is the total number of international co-authored papers by the 18 countries;

C_x is the total number of internationally co-authored papers by country x; and

C_y is the total number of internationally co-authored papers by country y.

A preferential partnership can be considered to exist when the PAI index between two countries is greater than 1 and thereby stronger than what would be expected from the publication output. The PAI index was also used to identify preferential collaborations in the FPs.

Outliers and extreme values are identified for each individual country. These values indicate whether an individual country appears as an extreme or outlier in comparison to the average rate of bilateral co-authorship with other countries. The outlier and extremes are calculated as follows:

extreme (*) > $UBV + *o.c.*(UBV - LBV)$ or extreme < $LBV - *o.c.*(UBV - LBV)$

outlier (o) > $UBV + 2*o.c.*(UBV - LBV)$ or outlier < $LBV - 2*o.c.*(UBV - LBV)$

where UBV = the mean + standard error or the 75th percentile;

LBV = the mean − standard error or the 25th percentile; and

$o.c.$ = outlier coefficient set to 1.5.

Results

Bilateral international collaborations are more common than multilateral ones but the latter have increased more

Both multilateral and bilateral international co-authorship have increased over the last decade (Figure 1). Multilateral co-authorship has increased more than bilateral in all studied countries (Figure 1) and in 2003–2005 accounted for 36% of all co-authorships (Figure 2). Still the majority, 64%, of international co-authored articles involve only two countries. The results are in line with Hypothesis 1.

Researchers tend to collaborate more in global multilateral collaborations rather than European ones

Studying the FP data we find that the number of involved non-European countries has increased over the different FPs. In FP4 only 4% of projects involve non-European countries and Israel accounted for 86% these, possibly because it is the most

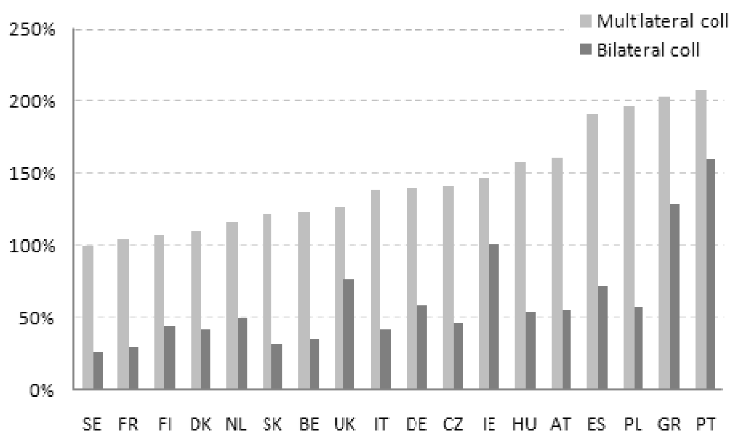

Figure 1. Relative increase of multilateral and bilateral collaborations between the periods 1995–1997 and 2003–2005

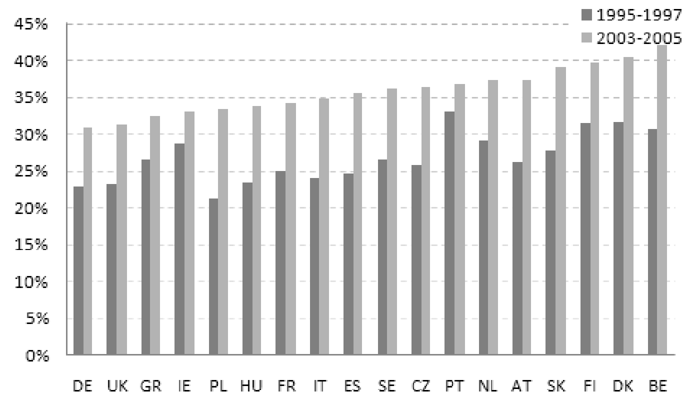

Figure 2. The multilateral share of total number of co-authored papers in the periods 1995–1997 and 2003–2005

scientifically active of the associated countries. In FP5 the participation rate of non-European countries increased to 8%, with Israel accounting for a smaller share of 69%. In FP6 the non-European participation rate increased even further and reached 31%, Israel accounting for 54% of these.

The most likely explanation for the sharp increase of participating non-European countries is that specific international scientific cooperation activities (INCO), targeting collaboration with third countries, were integrated into the thematic areas in FP6. Looking at the United States, the non-European country with which EU member states co-publish most, the FP participation rate of this country is between 1% and 2% when taking all projects into account. If only projects that involve non-European countries are considered, the USA participated in 1% of such projects under FP4, in 15% under FP5, and in 7% under FP6. Other non-European industrial countries such as Japan, Canada and Australia also exhibit relatively low participation rates.

In Figure 3, the relative importance of European compared to global multilateral co-authorship as a share of the total (multi + bilateral) number of international co-authored papers is illustrated. It is evident that global multilateral co-authorship is relatively more common than pure European multilateral co-authorship and has also significantly ($p < 0.05$) increased since 1995–1997. Bigger countries (France, Germany, UK) were the countries

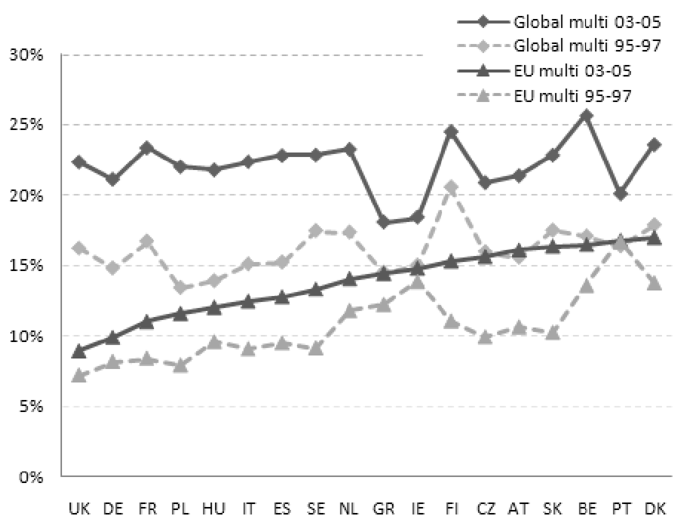

Figure 3. Share of global and EU multilateral co-authored papers of total numbers of co-authored papers

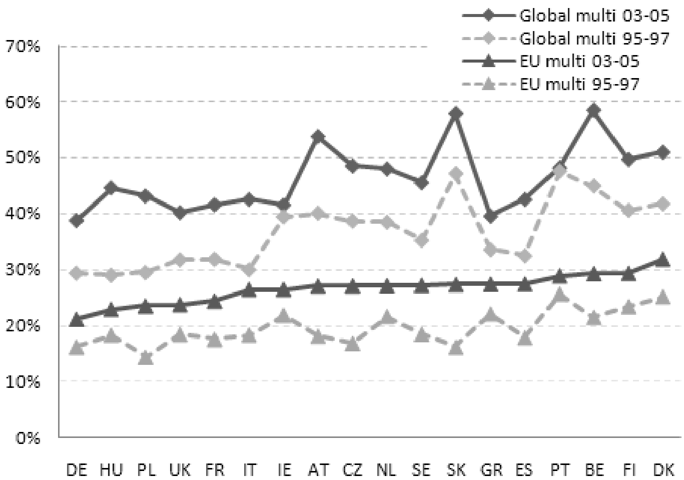

Figure 4. Share of multilateral co-authored papers of total global and EU co-authored papers

with least European multilateral co-authorship both in 1995–1997 and 2003–2005. In global multilateral collaborations there was no significant difference between countries of different sizes. Similar results were obtained by Mattsson *et al* (2008) where smaller countries were found to co-author more with other EU countries than bigger ones while the co-authorship rate with extra-EU countries was not dependent on the country's size. The relative importance of European vs. global multilateral co-authorship is illustrated in Figure 4. Multilateral collaborations are more common in global co-authored articles where it make up on average 45% compared to 27% in only EU co-authored papers. The figure also shows that global multilateral co-authorship has increased more.

Geographical proximity has an influential role in both co-publication and Framework Programme data

The next objective was to focus on the European research landscape and examine the pattern of scientific collaboration among European countries.

Figure 5 and 6 illustrate all preferential collaborations, based on co-publications, between EU countries during the two time periods (as defined earlier with PAI indexes greater than 1). All countries had at least one preferred collaboration partner. This can partly be explained by geographic, linguistic, or political proximity.

The major exceptions are Portugal and to a certain degree Greece. In 1995–1997, Portugal had preferential collaboration with several European countries.

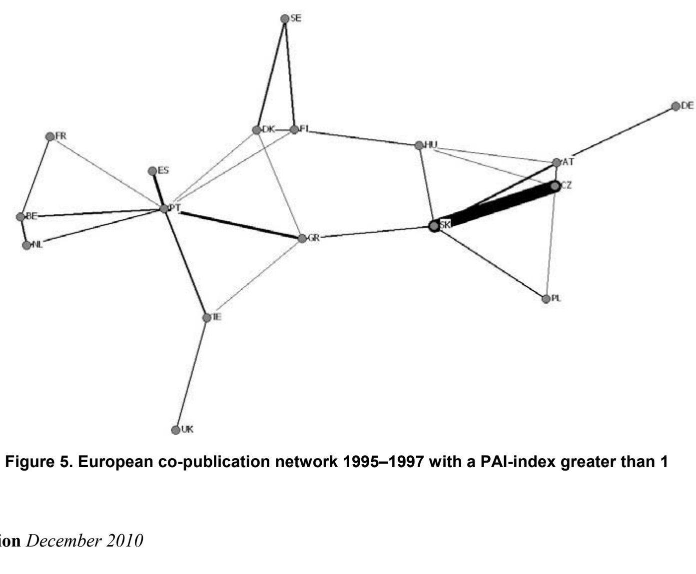

Figure 5. European co-publication network 1995–1997 with a PAI-index greater than 1

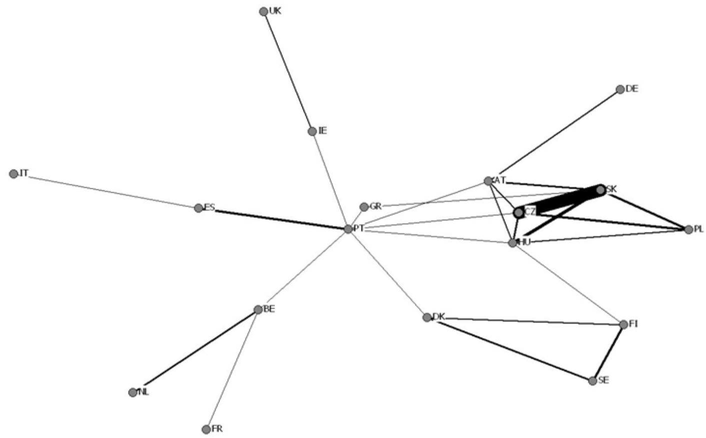

Figure 6. EU co-publication network 2003–2005 with a PAI-index greater than 1

In 2003–2005 these links returned to average co-authorship strength (PAI index around 1), probably explained by the overall increase of publication in the field. Portugal's international co-authorship has increased almost threefold compared to the average European country's growth of 85%. These differences should be considered with caution however, since the total number of publications in life sciences is low in both Portugal and Greece, making it easy to overestimate the importance of a few additional papers when comparing growth rates.

The average co-authorship indices, with EU member states, have increased, albeit non-significantly ($p > 0.05$), for all countries except Greece, Portugal and Ireland. Figures 5 and 6 also indicate that smaller countries and new member states are more likely to have preferential co-authorship links with other EU18 countries, illustrated by the centrality of these countries in the network.

The results suggest that preferential co-authorships exist and can in most cases be explained by geographical proximity. Co-publication rates between more recent EU member-state countries from the central/eastern part of Europe (Czech Republic, Hungary, Poland and Slovakia) have increased over the examined period and these countries constitute a cluster. Similarly, a Scandinavian cluster can be identified. In the next step we are investigating the level of multilateral vs. bilateral co-authorship between European countries. We are also studying whether preferential co-authorships are carried out as multi- or bilateral co-authorship. The relative share of bilateral collaborations (of the total co-authorship) between countries is shown in Figures 7 and 8.

The majority of preferential co-authorships have a higher degree of bilateral co-authorship in comparison to non-preferential collaborations.

In contrast, preferential co-authorship related to eastern European countries (Czech Republic, Hungary, Poland, and Slovakia) are carried out on a multilateral level. In cases where preferential co-authorship cannot be explained by geographical proximity and the number of co-authored papers is low, for example, Greece–Slovakia, Greece–Portugal, Ireland–Portugal, co-authorship is mainly carried out on a multilateral level. These collaborations could be a result of 'top-down' initiatives such as the FPs or large-scale 'bottom-up' initiatives.

Preferential collaborations do also exist in FPs; see Appendix 1, even though the distinctions between country-pairs are not as strong as in co-publications. The FP collaboration networks are denser than the co-publication ones because they involve more countries on average than co-publications. Geographical proximity has an influential role in many of the preferential collaborations. Bilateral projects, involving only two countries, are rare. Where they do exist they are to a great extent collaborations between preferred partners as identified by co-authorship, and collaboration of countries between which bilateral agreements exist.

Bilateral projects, involving only two countries, are rare. Where they do exist they are to a great extent collaborations between preferred partners as identified by co-authorship, and collaboration of countries between which bilateral agreements exist

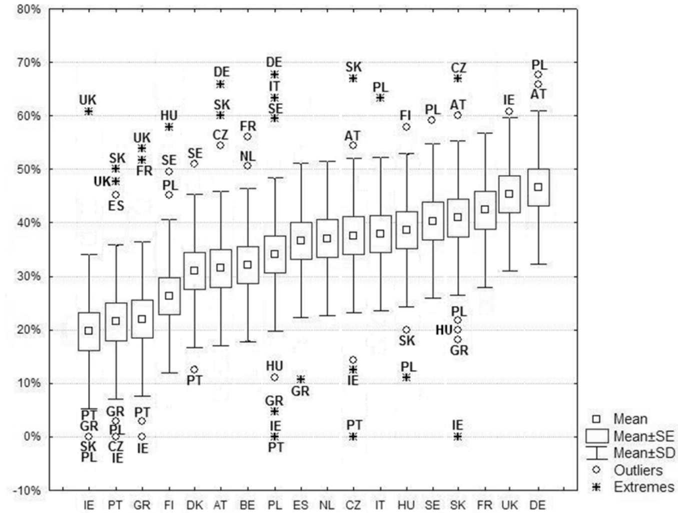

Figure 7. EU member-state bilateral share of total number of EU co-authored papers during 1995–1997

Notes: * and o depict extremes and outliers, respectively
For definitions, see analytical framework

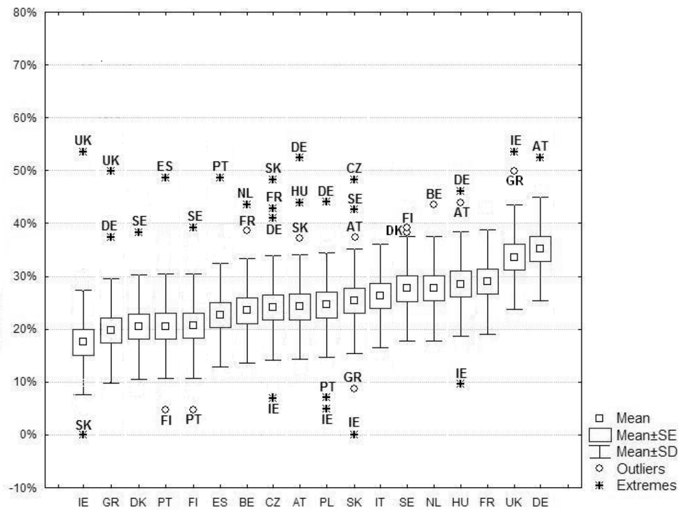

Figure 8. EU member-state bilateral share of total number of EU co-authored papers during 2003–2005

Note: For definitions of extremes and outliers, see analytical framework

Concluding discussion

This paper contributes to the understanding of international research collaborations and the interplay with funding mechanisms. This was done by studying the co-publication patterns and EC FP participation in the field of life sciences of 18 EU member states. We found that European co-authored papers were more likely to be the product of collaborations on a bilateral level, involving two countries. One possible explanation for this pattern is that the bulk of funding for international co-operation is coming from national sources, which are more likely to support bilateral agreements rather than large multilateral international collaborations.

At the same time, multilateral co-authorship, involving more than two countries, has grown more strongly, albeit from a smaller base, than bilateral co-authorship. Whether this growth is a direct result of existing ‘top-down’ funding programmes, such as the FPs, is difficult to say since most research is carried out with funding from both national and international sources. Overall, it can be concluded that science is becoming more international, drawing on competences from different national sources.

A more specific comparison of these collaboration types with regard to the involvement of non-European countries also provides interesting insights into the possible influence of funding and the differences between modes of collaboration networks. In Europe, a number of ‘top-down’ funding initiatives have been put in place in support of the ERA, with the aim of creating an ‘integrated knowledge creation network’. The existing funding (but not participation in general) is restricted to European member states, accessing and candidate countries. These programmes often require involvement from several countries, sometimes preferably with new member-state or candidate countries. These funding requirements are found to exert a stronger influence on related partnerships than the principal openness of FP-supported networks to non-European partners. Examining the research partnerships in life science projects funded under the FPs, we find that the involvement of non-European countries has been very limited.

In contrast, the analysis of co-authorship data showed the opposite. Here, researchers are acting in favour of *globalised*, rather than purely European multilateral co-authorship. These findings support the argument by Wagner and Leydesdorff (2005) that growth in international collaborations can be explained by the self interest of individual researchers rather than by structural, institutional or policy-related factors.

In addition it could be argued that the share of papers resulting from FPs supporting projects is too low to influence the overall publication pattern. Similarly, FP projects may result in other types of joint productions such as exchange of material and data, development of methods, databases and software, as well as the setting-up of conferences, and mobility programmes (Georghiou, 1998). For an exact measurement of the outputs of FPs you would have to study individual projects but also this has its limitations since nearly all research today is performed with multiple funding. There are very few research groups where funding comes from only one source.

Our research also shows that geographical proximity plays an important role in selection of collaboration partners. The persistence of the other factors identified is evident among preferential co-authorship between countries and can first and foremost be explained by geographical proximity. The results confirm that proximity has an influential role in both FPs and co-authorship. Clusters and preferential partners do still exist, both in co-authorship as well as in FPs. Similar studies in the USA, using mileage indicators between universities (Adams *et al*, 2005), confirm that distance matters, everything else being equal (language, culture, and policy-setting). Research feeds on interdisciplinarity and different experiences, which are heavily related to culture and national borders. Instead of trying to make the ERA fully integrated by design, programmes could be more accommodating to existing differences and experiences and to self-selected collaborations.

Looking at patterns of co-authorship, it is clear that researchers want to have access to global knowledge and global networks. Although the EC highlighted the importance of a greater concerted S&T cooperation with the rest of the world in its vision for ERA and although it proposed a more proactive policy to broaden access to world-wide knowledge, this global outlook is not visible in existing FP-supported collaboration networks. Maybe it is time for existing European programmes to open up more proactively to non-European partners and also to provide financial support for global collaborations.

As a qualifying factor it should be kept in mind that we have studied only the pattern of life science-related articles and initiatives and that the pattern of co-authorship varies widely among scientific fields. In physics the rate of international co-authorship is very high, mostly due to the sharing of expensive large-scale infrastructure and time-consuming experiments involving many researchers. A number of global, successful initiatives have served as inspiration for the setting-up of other scientific organisations, such as the European Organisation for Nuclear Research (CERN) and the European Southern Observatory. Therefore the patterns that have been observed in life sciences cannot be directly transferred to other fields.

Notes

1. Identified as a system with internal coherence, that has identifiable boundaries and interacts as a distinct entity with other bodies.
2. FP4 BIOMED2 (research into major illnesses, pharmaceuticals research, research into biomedical technology and engineering, brain research, human genome research, public-health research) and BIOTECH2 (cell factories, genome analysis, plant and animal biotechnology, cell communication in neurosciences, immunology and trans-disease vaccinology, structural biology, prenormative research , biodiversity and social acceptance, infrastructures), in FP5 (food, nutrition and health, control of infectious diseases, the ‘cell factory’, environment and health, sustainable agriculture, fisheries and forestry, generic activities), and in FP6 (fundamental genomics, applied genomics and biotechnology, genomic approaches to health and disease, cancer, HIV/AIDS, malaria and tuberculosis).
3. FP4 BIOMED2 and BIOTECH2, in FP5 QUALITY OF LIFE, and in FP6 life sciences, genomics and biotechnology for health.

References

Adams, J D, G C Black, J R Clemmons and P E Stephan 2005. Scientific teams and institutional collaborations: evidence from U.S. universities 1981-1999. *Research Policy*, **34**(3), 259–285.

Appendix 1. Preferential collaborations in Framework Programmes

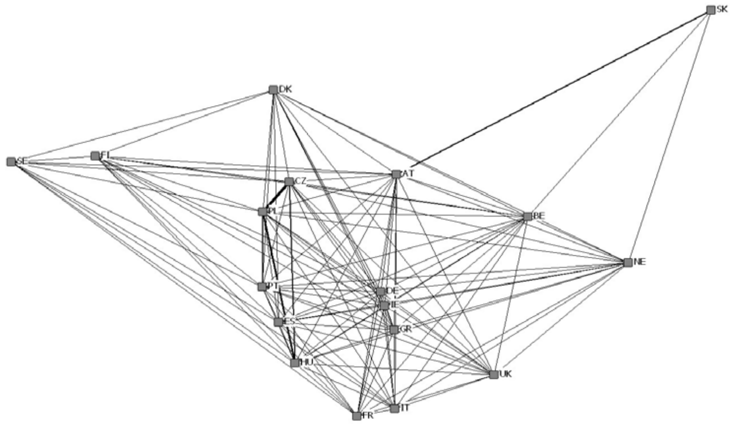

Figure A1. Project collaborations between countries in Framework Programme 4

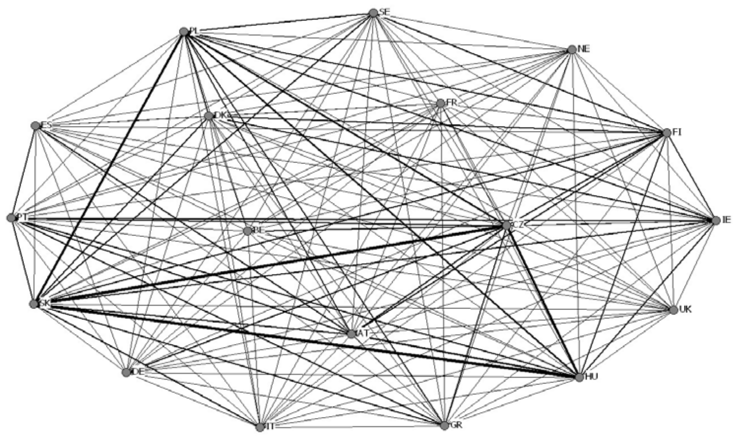

Figure A2. Project collaborations between countries in Framework Programme 5

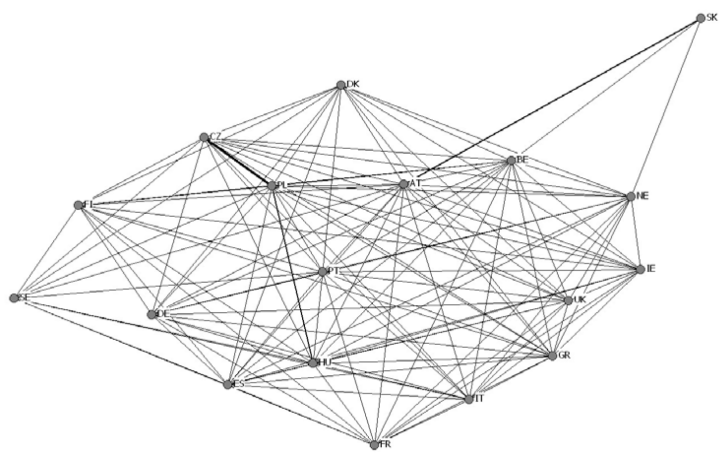

Figure A3. Project collaborations between countries in Framework Programme 6

Arnold, E, T Astrom, P Boekholt, N Brown, B Good, R Holmberg *et al* 2008. *Impact of the Framework Programme in Sweden.* Stockholm: VINNOVA.

Bammer, G 2008. Enhancing research collaborations: three key management challenges. *Research Policy*, **37**(5), 875–887.

Banchoff, T 2002. Institutions, inertia and European Union research policy. *Journal of Common Market Studies*, **40**(1), 1–21.

Bozeman, B and C Boardman 2003. Managing the New Multipurpose, Multidiscipline University Research Centers: Institutional Innovation in the Academic Community. IBM Center for the Business of Government.

Breschi, S and L Cusmano 2004. Unveiling the texture of a European research area: emergence of oligarchic networks under the EU framework programmes. *International Journal of Technology Management*, **27**(8), 747–772.

EC, European Commission 2000. *Towards a European Research Area*, Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions. Brussels, COM (2000) 6.

Eurostat yearbook 2009.

Frenken, K 2002. Europeanisation of science. *Tijdschrift voor Economische en Sociale Geografie*, **93**(5), 563–570.

Frenken, K, J Hoekman and F van Oort 2007. *Towards a European Research Area*. Rotterdam: NAi Publishers.

Georghiou, L 1998. Global cooperation in research. *Research Policy*, **27**, 611–626.

Glänzel, W 2001. National characteristics in international scientific co-authorship relations. *Scientometrics*, **51**(1), 69–115.

Greunz, L 2003. Geographically and technologically mediated knowledge spillovers between European regions. *The Annals of Regional Science*, **37**(4), 657–680.

Jaffe, A B 1989. Real effects of academic research. *American Economic Review*, **79**(5), 957–970.

Katz, J S 1994. Geographical proximity and scientific collaboration. *Scientometrics*, **31**(1), 31–43.

Leydesdorff, L 2000. Is the European Union becoming a single publication system? *Scientometrics*, **47**(2), 265–280.

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

Lundvall, B-Å, B Johnson, E S Andersen and B Dalum 2002. National systems of production, innovation and competence building. *Research Policy*, **31**(2), 213–231.

Luukkonen, T 2002. Technology and market orientation in company participation in the EU framework programme. *Research Policy*, **31**, 437–455.

Mattsson, P, P Laget, A Nilsson and C J Sundberg 2008. Intra-EU vs. extra-EU scientific co-publication patterns in EU. *Scientometrics*, **75**(3), 555–574.

NSF, National Science Foundation 2007. Asia's Rising Science and Technology Strength: Comparative Indicators for Asia, the European Union, and the United States. NSF 07-319. Arlington, VA.

Okubo, Y and M Zitt 2004. Searching for research integration across Europe: a closer look at international and inter-regional collaboration in France. *Science and Public Policy*, **31**(3), June, 213-226.

Pond, R, F van Oort and K Frenken 2007. The geographical and institutional proximity of research collaboration. *Papers in Regional Science*, **86**(3), 423–443.

Rigby, J and J Edler 2005. Peering inside research networks: some observations on the effect of the intensity of collaboration on the variability of research quality. *Research Policy*, **34**, 784–794.

Roediger-Schulga, T and M J Barber 2007. *R&D Collaboration Networks in the European Framework Programmes: Data Processing, Network Construction and Selected Results.* Maastricht: United Nations University.

Scherngell, T and M J Barber 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.

Stein, J A 2004. Is there a European knowledge system? *Science and Public Policy*, **31**(6), December, 435–447.

Sörlin, S 2004. Europas idehistoria. 1492-1918. *Natur och Kultur.* Sverige.

Wagner, C S 2002. The elusive partnership: science and foreign policy. *Science and Public Policy*, **29**(6), December, 409–417.

Wagner, C S 2006. International collaboration in science and technology: promises and pitfalls. In *Science and Technology Policy for Development: Dialogues at the Interface*, eds L Box and R Engelhard, pp 165–177. London: Anthem Press.

Wagner, C S and L Leydesdorff 2005. Network structure, self-organization and the growth of international collaboration in science. *Research Policy*, **34**(10), 1608–1618.

Widhalm, C, M Topolnik, A Kopcsa, E Schiebel and M Weber 2001. Evaluating patterns of co-operation: application of a bibliometric visualisation tool to the Fourth Framework Programme and the Transport Research Programme. *Research Evaluation*, **10**(2), August, 129–140.

Zitt, M and N Teixeira 1996. Science macro-indicators: some aspects of OST experience. *Scientometrics*, **35**(2), 209–222.

Zitt, M, R Barre, A Sigogneau and F Laville 1999. Territorial concentration and evolution of science and technology activities in the European Union: a descriptive analysis. *Research Policy*, **28**, 545–562.

III

Scientometrics
DOI 10.1007/s11192-010-0310-9

Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position

Pauline Mattsson · Carl Johan Sundberg · Patrice Laget

Received: 21 July 2010

Abstract Bibliometric indicators are increasingly used to fund and evaluate scientific research. Since the number of authors in a paper and the number of has increased it is difficult to determine the individual contribution of authors. Suggested approaches include the study of author position or the corresponding author. Our findings show that the corresponding author is most likely to appear first and then last in the byline. The results are dependent on number of authors in a paper and national differences exist. This underscores the need to take into account both the number of authors on a paper and their position in the byline to be accurate when measuring author contribution.

Keywords Co-authorship · Corresponding author · Author position · Author impact

Introduction

Bibliometric indicators have gained increasing importance and use in science policy and science management where it has been frequently used in the domain of research evaluation and assessment (Moed 2009). The development of performance indicators to respond to science policy questions and as a tool for distribution of resources has been the most common applications (Adam 2002). When individual researchers are to be assessed, publications in international peer-reviewed journals have been an important indication of visibility and international acknowledgment. Since the number of authors per paper has increased and the share of publications with multiple institutions grew from 40 to 61% between 1988 and 2005 (NSF 2008), it has become more difficult to determine the individual contribution to the work performed. Suggested approaches include the study of author position (Rennie et al. 1997) or the corresponding authorship (Wren et al. 2007).

P. Mattsson (✉) · C. J. Sundberg
Unit for Bioentrepreneurship, LIME, Karolinska Institutet, 171 77 Stockholm, Sweden
e-mail: pauline.mattsson@ki.se

P. Laget
Inserm Unit 646, 10, rue Andre Boquel, 49100 Angers, France

Earlier research has with qualitative methods attempted to determine the degree of contribution in relation to author position. The main notion has been that only first and last authorship have any significant meaning (Rennie et al. 1997; Burman 1982; Reisenberg and Lundberg 1990; Drenth 1998; Kennedy 2003). The first author is usually the person that has taken the main responsibility and carried out most of the work in a project (Reisenberg and Lundberg 1990).

In 1991, the International Committee of Medical Editors (ICMJE, also known as the Vancouver group) decided on a number of authorship criteria that should be met in submitted manuscript. Each author should have participated sufficiently in the work to take public responsibility for the content. Authorship credit should be based on (i) conception and design, analysis and interpretation of data (ii) the drafting or reviewing of the article (iii) final approval of the version to be published. There are no clear guidelines made about the order of authorship only that it should be a joint decision by the involved co-authors. These guidelines have been incorporated into the "uniform requirements for the submission of manuscripts to biomedical journals" to which almost 700 journals have subscribed according to ICMJEs website (ICMJE 1997).

The majority of journals require that each article should identify one of the authors as the corresponding author. The intention is that readers should be able to comment and ask questions about the published article. The role of the corresponding author is therefore to respond to these questions but also to be responsible for correspondence with the journal before acceptance and publication. The corresponding author should also be able to declare any competing or conflicting interest and to explain the presence and order of co-authors (ICMJE 1997).

Against this background we further investigated the meaning of author position in internationally co-authored papers. With quantitative methods we studied the relation between corresponding author and byline position and whether the number of authors assigned to a publication makes a difference. We also tested for differences among countries.

Methods

The data used in this study was retrieved from the ISI Web of Knowledge of Thomson Reuters. Only papers published as notes and articles in 2003 were taken into account. Scientific field included Fundamental Biology and Medicine, according to the classification done by Observatoire des Sciences et Techniques (OST) (Zitt and Teixeira 1996). These fields cover the journals subscribing to the ICMJEs guidelines. A selection of the most productive 18 European countries, accounting for 99% of the total EU-25 publication production, were made ($n = 39272$ articles). A difference between international co-authored papers, defined as articles with author addresses from more than one country, and national co-authored papers, defined as articles with author address from only one country, was made (Table 1).

Information such as author name, author position, and author address were collected. In the address field of SCI the corresponding author is labelled reprint author, in this paper referred to as corresponding author. Before 1998, less than 60% of the publications had this tag while from 1998 and onwards on average 98% include the reprint label.

Finally, the first author's institutional origin was manually compared to the last author's using a probability sampling method ($\alpha = 0.05$), by random numbers, to select a representative sample of 1000 articles proportional for the selected countries.

Table 1 Countries co-publications included in the study

Abbreviation	Country
AT	Austria
BE	Belgium
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
NL	Netherlands
PL	Poland
PT	Portugal
SE	Sweden
SK	Slovakia
UK	United Kingdom

We analysed the data using the software programme Statistica. To examine if there are any differences between countries we used the Wilcoxon Matched Pair Test.

Results

Overall, when the number of authors in an article where more than two, the first author accounted for the majority (52%) ($n = 20356$) of corresponding authors, of all internationally co-authored papers. The last author appeared as corresponding author in 39% ($n = 15470$). Only 9% ($n = 3446$) of internationally co-authored papers have a corresponding author positioned elsewhere in the byline.

Next we tested whether the number of authors in a publication had an impact on the position of the corresponding author. Corresponding author appearing as first author in the byline was more common when the number of authors was less than seven. When the number of authors are more than seven there was no difference between first and last appearing as corresponding author, see 95% confidence interval in Fig. 1.

The frequency of first versus last author having the same address was examined using a sub-sample of 1000 articles. We found that in 54% of the selected articles first and last author had the same address. When the last author is the corresponding author 73% of the articles have the same first versus last author addresses. In the case of first author being the corresponding author the equivalent frequency is 42%.

We also explored whether any differences between countries could be discerned due to different scientific publication cultures. If we compare the byline position of corresponding author between countries we find that there was a significant difference (Wilcoxon Matched Pair Test: $p > 0.001$) between countries. When researchers based in Scandinavia, the Netherlands, and former Eastern countries were corresponding authors it was more common that the corresponding author appears first in the byline (Fig. 2).

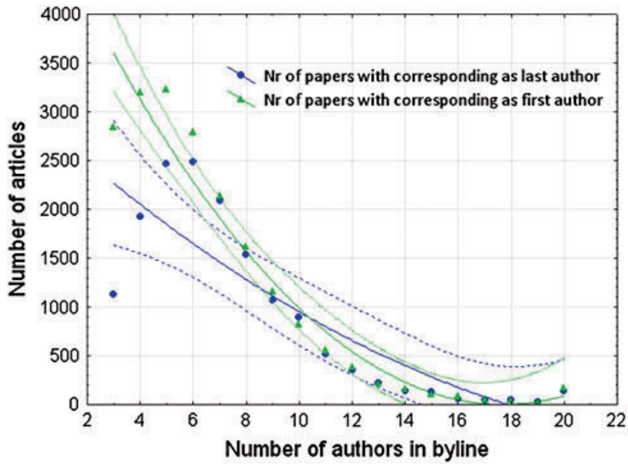

Fig. 1 Distribution of papers by position of corresponding author

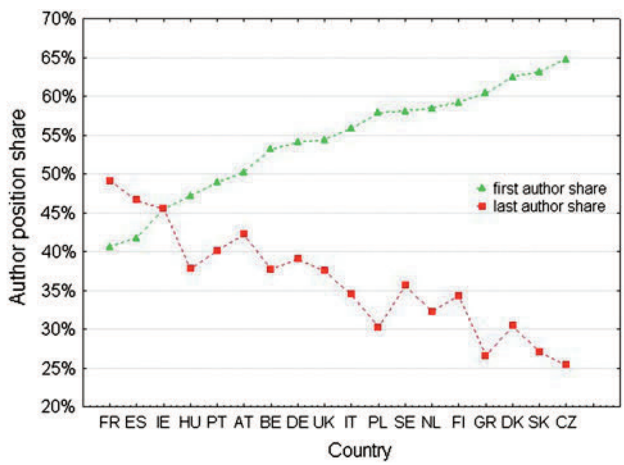

Fig. 2 Byline position in international co-authored articles, differences between countries

A similar pattern as for international co-authorship can be discerned in national co-authored papers. For France and Spain it is more common that corresponding author appears last in the byline for international co-authored papers while for national co-authored papers corresponding author, more commonly, appears first in the byline. On average, in national co-authored papers the corresponding author appears more often as a first in the byline (60%) than in international co-authorship (54%) (Fig. 3).

Discussion

The findings in this study provide evidence that theories stating that (1) corresponding author is the author contributing more to the article (Wren et al. 2007) and (2) first and last

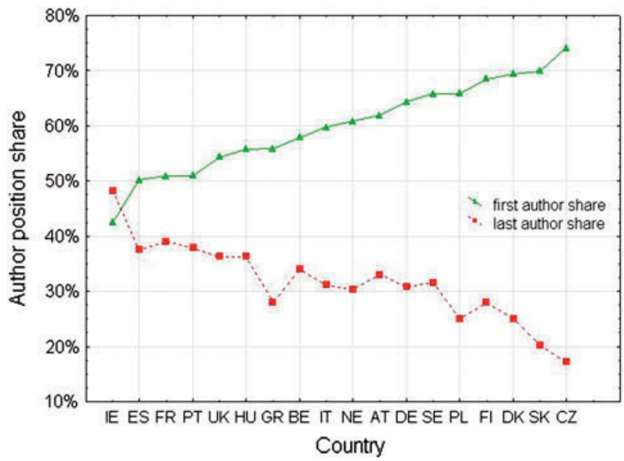

Fig. 3 Byline position in national co-authored articles, differences between countries

authorship have any significant meaning (Rennie et al. 1997; Burman 1982; Reisenberg and Lundberg 1990; Drenth 1998; Kennedy 2003) are complementary. This argument is supported by the result that the majority of corresponding authors appear as either first or last authors. We found that corresponding authors were more likely to appear first in the byline, which is in accordance with ICMJEs guidelines, i.e. the author contributing most to the article should be positioned first.

By tradition, in the cases were the first author is a junior researcher the final author is traditionally the mentor/supervisor (Rennie et al. 1997; Bhopal et al. 1997; Burman 1982; Reisenberg and Lundberg 1990). This has been supported by a study showing that the share of last authors with a professor rank is significantly higher than among first authors even though the number of professors appearing as first authors has increased (Drenth 1998). The last author position can also be reserved for the director of the laboratory or department, who played no direct role in the specific research, so-called gift or honorary authorship (Bhopal et al. 1997; Burman 1982). The origin of the tradition of the senior author occupying the final author position is believed to arise from the obligation of senior scientists' to recognise younger colleagues (Rennie et al. 1997; Shapiro et al. 1994). Whatever its origin, it has become a strong tradition in scientific publication (Buehring et al. 2007). The sequence of authors may also reflect the contribution of work with first author contributing most and last author contributing the least (Gaeta 1999).

This traditional way of structuring a paper could partly be supported since the majority of an article's first and last authors have the same address. In the case of last author being the corresponding author the likelihood of first and last authors having the same address was significantly higher than when the first author appeared as corresponding. A qualitative study investigating the seniority of authors would be needed to examine this issue further.

According to earlier research (Yank and Rennie 1999) the first author was twice as likely as the other contributors to have coordinated the study. The first and last contributors on the byline were also more likely to have written the paper, designed the study, and analyzed the data. In a survey (Wren et al. 2007) carried out with promotion and tenure committee chair people the aim was to assess how they value author contributions (according to initial conception; work performed; and supervision) determined by author position on a hypothetical manuscript where the correspondent author appeared as either last or middle author. The results indicate that respondents considered that the first author in a three-person byline had made the greatest contribution to the work performed (57%), whereas the last author deserved most credit for both the initial conception (49%) and supervision (54%) of the project. When the last author appear as the corresponding author, there was no significant difference in three-author compared with five-author bylines for the credit apportioned to the last author for initial conception, work performed or supervision. By contrast, the first author's relative contributions decreased significantly for initial conception and for work performed but not for supervision. When the correspondent author in a five-author byline paper was changed from last to middle author, the latter received increased credit for initial conception (34% versus 6%), work performed (20% versus 11%) and supervision (33% versus 7%), and the last author's overall credit decreased from 38 to 16%. This indicated that the corresponding author is perceived as the author contributing more to the article independently of the author position.

The results in this study combined with earlier research suggest that candidates for promotion or tenure would be well advised to highlight publications on which they acted as corresponding author, especially if they were not the first or last author. Our observation that the first and last author account for the majority of correspondence suggest that some of the recent efforts to develop bibliometric formulae to measure author impact (Ball 2005; Gomez-Alonso 2004) would need to take into account both the number of authors on a paper and their position in the byline to be accurate.

Different national scientific cultures seemed to influence the corresponding author position. A call for a shift in national publication culture towards common international standards could make the research better visible and comparable. Secondly an increasing harmonization would support the development of infrastructure such as databases and methods that can be used for observing and analysing scientific publications.

Limitations of the study

This study has focused on articles in the fields of Life Sciences. It is important to remember that the order of authorship could be different in other fields such as for example Physics. The results should therefore not be transferable to other fields.

The study has only investigated a number of countries. The results could therefore vary depending on the countries studied. The authors are addressing this issue and the role of countries in an ongoing study.

References

Adam, D. (2002). The counting house. *Nature, 415*, 726–729.

Ball, P. (2005). Index aims for fair ranking of scientists. *Nature, 436*, 900.

Bhopal, R., Rankin, J., McColl, E., Thomas, L., Kaner, E., Stacy, R., et al. (1997). The vexed question of authorship: Views of researchers in a British medical faculty. *BMJ, 314*, 1009–1012.

Buehring, G. C., Buehring, J. E., & Gerard, P. D. (2007). Lost in citation: Vanishing visibility of senior authors. *Scientometrics, 72*(3), 459–468.

Burman, K. D. (1982). Hanging from the masterhead—reflection on authorship. *Annals of Internal Medicine, 97*, 602–605.

Drenth, J. P. (1998). Multiple authorship: The contribution of senior authors. *JAMA, 280*, 219–221.

Gaeta, T. J. (1999). Authorship: "law" and order. *Academic Emergency Medicine, 6*(4), 297–301.

Gomez-Alonso, J. (2004). Author! Author! *JAMA, 292*, 1815.

ICMJE. (1997). Uniform requirements for manuscripts submitted to biomedical journals. *JAMA, 277*, 927–934.

Kennedy, D. (2003). Multiple authors, multiple problems. *Science, 301*, 733.

Moed, H. F. (2009). UK research assessment exercises: Informed judgments on research quality or quantity? *Scientometrics, 74*, 153–161.

National Science Board. (2008). *Science and engineering indicators*. Arlington, VA: National Science Foundation.

Reisenberg, D., & Lundberg, G. D. (1990). The order of authorship: Who's on first? *JAMA, 264*, 1857.

Rennie, D., Yank, V., & Emanuel, L. (1997). When authorship fails—a proposal to make contributors accountable. *JAMA, 278*, 579–585.

Shapiro, D. W., Wenger, N. S., & Shapiro, M. F. (1994). The contributions of authors to multiauthored biomedical research papers. *JAMA, 271*, 438–442.

Wren, J. D., Kozak, K. Z., Johnson, K. R., Deakyne, S. J., Schilling, L. M., & Dellavalle, R. P. (2007). The write position. *EMBO reports, 8*(11):988–991.

Yank, V., & Rennie, D. (1999). Disclosure of researcher contributions: A study of original research articles in the lancet. *Annals of Internal Medicine, 130*, 661–670.

Zitt, M., & Teixeira, N. (1996). Science macro-indicators: Some aspects of OST experience. *Scientometrics, 35*(2), 209–222.

IV

Jointly published by Akadémiai Kiadó, Budapest
and Springer, Dordrecht

Scientometrics, Vol. 80, No. 2 (2009) 441–455
DOI: 10.1007/s11192-008-2081-0

What do patent examiner inserted citations indicate for a region with low absorptive capacity?

JOAQUÍN M. AZAGRA-CARO,[a,b] IGNACIO FERNÁNDEZ-DE-LUCIO,[b]
FRANÇOIS PERRUCHAS,[b] PAULINE MATTSSON[a,c]

[a] *European Commission – Joint Research Centre, Institute for Prospective Technological Studies (IPTS)*
Edificio Expo, C/Inca Garcilaso, s/n, E-41092 Sevilla, Spain
[b] *INGENIO, CSIC-UPV, Valencia, Spain*
[c] *UBE-LIME, Karolinska Institutet , Stockholm, Sweden*

Most studies of patents citations focus on national or international contexts, especially contexts of high absorptive capacity, and employ examiner citations. We argue that results can vary if we take the region as the context of analysis, especially if it is a region with low absorptive capacity, and if we study applicant citations and examiner-inserted citations separately. Using a sample from the Valencian Community (Spain), we conclude that (i) the use of examiner-inserted citations as a proxy for applicant citations, (ii) the interpretation of non-patent references as indicators of science-industry links, and (iii) the traditional results for geographical localization are not generalizable to all regions with low absorptive capacity.

Introduction

Citations in patents can be seen as knowledge footprints and can be used to trace the information sources on which the invention is built. Further, they can illustrate the relations with other inventions e.g. geographic, sectoral and technological linkages. Many existing citation studies use citing-cited patents to analyse knowledge flows from company to company, or from other sectors, e.g. research institutes and academia to companies [MEYER, 2002, LEYDESDORFF & MEYER, 2003].

When a researcher or a company applies for a patent, it has to provide a description of the invention, proving its novelty and utility. Many applications include 'prior art' in the form of previous inventions or other relevant scientific information, information aimed at defining the differences from existing patents. A patent that includes reference to prior art, from hereon referred to as a citation, within a specific document can be said to build on the knowledge in the document(s) it cites.

Existing studies mainly analyse citations from the perspective of what we might call the 'research-intensive environment', i.e. leading world zones or countries, or high technologies, such as nanotechnology, pharmaceuticals, chemicals, etc. In this study, we focus on the regional dimension, which is a crucial unit of observation in terms of its

Received April 25, 2008; Published online April 16, 2009

Address for correspondence:
JOAQUÍN M. AZAGRA-CARO
E-mail: joaquin-maria.azagra-caro@ec.europa.eu

0138–9130/US $ 20.00

capacity to implement science and technology policies and embed an idiosyncratic culture [COOKE, 1992]. Consideration of the regional dimension allows us to question some of the findings from earlier studies.

When the country is the unit of analysis, the results reflect the aggregated outputs of regions. Some regions may display a higher propensity than others to include citations. The aggregated result will thus tend to reflect the properties of these regions. What kinds of regions are likely to include greater numbers of citations? Our assumption is that regions with high absorptive capacity are more likely to cite,[1] because one of the characteristics of production of goods and services in such regions is that it relies on a larger explicit knowledge base, which is relatively easy for firms to identify and absorb.

Therefore, when we examine patent citations at national or international level, what we are really observing is the pattern in regions with high absorptive capacity. Would the pattern be the same in regions with low absorptive capacity?

Similarly, the results from studies of advanced technologies may not be the same as the results from studies of low technology levels, or regions where low levels of technology predominate, i.e. regions with low absorptive capacity. Thus our interest in focusing on the latter type of region. Whereas the results from studies of high-tech regions are coherent, this provides the motivation to conduct similar studies on regions with low absorptive capacity in order to understand the relevance of citations as an indicator of science-industry relations.

The first concern in this paper is on the regional dimension of patent citations; the second is on the incipient, but increasing debate on the differences between applicant and examiner citations in patents, and we will argue that it is intrinsically linked to the interpretation of findings at regional level.

The basis of this debate is that citations can be included by the 1) inventor, 2) the legal expert, 3) the applicant, and 4) patent examiner. In studies based on patent citations, the assumption is made that all citations were included by the inventor, and thus they are an indicator of the knowledge that he/she possesses. However, in the course of a patent application, citations can be added first by the applicant, who is not necessarily the inventor, but might be the company that owns the property rights and/or second by the applicant's legal advisers. It is not possible to distinguish among these citations.[2] These citations are reviewed by the patent examiner to determine the novelty of the invention and to decide whether to accept or reject the application. The examiner compares the invention with the prior art and can insert on the front page of the patent application, what in his/her opinion are other relevant references. Of course, an applicant can choose to leave out some of the knowledge base to make the invention

[1] We follow COHEN & LEVINTHAL's [1990] definition of absorptive capacity: "a limit to the rate or quantity of scientific or technological information that firm can absorb". To justify the extension of the concept of absorptive capacity from firms to regions, see NIOSI & BELLON [2002] and AZAGRA & AL. [2006].

[2] This is why in this paper we use the term 'applicant citations' rather than the more common, but imprecise 'inventor citations'.

appear more novel. It is the examiners' responsibility to identify whether this has been done and to add relevant information and remove non-relevant citations. Therefore, the references on the front pages of patent applications are not necessarily the same as in the original application. In a study carried out by SCHMOCH [1993] only 8.4% of all the citations included on the front page of EPO originated from the inventor.

In the next section we describe how the literature on patent citations has given rise to frequently reproduced findings and how the debate on the differences between applicant and examiner citation makes sense in a regional setting, and especially regions with low absorptive capacity. Some hypotheses are proposed. The third section describes the context in which they are tested – Spain's Valencian Community – and the methodology employed. The presentation of the results follows and the last section provides some conclusions.

What we can learn from patent citations at regional level, in a low-tech context, based on full-text citations and a stronger European focus

Citations as an indicator of the inventor's knowledge base: applicant vs examiner citations

Patent citations have been criticized by several scholars as being 'noisy' when used to measure knowledge transfer, with the examiner added citations producing most of the 'noise'. In a study by ALCÁCER & GITTELMAN [2006], however, these criticisms were rejected; the authors concluded that the changes implemented by examiners are numerous and non-random. JAFFE & AL. [2000] differentiated between applicant and examiner added citations and found that one-third of applications included a high level of the knowledge from cited works, one-third learned more about related knowledge during completion of the application and one-third had no prior knowledge about the cited information. Both these studies indicate that examiners add an important proportion of the citations included, and that aggregated citations can be a misleading measure of inventor knowledge transfer. We prefer to differentiate between applicant and examiner citations and to analyse them separately. ALCÁCER & GITTELMAN [2006] tested whether applicants and examiners cite similar patents, analysing forward citations in highly cited patents. They found that inventors are most likely to cite highly cited patents but that examiners also cite a large group. However both inventors and examiners forward citations, converge over time.

When comparing citation patterns among patents issued in Europe/Japan compared to the US, care is needed in the interpretation of results. Since 2001, the applicant and/or its attorney for a US Patents and Trademark Office (USPTO) patent have a 'duty of candour', which implies that they must, by law, include any prior art of an invention.

This is not the case in Europe and Japan. Applicants to the USPTO, therefore, tend not to limit their references, but rather to include citations that are in any way relevant to their invention in order to minimize the risk of their application being rejected. Therefore, USPTO patents on average include more citations than Japanese or European ones [ALCÁCER & GITTELMAN, 2006]. In a study carried out by MICHEL & BETTELS [2001], American patents include three times as many patent references and half as many non-patent references as the EPO patents. THOMPSON [2006] found, using USPTO data, that examiners added 41% of all citations and that examiner added citations accounted for all of the citations made by 38% of the citing patents, compared to 8.5% in the case of applicants. In applications to the European Patent Office (EPO), examiners rather than applicants add the majority of citations [CRISCUOLO & VERSPAGEN, 2005; ALCÁCER & GITTELMAN, 2006]. MEYER [2000] discusses other differences between the US and European systems, such as the generally lower education levels of US examiners, which results in the inclusion of citations that are not always directly relevant to the patent. In addition, the heavy workload of US examiners and the focus on English language documents could have a negative impact on the search for relevant documents. EPO patents' citations are more focused towards a limited number of fields and therefore EPO applicant citations might be closer to real knowledge transfer than USPTO applications. This has further been supported by MICHEL & BETTELS [2001] that argue that the general coverage of underlying technology in the US search report is much broader than in the EP. These differences should be taken account of in choosing the most suitable data set.

In terms of analysing the differences between applicant and examiner citations, there are only a few existing studies, and, to the best of our knowledge, only one for the European case, i.e. CRISCUOLO & VERSPAGEN [2005]. There is a need for more empirical evidence, especially for Europe where the obligation to reveal information is not as compelling as in the US; CRISCUOLO & VERSPAGEN [2005] found a much lower share of applicant citations over all front-page citations than in the US (9% vs 63%). Also, all these studies analyse only front-page citations, so the applicant citations are only those that patent examiners consider relevant. The patent examiner's criteria for applicant citation relevance may not reflect actual knowledge flows and we would argue that full-text citations are a better indicator. ACOSTA & CORONADO [2003] justify this emphasis on full-text citations to avoid possible under-estimation of science-technology links, but they do not compare them with front-page citations. In a related paper, ACOSTA & CORONADO [2002] show that only 31% of Spanish patents include applicant full-text citations of patent references (PRs) and 10% of Spanish patents include applicant full-text citations of non-patent references (NPRs), i.e. a small number compared to front-page citations, which appear in all patents.

Hence, we would make another plea for the inclusion of the regional dimension and the notion of absorptive capacity in the analysis of citations in patents. If we study a region with low absorptive capacity, we can expect a relatively low number of applicant citations in patents because of the argument previously applied to NPRs – citations require qualified human capital able to decode the information embedded in another source, not just academic publications but also other firm patents. If there are fewer applicant citations in EPO patents and in full-text citations, then in any study of a region with low absorptive capacity in Europe through full-text citations, there will necessarily be a very small number of applicant citations.

Hypothesis 1. Patent examiner's citations in patents are a weak indicator of an applicant's knowledge flows, in regions with low absorptive capacity.

Non-patent references are useful to trace science-industry links

A patent can contain both PRs and NPRs. The latter can include books, articles in scientific journals, newspaper articles, etc. CALLAERT & AL. [2006] and HARHOFF & AL. [1999] found that in most cases, NPRs are references to scientific journal articles. NPRs in the form of scientific references in patents, have been argued to reveal a direct influence of science on technology [NARIN & AL., 1997], while MEYER [2000] does not claim such a strong link. NARIN & NOMA [1985], in an investigation of the science-technology links, introduced the concept of NPRs. While in the literature, PRs are considered to be indicators of the technological knowledge embodied in patents, NPRs are mainly seen as indicating scientific proximity and a measure of the knowledge flow between science and technology. However, this link is dependent on the geographic context, e.g. CALLAERT & AL. [2006] found that EPO patents are more likely to include NPRs, suggesting that European patents have a closer linkage with science.

SAMPAT [2004] studied the differences between applicants' and examiners' tendencies to cite patents vs NPRs. He found that examiners were less likely to include NPRs than applicants. When only patents including NPRs were considered the majority of the citations came from the applicants. The explanation given was that applicants have a better state-of-the-art knowledge and keep track of developments in the field.

So, are NPRs a useful indicator of science-industry links in a region with low absorptive capacity? We should take into account the fact that science-industry links take a variety of forms, and have different costs, e.g. having a R&D department or engaging in a research cooperation agreement tends to be more costly for a company than outsourcing a prototype design or having an informal conversation. We assume that patents resulting from more costly links will contain a higher number of NPRs because such investment will increase the firms' capacity to use the science base and codify it through a citation. Implicit in this is the assumption that if the average type of firm in the region engages into less costly links with science, we may not observe NPRs although the links exist. Therefore, NPRs would not be useful to trace science-industry links in a region with low absorptive capacity.

Hypothesis 2. NPRs are a less valid indicator of science-industry links in a region with low absorptive capacity than in a more research-intensive environment (region, nation or technology).

Geographic distribution of citations indicates regional concentration of knowledge flows

The geographic location of a citation can be used to trace the origin of a knowledge source and to determine how far awareness of an invention has diffused. Several authors have studied the differences between national and international citation patterns. JAFFE & TRAJTENBERG [1996] found that knowledge spillovers are geographically localized but that disparities fade over time. ALMEIDA & KOGUT [1999] concluded that local knowledge spillovers are more likely in high-tech regions such as the Silicon Valley, the Boston area and Austin in Texas. The idea is that local innovations are stimulated by nearby technology developments. THOMPSON [2006] tested the correlation between technology classes and geography and found national citations were more common when both the patent and the cited patent belonged to the same technology class.

When applicant and examiner added citations are studied separately, significant differences are found. THOMPSON [2006], using USPTO data, concluded that applicants are more likely than examiners to include national citations. He argued that localization of knowledge transfer decreases over time, but that national borders constitute more impermeable barriers to international knowledge outflow. These results are not surprising as researchers tend to move more frequently within national than international borders. CRISCUOLO & VERSPAGEN [2005] studied intra-European patent citations. They, like Thompson and others [SAMPAT, 2004; ALCÁCER & GITTELMAN, 2006], found that applicant citations were more geographically concentrated.

We would expect the results for regions with low absorptive capacity not to be similar to those found by the studies mentioned above. In regions with low absorptive capacity, physical proximity will not be a reason for accessing explicit knowledge, since firms within the region will use explicit knowledge incorporated in capital goods, mainly from other regions. We could hypothesize that in regions with low absorptive capacity citations will not present high geographical localization in the region.

Hypothesis 3. Patent citations will not be geographically localized in regions with low absorptive capacity, contrary to very research-intensive environments (regions, nations or technologies).

A low absorptive capacity regional sample

The research context: the Valencian Community

In an earlier study [AZAGRA & AL., 2006], the Valencian Community was characterized as a region with low absorptive capacity. The main features of the region are:

- The low-tech profile of its economic structure (predominance of microfirms in services and traditional manufactures).
- The weaknesses of its innovation activities (innovation occurs, but it is mostly incremental and through machinery and equipment acquisition, with little expenditure on R&D).
- The scarcity of qualified personnel at firms, even in knowledge-intensive sectors.
- Policy emphasis on enhancing technology transfer (similar to high-tech regions or countries).

We would stress that this context is compatible with the existence of important academia-industry links. A report for the Valencian R&D Council [ACCID, 2005], showed that 3% of Valencian firms' sales were due to product innovations that could not have been developed in the absence of academic research. This result is similar to the findings from other studies based on US and German data [MANSFIELD, 1998; BEISE & STAHL, 1999]. It also showed that industry funding of Valencian university R&D (6%–8%) was similar to the Spanish average and higher than the EU and OECD average and that Valencian firms tend to contract out to Valencian universities for low-tech, short-term oriented R&D. In other words, academia-industry links exist, because universities to an extent have adapted to the regional absorptive capacity level.

More specifically, we showed that most faculty members strongly support university-industry interaction [AZAGRA & AL., 2006], that firms do not show the same propensities to interact with universities and that some faculty members prefer to interact with firms outside the region [AZAGRA, 2007A] because this provides access to higher technology and larger firms [AZAGRA, 2007B]. Moreover, for science-dependent sectors we found an 'alocalization' effect in terms of university-industry links [TODT & AL., 2007], in contrast to current regional R&D policy, based on a linear vision of innovation and spillover effects of local knowledge production.

Finally, we should mention that the Valencian Community has been a pioneer in the establishment of a network of technology centres, focused on development rather than research, but at some point contributing to science-industry links.

It is important to keep this in mind to provide a context for some of the results in the next section.

Methodology and data

The source of patent application information for the Valencian Community is the Spanish Patent and Trademark Office (OEPM), which selects patents through the field 'province', and includes the codes for the three Valencian NUTS 3 regions: Alicante, Castellon and Valencia. We recovered patents from 1999 to 2003.[3]

The OEPM database includes a field for the name(s) of the patent applicant. To identify their correspondence with firms is not straightforward and involved checking each patent to classify and standardize it. We classified the 1,382 patents of the Valencian Community registered between 1999 and 2003, distinguishing between firm applicants and other types of applicants based on an acronym for firm type in the name of the applicant, i.e. SL – limited society, SA – anonymous society, etc.

For this group of patents, we constructed a database to include citations. We studied the full-text of the patent application form for each patent, taking particular notice of the description field where the applicant includes the prior art, and counted the number of citations.[4] Then we went through the same procedure for the citations in the prior art report (included by the examiner). This task was made easier because some of the citations were already present in the OEPM database. The result was 712 applicant citations and 2,849 examiner citations, which we further classified according to two criteria:

- Scientific or technologic nature, i.e. NPRs or PRs, respectively.
- Geographic location of PRs and Thomson Scientific's Science Citation Index (SCI)-NPRs: the OEPM database does not contain a field with this information, so we had to search all the PRs and some of the NPRs. Because of cost constraints we did not search for all NPRs, but PRs plus SCI-NPRs represent almost 90% of all references.

In the case of the PRs, we searched for the location of the applicant in the Cibepat[5] database or applied to the relevant national office in the case of patents from other national patent offices. It would have been better to identify the location of all the applicants, but we assume that inventors are more likely to apply for patents in their own countries.[6] In the case of the Spanish patents, we used the field 'province' to

[3] OEPM updates online data on a regular basis. The date of the extraction for this paper is 23 June 2006.

[4] In the full-text of the application form, applicants provide a description of their invention to demonstrate its novelty, to describe it fully and to explain how it is made. Although most include a section on prior art, it is not obligatory to include citations. Consequently, a patent without citations does not mean that there is no description, but only that it is a tacit description of the prior art.

[5] For international patents, information was completed from the EPO database (*esp@cenet*) and the World International Patent Organization (WIPO) database.

[6] Other evidence from the sample reinforces this idea: the expectation that the origin of a cited EPO or PCT patent is mostly abroad, was widely confirmed, thus it confirms the existence of such a bias.

distinguish between patents from the Valencian Community and those from the rest of the Spain.

For the SCI-NPRs, we used the Thomson Scientific database to recover the geographic location of the first author to set it against the citation. It would have been better to consider all the authors, but in the case of a small international cooperation this would not have been problematic. Moreover, we assume that if the same group of authors published another paper, the names might appear in a different order, so on average the different geographical affiliations would be cancelled out.

Results

Citations used to indicate an invention's knowledge base – not in the Valencian Community

As Table 1, panel on applicants shows, more than 70% of patent applicants have no citations. The average number of applicant citations per patent is 1.17, but the most frequent mode is zero. Moreover, the 1.17 average hides an important disparity between patents with applicant citations and patents without applicant citations. The 30% of patents with applicant citations had an average of 4.13 applicant citations per patent, with a still low mode of 1, and observable in almost one-third of cases.

Table 1. Patents from the Valencian Community with firm applicants and their citations

Applicants

	Nº of patents	Nº of applicant citations	Average	Mode	Frequency
All patents	571	669	1,17	0	72%
Patents without applicant citations	409	–	–	–	–
Patents with applicant citations	162	669	4,13	1	32%

Examiners

	Nº of patents	Nº of examiner citations	Average	Mode	Frequency
All patents	571	2 758	4.83	4	23%
Patents without applicant citations	409	1 987	4.86	4	22%
Patents with applicant citations	162	771	4.76	4	24%

Source: Own elaboration from OEPM: Cibepat

Panel on examiners of Table 1 shows that the average goes up to 4.83 citations per patent for examiners. The mode is 4 citations per patent, and this occurs in less than a quarter of cases. Therefore, most examiner citations originate with the examiners themselves. This is especially true for patents without applicant citations, since the difference between patents with and without applicant citations does not significantly change the average number, the mode, or the frequency for examiners.

Another way of analysing the differences between applicants and examiners is to look at the correlation between the number of citations from each: the correlation coefficient is close to zero even if we restrict the sample to patents with a positive number of applicant citations.

These findings support Hypothesis 1, i.e. examiner citations are not very representative of applicant knowledge flows in a region with low absorptive capacity such as the Valencian Community.

NPRs are useful to trace science-industry links – not in the Valencian Community

As we can see from Table 2, 98% of examiner citations are PRs, implying that only 2% are NPRs. Through examiners citations, we identify a very small link between firm patents and their scientific base.

Table 2. References in Valencian Community patents with firm applicants, according to their patent or non-patent type

Type of cited reference	Nº of examiner citations	Nº of applicant citations
PR	2 707 (98%)	547 (82%)
NPR	51 (2%)	122 (18%)
Total	2 758	669

Source: own elaboration from OEPM: Cibepat and Thomson Scientific: Web of Knowledge

Consistent with the literature, applicants tend to include more NPRs (18%) than examiners. However, the majority of citations are still PRs (82%). Therefore, the largest part of the knowledge base of Valencian firm patents is technological rather than scientific. This result, given the existence of the academia-industry links provides evidence to support Hypothesis 2.[7]

Geographical localization – not in the Valencian Community

Table 3 reports the identification of the origin of PRs cited by Valencian patents. If we accept the total in the last column to be an approximately, 84% of the PRs cited

[7] As is usual in this type of study, we also identified publications in the SCI. Surprisingly, most citations – both applicants and examiners – to NPRs were publications in SCI journals. Therefore, although the link between Valencian firms and science is probably weak, it is not particularly so in terms of links with the high-quality science. However, we must take this result cautiously because it relies on a very low number of patents, concentrated in the pharmaceutical and biotechnology sectors.

by examiners in the panel on applicants correspond to knowledge generated abroad.[8] A large proportion of these are from the USA (42%), which reflects the US position as technology leader, with the European Union accounting for 36%, and if we consider international patents the proportions are 53% and 24% respectively. We should emphasize that foreign patents account for 8% of total Spanish patents.

Table 3. Patent references in patents from the Valencian Community with firm applicants, according to their geographic origin

Examiner citations

Geographic origin of cited patents	Nº of citations of other national patents	Nº of citations of other international patents	Total
Foreign	**1 811 (83%)**	**475 (92%)**	**2 286 (84%)**
European Union countries	568 (31%)	252 (53%)	830 (36%)
United States	856 (47%)	113 (24%)	969 (42%)
Japan	164 (9%)	47 (10%)	211 (9%)
China	16 (1%)	1 (0%)	17 (1%)
Rest of the world	31 (2%)	62 (13%)	93 (4%)
Spanish, with foreign applicant	176 (10%)	-	176 (8%)
Spanish	**382 (17%)**	**21 (4%)**	**403 (15%)**
From the Valencian Community	230 (60%)	10 (48%)	240 (60%)
(of which, self-quotations)	(53)	(2)	(55)
Outside Valencian Community	149 (39%)	11 (52%)	160 (40%)
Undetermined	3 (1%)	0	3 (1%)
Undetermined	**0**	**18 (4%)**	**18 (1%)**
Total	**2 193**	**514**	**2707**

Applicant citations

Geographic origin of cited patents	Nº of citations of other national patents	Nº of citations of other international patents	Total
Foreign	**258 (64%)**	**131 (90%)**	**389 (71%)**
European Union countries	47 (18%)	66 (50%)	113 (29%)
United States	157 (61%)	33 (25%)	190 (49%)
Japan	7 (3%)	13 (10%)	20 (5%)
China	3 (1%)	0 (0%)	3 (1%)
Rest of the world	0 (0%)	19 (15%)	19 (5%)
Spanish, with foreign applicant	44 (17%)	-	44 (11%)
Spanish	**144 (36%)**	**6 (4%)**	**150 (27%)**
From the Valencian Community	74 (51%)	5 (83%)	79 (53%)
(of which, self-quotations)	(23)	(3)	(26)
Outside Valencian Community	47 (33%)	1 (17%)	48 (32%)
Undetermined	23 (16%)	0	23 (15%)
Undetermined	**0**	**8 (6%)**	**8 (1%)**
Total	**402**	**145**	**547**

Source: own elaboration from OEPM: Cibepat

[8] Strictly speaking, it is not correct to sum the second and third columns for the reasons given in the methodology: in framed cells, patents may have some Spanish applicants. However, the ratio of foreign applicants is so large in all cases that it is doubtful whether greater refinement would change the results, even without the supposition of a national bias.

Of the PRs cited by examiners 15% are national patents with Spanish applicants. The majority (60%) are Valencian applicants –one-fifth self-citations (the citing and cited patent applicant is the same), so the proportion of citations to patents from other Spanish regions is 40%. Consequently, although most of the technological knowledge base of Valencian firm patents is foreign in origin, we cannot find a preference for national over regional boundaries.

In terms of the geographical origin of the PRs or patents cited by applicants, the panel on examiners shows that proportion of foreign ones is high (71%), but not so high as for examiners. This is consistent with the studies referred to earlier. Similar to the case of examiners, the USA accounts for the largest proportion, but the distance to other regions is smaller, e.g. to the European Union.

Among the national PRs cited by applicants, the largest proportion is Valencian PRs (53%) – one-third self-citations. While applicants might overestimate the local component, we cannot find a significant preference for the national.

Because of the small number of NPRs and SCI-NPRs in our sample, we refrain from making detailed claims as to their geographical distribution. However, it seems that, similar to PRs, the foreign knowledge base is much larger than the national one, in the cases of both examiner and applicants.

The above results support Hypothesis 3. In other words, where we have a low absorptive capacity region sample, we cannot confirm the geographical localization of citations found in other studies, but rather the reverse, that there is geographical alocalization.

Conclusions

In this article, we discussed the utility of standard uses of citations in patents in regions with low absorptive capacity and presented some evidence from such a region, the Valencian Community. The results have several implications.

First, the few patents with applicant citations imply that examiner citation analyses overestimate the weight of the explicit knowledge base of patents in regions such as the Valencian Community and hide one of the region's weaknesses: the scarcity of qualified personnel at firms, which makes it more difficult for firms to use explicit knowledge. Examiner citation analyses would also overestimate the importance of knowledge flows for innovation. There is therefore a need to separate between examiner and applicant citations. In this region and, in general, in Spain, the process of innovation is mainly built on the internal capabilities of firms. In this sense, neither the externalisation of R&D nor the cooperation with scientific actors appear to be determining for innovative performance [VEGA & AL., 2007].

Second, the low number of scientific citations points to its lack of utility as an indicator of science-technology links in low absorptive capacity regions such as the Valencian Community, where other signs would suggest that these links exist. If the science-technology links in such region are considerable, the fact that firm patents quote only a few scientific publications may suggest that these links may be technological in nature. It may also be that universities contribute in a useful, but short-term oriented way, with low-level knowledge, which confirms the current position of Valencian firms as technology followers rather than leaders. This is coherent with the low degree of novelty of Valencian patents and with the predominance of incremental over radical innovation detected in other studies of the Valencian Community.

Third, we can see that the largest part of the knowledge base of Valencian firm patents is foreign in origin. One explanation for this may be that ideas are created in every part of the world, and to justify a certain degree of novelty, firms develop international search strategies.

This suggests that some policies to support innovation in regions with a low absorptive capacity may be less successful, e.g. supporting regional R&D through subsidies rather than the incorporation of qualified personnel at firms or the increase of local university-industry links. Policymakers have a limited vision of the process of technological development and its socioeconomic consequences and consider it as a linear, automatic and geographically localized process. In this sense, policymakers should be better informed about the determining importance of internal sources of firms, the scarce relation among the actors of the innovation system and the alocalization of knowledge that favours epistemological communities rather than local networks.

To sum up, the traditional use of patent citations in regions with low absorptive capacity to trace knowledge flows and science-technology links has many flaws, as we have shown in the case of the Valencian Community. Nevertheless, we would support the utility of patent citations to identify knowledge alocalization in such regions compared with more research-intensive environments, taking into account the separation of applicant vs. examiner citations and the study of full text reports.

*

Thanks are due to Martin Meyer, for his careful support and revision; to the Valencian R&D Consultancy Council for funding the research on which the study relies; to Juan Alcácer and Julie Callaert for their responses to our queries.

Since 1 December 2006 Joaquín M. Azagra-Caro has been employed by the JRC IPTS. However, the research described in the paper was performed in the main during the time he was employed by INGENIO. The views expressed in this article are the authors' and do not necessarily reflect those of the European Commission (EC). Neither the EC nor anyone acting on behalf of the EC is responsible for the use that might be made of the information.

References

ACCID (2005), *Análisis de la contribución de la I+D universitaria al desarrollo económico de la Comunidad Valenciana*. In: Alto Consejo Consultivo en Investigación y Desarrollo de la Presidencia de la Generalitat Valenciana, Informe anual sobre el estado de la investigación, el desarrollo y la tecnología en la Comunidad Valenciana. Valencia, Generalitat Valenciana.

ACOSTA, M., CORONADO, D. (2002), Las relaciones ciencia-tecnología en España. Evidencias a partir de las citas científicas en patentes. *Economía Industrial*, 346 : 27–46.

ACOSTA, M., CORONADO, D. (2003), Science-technology flows in Spanish regions. An analysis of scientific citations in patents. *Research Policy*, 32 : 1783–1803.

ALCÁCER, J., GITTELMAN, M. (2006), Patent citations as a measurement of knowledge flows: The influence of examiner citations. *Review of Economics and Statistics,* 88 (4) : 774–779.

ALMEIDA, P., KOGUT, B. (1997), The Exploration of Technological Diversity and the Geographic Localization of Innovation. *Small Business Economics*, 9 : 21–31.

ALMEIDA , P., KOGUT, B. (1999), Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, 45 (7) : 905–917.

AZAGRA-CARO, J. M., *The regional dimension of university-industry interaction*. In: SURIÑACH, J., MORENO, R., VAYÁ, E. (Eds), Knowledge Externalities, Innovation Clusters and Regional Development. Cheltenham and Northampton, Edward Elgar, (2007a).

AZAGRA-CARO, J. M. (2007b), What type of faculty member interacts with what type of firm? Some reasons for the delocalisation of university-industry interaction. *Technovation,* 27 : 704–715.

AZAGRA-CARO, J. M., ARCHONTAKIS, F., GUTIÉRREZ-GRACIA, A., FERNÁNDEZ-DE-LUCIO, I. (2006), Faculty support for the objectives of university-industry relations versus degree of R&D cooperation: the importance of regional absorptive capacity. *Research Policy,* 35 (1) : 37 – 55.

BEISE, M., STAHL, H. (1999), Public research and industrial innovations in Germany. *Research Policy*, 28 (4) : 397-422.

CALLAERT, J., VAN LOOY, B., VERBEEK, A., DEBACKERE, K., THUS, B. (2006), Traces of prior art: An analysis of non-patent references found in patent documents. *Scientometrics,* 69 (1) : 3–20.

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

COOKE, P. (1992), Regional innovation systems: competitive regulation in new Europe, *Geoforum*, 23 (3) : 365–382.

CRISCUOLO, P., VERSPAGEN, B. (2005), *Does It Matter Where Patent Citations Come From? Inventor versus Examiner Citations in European Patents*. ECIS Working Papers 05.06.

HARHOFF, D., NARIN, F., SCHERER, F. M., VOPEL, K. (1999), Citation frequency and the value of patented inventions. *Review of Economics and Statistics*, 81 (3) : 511–515.

JAFFE, A., TRAJTENBERG, M., FOGARTY, M. S. (2000), Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review*, 90 (2) : 215–218.

JAFFE, A., TRAJTENBERG, M. (1996), Flows of knowledge from universities and federal laboratories: Modeling the flow of patent citations over time and across institutional and geographic boundaries. *PNAS*, 93 : 12671–12677.

LEYDESDORFF, L., MEYER, M. (2003), The Triple Helix of university-industry-government relations. *Scientometrics*, 58 (2) : 191–203.

MANSFIELD, E. (1998), Academic research and industrial innovation: An update of empirical findings. *Research Policy*, 25 (3) : 773–76.

MEYER, M. (2002), Tracing knowledge flows in innovation systems. *Scientometrics*, 54 (2) : 193–212.

MEYER, M. (2000), What is special about patent citations? Differences between scientific and patent citations. *Scientometrics,* 49 (1) : 93–123.

MICHEL, J., BETTELS, B. (2001), Patent citations analysis: A closer look at the basic input data from patent search reports. *Scientometrics,* 51 (1) : 185–201.

NARIN, F., HAMILTON, K. S., OLIVASTRO, D. (1997), The increasing linkage between U.S. technology and public science. *Research Policy*, 26 (3) : 317–30.

NARIN, F., NOMA, E. (1985), Is technology becoming science? *Scientometrics*, 7 (3-6) : 369–381.

NIOSI, J., BELLON, B. (2002.), The absorptive capacity of regions. *Colloque Economie Méditerranée Monde Arabe*. Sousse 20-21 September.

SAMPAT, B. (2004), *Examining Patent Examination: An Analysis of Examiner and Applicant Generated Prior Art*. Working Paper, School of Public Policy, Georgia Institute of Technology.

SCHMOCH, U. (1993), Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics,* 26 (1) : 193–211.

THOMPSON, P. (2006), Patent citations and the geography of knowledge spillovers: Evidence from inventor- and examiner-added citations. *Review of Economics and Statistics*, 88 (2) : 383–388.

TODT, O., GUTIÉRREZ-GRACIA, A., FERNÁNDEZ-DE-LUCIO, I., CASTRO-MARTÍNEZ, E. (2007), The regional dimension of innovation and the globalization of science: the case of biotechnology in a peripheral region of the European Union. *R&D Management*, 37 (1) : 65–74.

VEGA-JURADO, J., GUTIÉRREZ-GRACIA, A., FERNÁNDEZ-DE-LUCIO, I., MANJARRÉS-HENRÍQUEZ, L. (2007). The effect of external and internal factors on firms product innovation. *Research Policy*, forthcoming.

V

Submitted to Journal of the American Society for Information science and Technology (under revision)

Smoothing the lies: Do patent examiners take notice if applicants include citations?

Joaquín M. Azagra-Caro*

*European Commission, Joint Research Centre (JRC), Institute for Prospective Technological Studies (IPTS), Edificio Expo, C/Inca Garcilaso 3, E-41092 Sevilla, Spain***

CSIC - Consejo Superior de Investigaciones Científicas, Institute of Public Goods and Policies (IPP-CCHS), C/Albasanz 26-28, E-28037 Madrid, Spain. E-mail: joaquinmaria.azagra@cchs.csic.es

INGENIO (CSIC-UPV), Polytechnic University of Valencia, Edif. 8E 4º, Camino de Vera s/n, Valencia, Spain. E-mail: franperr@ingenio.upv.es

Pauline Mattsson

Unit for Bioentrepreneurship (UBE) - Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, SE-171 77 Stockholm, Sweden. E-mail: pauline.mattsson@gmail.com

François Perruchas

INGENIO (CSIC-UPV), Polytechnic University of Valencia, Edif. 8E 4º, Camino de Vera s/n, Valencia, Spain. E-mail: franperr@ingenio.upv.es

* Corresponding author at: CSIC - Consejo Superior de Investigaciones Científicas, Institute of Public Goods and Policies (IPP-CCHS), C/Albasanz 26-28, E-28037 Madrid, Spain. Tel.: +34 916 022 370. E-mail address: joaquinmaria.azagra@cchs.csic.es.

** The views expressed in this article are those of the authors; they do not necessarily reflect those of the European Commission (EC). Neither the EC nor anyone acting on behalf of the EC can be held responsible for the use that might be made of the information.

Examiner patent citations are a popular source of indicators of technological impact and knowledge flows, despite various critiques. We analyse the distribution of examiner patent citations according to patent characteristics, to show their comparative meaningless. Our findings show that it is the science-base of the technology that determines the inclusion of applicant citations. However, this gets masked by the citations added by patent examiners, who smooth the distribution of citations across technology classes and include the 'standard' knowledge bases regardless of which references applicants cite. Some researchers have called for the use of applicant rather than examiner patent citations to build indicators of technology impact and knowledge flows. However, we show that the former are not necessarily 'better' than the latter, because applicants may 'inflate' the numbers in international patents especially when there are co-applicants. The implications are that analysts should consider alternative uses of patent citations e.g. to build indicators of trust within a research system.

Introduction

'Generation, diffusion and exploitation of knowledge are at the core of the research system' (EC, 2007: 16). This official recognition of a social interest underlies much academic research on the origins and the destiny of the knowledge produced. When the focus is on technological applications, key data for quantitative analyses are patent citations. Explaining the state-of-the-art requires some differentiations, starting with this:

- according to direction of the citation, we can classify patent citations as *forward* or *backward* citations.

The use of forward citations is customary to express the technological impact of the patented invention (Noma & Olivastro, 1984), often as a function of the characteristics of the patent (Allison & Sager, 2007). The use of backward citations is customary to express the knowledge base of the patented invention. This introduces another important distinction:

- according to the type of cited document, we can classify patent citations as *patent-to-patent* or *patent-to-paper* citations (also known as patent vs non-patent references or literature).

Patent-to-patent citations are the most frequent and often serve as a proxy for the whole knowledge base of the invention, more properly specified as the technological knowledge base. Applications include analysis of whether they are from the same country as the applicant, which measures geographical spillovers (Jaffe & Trajtenberg, 1996). Patent-to-paper citations frequently serve as proxies for the scientific knowledge base (Narin & Noma, 1985; Hassan, 2003; Leydesdorff, 2004), usually leading to some justification of the importance of science, or at least of some cutting-edge technologies.

As Chen (2003) accurately notes, patent-to-patent backward citations are used by economists to explore knowledge spillovers, while their wider application - especially, of patent-to-paper citations – is generally in scientometrics. However, classical works use similar wording to suggest a causal effect from citation to patent: ‘knowledge diffusion’, ‘utility of basic research to technology’, etc.

The increasing use of these techniques among researchers has developed in parallel with advocates of their application to justify research funding (Kostoff, 1994). However, some qualitative and case studies recommend caution in the interpretation of results, based on another distinction among patent citations:

- according to who inserted the citation, we can classify patent citations as *examiner* or *applicant* citations (the latter are also, somewhat improperly, referred to as inventor citations).

The traditional studies on patent citations rely on patent examiner citations that appear on the front pages of patent documents. Their use introduces two main problems (see e.g. Meyer, 2000; Michel & Bettels, 2001). First, examiner citations may provide biased information about knowledge flows, since numbers might vary for administrative reasons. Second, unlike the US, the patent system in Europe does not compel patent applicants to disclose complete information in patent documents, making it much more probable that patents examiners will add citations.

These critiques have not deterred traditional quantitative studies, perhaps because qualitative evidence is not sufficiently convincing. And some academics are claiming that patent citations are useful because they are more credible than paper citations (Lai & Wu, 2005) and that more efforts should be devoted to producing better-codified data on patent citations (Stock & Stock, 2006) to facilitate its use. It has been proposed that examiner forward citations should be used to build indicators such as h-indexes of firms' technological performance (Guan & Gao, 2008).

The first contribution of this paper is to provide quantitative evidence of possible inconsistencies in examiner citations that should prevent from extended use.

Nevertheless, the critiques made about the use of patent citations have had some consequences and have inspired several quantitative research lines. The first is to adopt more careful wording to refer to the relation between patents and their citations: 'interactions', 'links' or 'linkages', 'vicinity', etc. (Tijssen, 2001; Callaert et al., 2006). A second is to use patent data in alternative ways to visualise relations within the knowledge base, e.g. through co-classification in technology classes to show perhaps that countries are not an appropriate unit of observation (Leydesdorff, 2008). A third is to promote a quantitative approach to the difference between examiner and applicant citations. This work is confirming that the use of examiner citations biases the interpretation of findings, for instance, because the knowledge base appears to be more localised if measured through applicant citations (Criscuolo & Verspagen, 2008). The degree of localisation and of differences between examiner and applicant citations depend most likely on the absorptive capacity of the region (Azagra et al., 2009).

This last stream of research pleads for the use of applicant rather than examiner citations as a better expression of knowledge flows, and links directly to the second contribution of the present paper, which is to establish whether applicant citations are a better indicator of knowledge flows.

The next section builds a conceptual framework for the distribution of patent citations according to the characteristics of patents to help understand the weaknesses of examiner and applicant citations. Using testable hypotheses, having described the research context, we present the data and methodology, our results and some conclusions.

Some insights into the meaning of examiner and applicant patent citations

The distribution of examiner backward citations by patent characteristics

We have reported that: (i) quantitative studies on examiner forward patent citations explore how they vary in terms of patent characteristics; (ii) qualitative studies suggest that patent characteristics influence the inclusion of examiner backward patent citations. These two aspects justify quantitative study of the distribution of examiner backward citations according to the characteristics of the patents. However, earlier investigations have been incidental, and provides descriptive statistics rather than substantial analysis. Which characteristics are included? So far, the focus has been limited mainly to time effects and technology classes.

In terms of time effects, they seem to be influential in the distribution of examiner citations. US examiner time constraints have become tighter, with the increase in the number of patent applications outpacing increases in the number of examiners (Merrill et al., 2004). Time and resource constraints necessarily limit the comprehensiveness of examiners' prior art search, producing 'citation inflation' in US Patent and Trademark Office (USPTO) patents (Hall et al., 2000). Callaert et al. (2006) confirm an increase in USPTO citations between 1991 and 2001, while for European Patent Office (EPO) patents the average number of examiner citations per patent in the same period has been constant (or even slightly decreasing).

In terms of technology classes, Callaert et al. (2006) find a larger number of EPO examiner citations in patent-to-patent citations in Mechanical Engineering and Machinery and patent-to-paper citations in Chemistry and Pharmaceuticals. In the USPTO, Process Engineering and Social Equipment ranks first for patent-to-patent citations and Chemistry and Pharmaceuticals for patent-to-paper citations. Sampat (2004) finds that examiners in the US face particular challenges in identifying prior art in emerging technological fields, focusing on nanotechnology.

Although not touched on in descriptive analysis, there are some other characteristics of patents that may condition the distribution of citations.

First, given the influence of technology classes, it could be argued that related dimensions, such as geographic region or economic sector, may be influential, i.e. the more specialised regions and sectors are in technologies where patent examiners insert more citations, the higher will be the number of examiner citations in these regions and sectors.

Second, there may be administrative reasons for the inflated number of citations. For instance, protection means and ownership regimes.

In relation to means of protection, many works consider only one protection alternative, national (USPTO) or international (EPO), probably because of the major focus on the US and Europe generally. However, when studying a single European country or region, several alternatives may be relevant, national and international, because of the home advantage effects (Criscuolo, 2006) and, also, alternative routes may be indicators of geographical patterns or technological protection (Azagra et al., 2006). Moreover, 'any non-national language documents are only cited when already having been quoted by the applicant' (Michel & Bettels, 2001: 198). Callaert et al. (2006) show that USPTO examiners include a higher number of citations per patent than do EPO examiners. The reason for this for could be that USPTO applicants have the legal obligation to provide full lists of prior art which tend to accepted by examiners, whereas there is no such obligation in the case of the EPO (Michel & Bettels, 2001).

Regarding the ownership regime, although there is little evidence about the impact of co-ownership on the number of citations in patents, it could be argued to be influential, since institutions may differ in their tendency to include citations. For instance, universities and public research organisations may show a higher propensity for citation, so co-applications involving firms and these institutions may be more likely to incorporate prior art than patent applications from firms alone.[1]

It is useful to synthesise this information in the following hypothesis:

Hypothesis 1. Patent characteristics play an influential role in the distribution of examiner citations.

[1] As for forward citations, Jaffe & Trajtenberg (1996) have shown that universities on average receive more citations, followed by corporations and government institutions.

Front-page vs full-text citations: Examiners trust applicants

Another common source of criticism of information derived from examiner patent citations is that it is the patent applicants that really know the base of the invention (Jaffe et al., 2000). However, the codification of applicant citations is lagging behind that of examiner citations, making quantitative research on the former less easy. It is only recently that quantitative studies have begun to investigate examiner and applicant added citations separately, and this is due to improvements in computation facilities for identifying citations on front pages of patent documents (Sampat, 2004; Thompson, 2006; Alcácer & Gittelman, 2006; Criscuolo & Verspagen, 2008).

However, following the economic tradition of measuring knowledge spillovers referred to in the introduction, most of these studies focus on citing-cited pairs. They analyse localisation effects, regardless of whether the citations are from examiners or applicants. They are rarely interested in differences in patent characteristics although such differences are relevant, since they may be a basic reason for the different sets of patents included by applicants and examiners.

Another characteristic of these studies is that they look only at applicant citations *that the examiner considers relevant*, i.e. those on the front pages. However, the applicant citations which are included in the patent text, may be much closer to the knowledge base, since they have escaped the examiners' editing. Because of the time consuming process of retrieving the information, only a few studies consider this aspect (see exceptions such as Acosta & Coronado, 2003)[2]

[2] Acosta & Coronado (2003) analyse full-text applicant citations. However, they add (Acosta & Coronado, 2003: 1794) that they also include examiner citations, which is rather confusing. What is clear is that the weight of full-text citations in their study is higher than in many other studies.

So, there do exist some results related to the influence of patent characteristics on applicant citations. For instance, regarding time effects, Criscuolo & Verspagen (2008) find that the share of applicant citations in EPO patents decreased between 1985 and 1999. In terms of route of protection, the same authors conceptually justify the fact that both examiners and applicants tend to add more citations in patents applied for through the Patent Cooperation Treaty (PCT)-EPO procedure, compared to the direct-EPO procedure. The evidence from their econometric tests is inconclusive about whether this affects the probability of examiner-added citations.

Technoeconomic characteristics (region, technology, economic sector) are also worthy of attention. Focusing on regions, Acosta & Coronado (2003) observe the concentration of patent-to-paper citations in the more developed regions in Spain, such as Madrid, Catalonia and the Basque Country, and their scarcity in regions with GDP below 75%[3] of the EU average. However, although Acosta & Coronado do not stress this point, the distribution of patent-to-patent citations (which are more abundant than patent-to-paper citations) is relatively even across regions.

For technology classes, Sampat (2004) provides some empirical evidence and suggests that applicants are more likely to include citations in fields where patenting is important such as chemicals and pharmaceuticals. Drawing on the previous literature, Sampat argues that in fast-moving, more technology intensive classes, the ability of examiners to access current information is limited, and applicants are better informed about the closely related prior art and therefore include more citations than examiners. In technologies that are less science intensive, citation rates between examiners and applicants will generally be more similar. In the case of EPO patents, Criscuolo & Verspagen (2008) find that the share of applicant

[3] Note that the present empirical analysis studies the Valencian Community, which is included in this group.

citations is higher in chemistry and materials and lower in semiconductors, telecommunication, audiovisual and information technology.

For economic sectors, Acosta & Coronado (2003) show that, in Spain, 85% of patent applicants' patent-to-paper citations are concentrated in only three sectors (chemistry, pharmaceuticals and biotechnology). However, patent-to-patent citations predominate in high and medium-high technology sectors, such as electrical engineering and instruments, which is why Leydesdorff (2004) suggests that a sector such as biotechnology is not a valid model for how university-industry interactions occur in general.

Given the scarcity of direct comparisons of citations along patent characteristics, it is difficult to establish any a priori expectations. Assuming that the administrative and technoeconomic reasons for increasing the number of citations affect examiners and applicants equally, we can start by formulating a cautious hypothesis:

Hypothesis 2. Applicants and examiners include comparable distributions of citations independent of patent characteristics.

Within front-page citations: Examiners behaviour is different depending on whether applicants disclose some knowledge or not

The best conceptual background to understand the differences between applicant and examiner citing-cited pairs is in Alcácer & Gittelman (2006). They investigate different scenarios in terms of examiners' decisions about whether to complement or supplement inventor citations. Implicitly, this approach assumes that examiners have something to complement or supplement because inventors disclose some knowledge.

This may not be an issue in the US, where applicants are subject to 'duty of candour', which forces them to be exhaustive in their inclusion of references to prior art (Meyer, 2000;

Michel & Bettels, 2001). However, it is important in Europe, since there is no imposition of that kind on applicants, meaning that they are free to decide whether or not to include references. For instance, Acosta & Coronado (2002) show that only 31% of Spanish patents include applicant full-text patent-to-patent citations and 10% of Spanish patents include applicant full-text patent-to-paper citations, i.e. a small number compared to front-page citations, which appear in all patents. Similarly, Azagra et al. (2009) find that only 30% of patents from the Valencian Community (a Spanish NUTS 2 region) includes at least one applicant citation.

Thus, in the European case, it is worth studying whether examiners treat patents with and without citations equally. If treatment is equal in that in both cases the examiner transposes a standard body of references, then it is more difficult to uphold their objectivity. If treatment is not equal, we would expect the distribution of examiner citations according to the characteristics of the patent, to differ between patents with and without citations, because each will be subject to a case-by-case search report. Let us assume this situation as a starting point:

Hypothesis 3. The distribution of examiner citations is different according to whether the analysis includes patents with applicant citations or all patents.

Within full-text citations: Applicants who disclose some knowledge are representative of all applicants

Criscuolo & Verspagen (2008) conclude their study with a plea for greater use of applicant rather than examiner citations, as indicators of knowledge flows. They rely on the assumption that their observation of applicant citations is representative of the behaviour of all applicants. The problem, which the authors are aware of since they study the European case, once again, is lack of applicant citations. What about applicants who do not include citations?

In this situation, it is questionable whether we should consider even full-text applicant citations as being representative of the knowledge base. A good indication might be if the number of patent citations was the same for different types of patents for both applicants who disclose some information and those who do not reveal any. To test for this, it is useful to formulate:

Hypothesis 4. The distribution of applicant citations is similar regardless of whether the analysis includes patents with applicant citations or all patents.

The research context: the Valencian Community

The Valencian Community is described as having low absorptive capacity (Azagra, 2007). The main features of the region are:

- low-tech profile of its economic structure (predominance of microfirms in services and traditional manufactures);
- weak innovation activities (innovation is mostly incremental and in the form of machinery and equipment acquisition, with little expenditure on R&D);
- scarcity of qualified personnel even in firms in the knowledge-intensive sectors;
- policy emphasis on enhancing technology transfer (similar to high-tech regions or countries).

Table 1 provides a more detailed description of the industrial structure of the Valencian Community.

{Table 1 around here}

The column presenting gross value added (GVA) shows the share of the average value of this variable for 1999-2003 (for comparison with patents), in constant prices, by NACE activity. The data are from the Spanish National Statistics Institute's Regional Accounts. We group economic branches following (and extending) the typologies proposed by Pavitt (1984) and Breschi and Malerba (1997). The classification is arguable, but it is not the objective of this paper to justify this, and alternative sector groupings would have led to similar conclusions regarding posterior analysis.

Supplier-dominated sectors predominate, especially 'construction' and services such as 'wholesale and retail trade', 'non-market services' and 'transport, storage and communication'. Within production-intensive sectors, it 'manufacturing of other non-metallic mineral products' stands out and is based on regional strength in ceramic tiles. The contribution of science-based sectors is relatively small. The case of 'real estate, renting and business activities' is rather special, since the high GVA weighting is due mainly to the activities in the supplier-dominated sectors. However, because they include 'computer and related activities' and 'research and development' we chose to classify them under science-based sectors, which is relevant for the information in subsequent columns on patents and patent citations. First, we explain the methodology and data.

Methodology and data

The source of patent application information for the Valencian Community is the database of the Spanish Patent and Trademark Office (OEPM), which allows us to select according to 'province', and includes codes for the three Valencian NUTS 3 regions: Alicante, Castellon and Valencia. We recovered patents from 1999 to 2003.[4]

---

[4] OEPM regularly updates online data. The date of the extractions used in this paper is 23 June 2006.

We chose patents identified as firm patents rather than including patents from all performance sectors, to allow for some institutional homogeneity and a focus on knowledge-industry interaction, and also because they represent the bulk of patenting activity. The OEPM database includes information on name(s) of patent applicant. Identifying their correspondence with firms is not straightforward and involved checking each patent individually. We classified the 1,382 patents registered for the Valencian Community between 1999 and 2003, distinguishing between firm and other types of applicants, based on an the acronyms SL (limited society) and SA (anonymous society) attached to the name.

We constructed a database to include citations for these firm patents. We studied the full-text of the patent application forms for every patent, especially the description field where the applicant includes the prior art, and counted the numbers of citations.[5] We conducted the same exercise for citations in the prior art report (included by patent examiners). This task was made easier since some of the citations already appeared in the OEPM database. This resulted in 712 applicant citations and 2,849 examiner citations, which we classified further according to the following characteristics:

- year of application: from 1999 to 2003;
- route of protection: there are three legal routes to protection of an invention in Spain included in the OEPM database: national, European, and PCT. Because of the few

[5] In the full-text of the application form, applicants provide descriptions of their inventions to demonstrate their novelty, to describe them and to explain how they were made. Although most include a section on prior art, it is not obligatory to include citations. Consequently, a patent with no citations does not mean there is no involvement of prior art, but only that it is implicitly referred to.

number of patents that apply for EPO protection, we combined them with PCT applications under the label 'international applications';[6]

- ownership: in the case of patents with several applicants where at least one is a firm, we defined cooperation type: with a firm, with a research centre, with an innovation/technology centre, and with an individual.

- NUTS 3 region: there are three regions in the Valencian Community –Alicante, Castellon and Valencia;

- technology class: each of the eight sections at the first level of the International Patent Classification (IPC), calculated through a fractional count.

We also attributed patents to economic sectors by linking them to the Analytical System on Spanish Balance Sheets (SABE). This database, which includes economic data on Spanish firms, includes a field for the Statistical Classification of Economic Activities (NACE), revision 1.1. We were able to match 92% of the firm applicants in our sample to the firms included in SABE and assigned to each of these patents the two-digit NACE activity code of the applicant. This allowed us to construct the following variable:

- economic sector: two-digit NACE activity of applicant firm, calculated with a fractional count, which provides information included in the last three columns of Table 1.

For each variable, we calculate whether the difference between numbers of applicant and examiner citations is significant, and whether there is a significant difference between variable categories using ANOVA and the non-parametric Kruskal-Wallis ANOVA.

---

[6] Note that PCT does not award patents: applications are subject only to international review and then filed at national patent offices, i.e. PCT applications become national (or not, if they are abandoned).

For the analysis, we divided the sample into sub-samples - with and without applicant citations. The comparison between the full sample and the sub-sample with applicant citations allows us to test some of our hypotheses. Of course, this implies that applicants that disclose some knowledge (i.e. include at least one applicant citation) hide less knowledge than applicants who do not disclose any knowledge (i.e. do not include any applicant citations). Although it is beyond the scope of this study to verify this, we provide some evidence that this is a reasonable assumption.

Results

There is generally not much relation between total number of patents and GVA in Table 1. The exceptions are 'wholesale and retail trade' and 'real estate, renting and business activities', which have a high share of both patents and GVA. Other large sectors, in terms of GVA ('construction', 'non-market services', 'transport, storage and communication'), do not have high shares of patents, while other active patenting sectors do not show high shares of GVA ('rubber and plastic products', 'machinery and equipment n.e.c.', 'chemicals and chemical products'). This suggests that the technological structure of the economy is determined by the type of sectoral innovation system rather than by type of industry.

The knowledge base, proxied by examiner citations, follows the number of patents distribution, but the relation to GVA is slightly larger –mainly due to the central categories of both distributions.

The share of applicant citations follows the share of patents, but is less related to the composition of GVA, because of its high concentration in science-based sectors. Thus, there is a higher number of examiner than applicant citations in sectors with high GVA such as 'wholesale and retail trade', while the opposite is true for low GVA sectors such as

John Wiley & Sons, Inc.

‘chemicals and chemical products’. This suggests that applicants in the science-based sectors tend to provide knowledge. This point is further addressed in the following section.

Are examiner citations based on administrative requirements? No, they reflect the technoeconomic specialisation

Table 2 shows that the average number of examiner patent citations decreases over time. So does the frequency of mode. This may be due to harmonisation with European standards, which tend towards fewer but more relevant citations, and has been encouraged since 1991 when the EPO gave the OEPM the responsibility for providing search reports for international patent applications.

With the exception of time effects, other inflationary characteristics (route of protection and ownership regime) do not influence the number of examiner citations.

There is also some regional variation: in the NUTS 3 region of Alicante, which has the lowest per capita income, patents include more examiner citations compared to the other two regions. This is because the technologies and sectors that include more examiner citations are present in Alicante.

{Table 2 around here}

Technological variation is also present and deserves some comment. On the one hand, the number of examiner citations is low in science-based technologies such as C. Chemistry; Metallurgy, G. Physics and H. Electricity. On the other hand, some non-science-based categories also have high numbers of average citations: A. Human Necessities, D. Textiles; Paper, B. Performing Operations and E. Fixed Constructions.

Table 3, by economic sector, shows that there is significant variation and, that, although differences across categories are not as clear-cut as in the case of technology classes, science-based sectors do not present many more examiner citations per patent than other sectors.

{Table 3 around here}

Overall, Hypothesis 1 is supported: examiner citations vary according to the distribution of patent characteristics. The pattern of this variation is interesting: (i) except for time effects, administrative reasons do not seem to be influential; (ii) less science-intensive technologies have fewer examiner citations, and for economic sectors, the opposite is not true. Hence, examiners are paying attention to the technoeconomic structure of the economy.

Do examiners trust applicants? No, they correct by region, technology class and economic sector

If we take single patents as the unit of observation, the correlation coefficient between number of applicants and examiner citations is close to zero. That is, there is no relation between the numbers of citations included by the two parties. This means that Hypothesis 2 is not supported. To dig deeper into the source of differences among patents grouped by characteristics, we focus on applicant citations (2nd last column in Table 2).

In terms of patent characteristics we find that the pattern of applicant citations is different from that of examiner citations: on the one hand, the average number of citations over time does not decrease significantly; on the other hand, applicants tend to include more citations if the patent application is international and has co-applicants.

In terms of technoeconomic characteristics, applicant citations and examiner citations are similar in that technological and economic heterogeneity and more especially regional variation are significant. However, the sources of variation are very dissimilar between examiners and applicants.

Castellon, the region with the smallest numbers of examiner citations, has the highest numbers of applicant citations. The region with the largest numbers of examiner citations, Alicante, ranks second for applicant citations. Thus, from the regional distribution of citations, we can say that to an extent, examiners complement applicant citations.

In the case of technologies, there are significant differences between examiner and applicant citations in all classes. The largest differences are in non science-based technologies, where numbers of applicant citations are much smaller compared to examiners' (D. Textiles; Paper, B. Performing Operations; Transporting and A. Human Necessities). The smallest differences are in some science-based technologies (clear in the case of C. Chemistry; Metallurgy; to a lesser extent, in the case of H. Electricity). Thus, to an extent, examiners counterweight the concentration of applicant citations in a few technologies.

These findings about complementarity are even clearer in the case of economic sectors (Table 3). Here, we can see significant differences between examiner and applicant added citations in almost all categories. In the supplier-dominated and production-intensive sectors, examiners cite more when applicants cite less, e.g. in Other supplier-dominated sectors, Manufacture of machinery and equipment n.e.c. and Manufacture of wood, paper, publishing, media. This inverse relation appears also in Manufacture of chemicals and chemical products, where most applicant citations are concentrated, and numbers of examiner citations are small. However, there are exceptions, such as Manufacture of electrical machinery and equipment n.e.c., a science-base sector with large differences between examiner and applicant citations.

These differences would seem to be the source of the overall lack of correlation between applicant and examiner citations and, overall, do not support Hypothesis 2.

Note that intergroup variation is always significant because examiners include more citations than do applicants – which is logical since the sample includes patents with no applicant citations. If we exclude this group, the results for intergroup variation are more interesting.

Do examiners behave differently if applicants include citations to disclose knowledge? No, they behave similarly

Table 4 presents the results for the sub-sample of patents with at least one applicant citation. The column labelled "average number of examiner citations" describes whether, compared to the columns with the same name in Table 2, examiners' behaviour is the same, irrespective of whether or not applicants include citations.[7]

{Table 4 around here}

In Table 4, we can see that examiners tend to include fewer citations over time than in Table 2, but that there is no significant variation between routes of protection; patents from Alicante include more examiner citations, similar to the numbers in non-science-based technologies and supplier dominated sectors (Table 5). This reflects the position for examiner citations for the overall sample.

{Table 5 around here}

The difference is for co-applications: they are not significant in the overall sample but show more examiner citations in the sub-sample. However, the last column in Table 5 indicates that

[7] Table 4 excludes technology classes with fewer than 10 patents.

variations between examiner and applicant citations are not significant, suggesting that intragroup variation is not important.

The findings indicate that, in general, examiners behave similarly regardless of whether or not applicants include citations, which does not support Hypothesis 3.

Are applicants who disclose some knowledge representative of most applicants? No

Even if we focus on applicants that include at least one citation, the correlation with numbers of examiner citations is not significantly different from zero. However, as the column labelled "average number of applicant citations" in Table 4 shows, there are more similarities with examiners than in the full sample.

The average number of applicant citations for patents in Table 4 decreases significantly over time and we do not observe differences among route of protection or ownership regime. These are the same as the results for examiner citations in the full sample. For the sub-sample, intergroup variation does not tend to be significant. The differences between examiners and applicants are small along the patent characteristics mentioned above.

The differences that do exist are based on technoeconomic reasons. Among technology classes, there is significant intra- and intergroup variation: the highest difference is C. Chemistry; Metallurgy, where examiners introduce fewer citations than applicants. In H. Electricity and in B. Performing Operations; Transporting, the reverse is true. Therefore, there are indications that examiners try to complement applicants' knowledge disclosure. The findings are similar for economic sectors (Table 5).

Comparison between applicant citations in the full sample and the sub-sample is interesting. The number of citations in the full sample (Table 2) is stable over time, but increases for international applications and applications with more than one applicant. Table 4 shows that the number of citations (in the sub-sample of patents with at least one applicant citation) significantly decreases, and does not vary based on route of protection or a positive number of co-applicants.

This does not support Hypothesis 4. Applicants that include at least one citation bias the emergent statistics on applicant citations, because they inflate the number of citations for reasons such as route of protection and ownership regime.

An unforeseen side result –a fixed number of examiner citations

The above results suggest that examiners merely replicate bodies of knowledge from patent to patent, ignoring the citations included by applicants. To verify this, we matched pairs of common examiner citations in different patents but did not find huge correspondence (less than 30%). However, the number of examiner citations in these patents was often the same. Therefore, it seems that examiners tend to add a 'fixed' number of citations, although these citations may be different. This contradicts the claim of replication of bodies of knowledge, but imply the existence of a standard practice in terms of number of citations. This result deserves further research.

Conclusions

The results in this paper suggest that it would be premature to defend the use of examiner citations in patents to justify research funding or evaluate performance. Our results also do not support the hypothesis that applicant citations are better sources of information. However, the results support the idea that patent citations have different meanings. We develop these ideas in the next paragraphs.

Examiner citations are related to the technoeconomic structure of the territory. If the local industry has very few leading and patenting sectors, there will be fewer examiner citations. For other sectors, even though the knowledge base exists, examiner patent citations will not capture it. Therefore, studies on examiner citations should consider the technoeconomic structure. Examiner citations may be representative of the knowledge base where there is strong industrial specialisation in highly patenting or leading sectors but require complementary analysis in every other case.

This paper looked at full-text applicant citations, which highlighted some practices of patent examiners related to adding citations, namely to include more (less) references than applicants, in patents for technology classes and economic sectors that rely less (more) on the science base. The result is a more homogeneous distribution of citations that masks the scarce importance of codified knowledge for traditional economic activities.

If we take into account the fact that examiners also assign IPC codes to patents (Kang et al., 2007), the 'noise' in citations grows exponentially. However, we also show that using applicant citations creates problems, since applicants may add large numbers of (hardly justified) citations in some patents.

Overall, this makes a case for increasing resource allocation to patent examiners' to allow for case-by-case search reports, which would increase applicants' incentives to disclose information to the EPO and the World Intellectual Property Organization -WIPO (and avoid artificial distortions among types of patents), and provide better legal standards for storing information, especially for analysing full-text applicant citations.

Many of the explanations for parts of the evidence from this study are related to the conduct of examiners and applicants. Interviews would be useful to check the consistency of explanations. Ongoing work suggests that patent examiners' personal characteristics also determine the inclusion of citations (Lemley & Sampat, 2008).

However, we would propose investigation of the differences between examiner and applicant citations as indicators of trust in the research system. For example, in Table 4 the case of technology classes: for E. Fixed Constructions, the difference is 0, possibly implying that examiners trust applicants; for C. Chemistry; Metallurgy, the difference in absolute value is equal to 2, possibly implying that examiners do not trust applicants. We would expect that the more developed the research system, the more often this indicator will take a zero value.

To what extent these results are idiosyncratic of the sample we analysed, which is based on a region with low absorptive capacity, is questionable. Elsewhere (Azagra et al., 2009), we emphasise those properties of the sample that the regional context influences more clearly. Here, we want to stress that, although some of our findings may be idiosyncratic, the importance of the hypotheses and their methodological implications should guide the debate on patent citations.

References

Acosta, M. & Coronado, D. (2002). Las relaciones ciencia-tecnología en España. Evidencias a partir de las citas científicas en patentes. Economía Industrial, 346, 27-46.

Acosta, M. & Coronado, D. (2003). Science-technology flows in Spanish regions. An analysis of scientific citations in patents. Research Policy, 32, 1783-1803.

Alcácer, J. & Gittelman, M. (2006). Patent citations as a measurement of knowledge flows: The influence of examiner citations. Review of Economics and Statistics, 88(4), 774-779.

Allison, J.R. & Sager, T.W. (2007). Valuable Patents Redux: On the Enduring Merit of Using Patent Characteristics to Identify Valuable Patents. Texas Law Review, 85, 1769-1797.

Azagra-Caro, J.M., 2007. What type of faculty member interacts with what type of firm? Some reasons for the delocalisation of university-industry interaction. Technovation, 27, 704-715

Azagra-Caro, J.M., Fernández de Lucio, I., Perruchas, F. & Mattsson, P. (2009). What do patent examiner inserted citations indicate for a region with low absorptive capacity? Scientometrics, 80(2), 441–455.

Azagra-Caro, J.M., Yegros-Yegros, A. & Archontakis, F. (2006). What do university patent routes indicate at regional level? Scientometrics, 66(1), 219-230.

Breschi, S. & Malerba, F. (1997). Sectoral Innovation Systems: Technological Regimes, Schumpeterian Dynamics, and Spatial Boundaries, C. Edquist (ed.): Systems of Innovation: Technologies, Institutions and Organizations, ch. 6. London and Washington: Pinter.

Callaert, J., van Looy, B., Verbeek, A., Debackere, K. & Thus, B. (2006). Traces of Prior Art: An analysis of non-patent references found in patent documents. Scientometrics, 69(1), 3-20.

Chen, C. (2003). Book Review: Patents, Citations & Innovations: A Window on the Knowledge Economy. Journal of the American Society for Information Science and Technology, 54(8), 802–803.

Criscuolo, P. (2006). The ‘home advantage’ effect and patent families. A comparison of OECD triadic patents, the USPTO and the EPO. Scientometrics, 66(1), 23–41.

Criscuolo, P. & Verspagen, B. (2008). Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. Research Policy, 37, 1892-1908.

EC (2007). Commission Green Paper ‘The European Research Area: New Perspectives’. COM(2007), 161.

Guan, J.C. & Gao, X. (2009). Exploring the h-Index at Patent Level. Journal of the American Society for Information Science and Technology, 60 (1), 35-40.

Hall, B.H., Jaffe, A. & Trajtenberg, M. (2000). Market value and patent citations: A first look. NBER Working Paper, 7441.

Hassan, E. (2003). Simultaneous Mapping of Interactions between Scientific and Technological Knowledge Bases: The Case of Space Communications. Journal of the American Society for Information Science and Technology, 54(5), 462-468.

Jaffe, A. & Trajtenberg, M. (1996). Flows of knowledge from universities and federal laboratories: Modeling the flow of patent citations over time and across institutional and geographic boundaries. PNAS, 93, 12671-12677.

Jaffe, A., Trajtenberg, M. & Fogarty, M.S. (2000). The meaning of patent citation: Report on the NBER/Case-Western reserve survey of patentees. NBER Working Paper, 7631.

Kang, I.S, Na, S.H., Kim, J. & Lee, J.H. (2007). Cluster-based patent retrieval. Information Processing & Management, 43(5), 1173-1182.

Kostoff, R.N. (1994). Federal Research Impact Assessment: State-of-the-Art. Journal of the American Society for Information Science and Technology, 45(6), 428-440.

Lai, K.K. & Wu, S.J. (2005). Using the patent co-citation approach to establish a new patent classification system. Information Processing and Management, 41, 313–330.

Lemley, M. & Sampat, B.N. (2008). Examiner Characteristics and the Patent Grant Rate, mimeo: http://www.nber.org/~confer/2008/si2008/IPPI/lemley.pdf, last access: 13/01/2009.

Leydesdorff, L. (2004). The University–Industry Knowledge Relationship: Analyzing Patents and the Science Base of Technologies. Journal of the American Society for Information Science and Technology, 55(11), 991-1001.

Leydesdorff, L. (2008). Patent Classifications as Indicators of Intellectual Organization. Journal of the American Society for Information Science and Technology, 59(10), 1582-1597.

Merrill, S.A., R.C. Levin & M.B. Myers (eds) (2004). A Patent System for the 21st Century. Washington, DC: The National Academies Press.

Meyer, M. (2000). What is special about patent citations? Differences between scientific and patent citations. Scientometrics, 49(1), 93-123.

Michel, J. & Bettels, B. (2001). Patent citations analysis: A closer look at the basic input data from patent search reports. Scientometrics, 51(1), 185-201.

Narin, F. & Noma, E. (1985). Is technology becoming science? Scientometrics, 7(3-6), 369-381.

Noma, E. & Olivastro, D. (1984). Are there enduring patents? Journal of the American Society for Information Science and Technology, 36(5), 297 – 301.

Pavitt, K. (1984): Sectoral patterns of technical change: Towards a taxonomy and a theory. Research Policy 13, 343-373.

Sampat, B. (2004). Examining Patent Examination: An Analysis of Examiner and Applicant Generated Prior Art, Working Paper, School of Public Policy, Georgia Institute of Technology.

Stock, M. & Stock, W.G. (2006). Intellectual Property Information: A Comparative Analysis of Main Information Providers. Journal of the American Society for Information Science and Technology, 57(13), 1794-1803.

Tijssen, R.J.W. (2001). Global and domestic utilization of industrial relevant science: patent citation analysis of science–technology interactions and knowledge flows. Research Policy, 30, 35–54.

Thompson, P. (2006). Patent citations and the geography of knowledge spillovers: Evidence from inventor- and examiner-added citations. Review of Economics and Statistics, 88(2), 383-388.

Tables

TABLE 1. Industrial structure, technological structure and knowledge base of the Valencian Community.

Description	NACE	Share of total gross valued added	Share of total patents	Share of total number of examiner citations	Share of total number of applicant citations
Supplier-dominated sectors					
Agriculture, hunting and forestry	01,02,05	3%	1%	1%	0%
Textiles, textile products, leather and leather products	17,18,19	3%	2%	2%	0%
Wood and wood products	20	1%	4%	4%	0%
Pulp, paper and paper products; publishing and printing	21,22	1%	4%	3%	1%
Rubber and plastic products	25	1%	10%	9%	5%
Manufacturing n.e.c.	36,37	2%	3%	4%	1%
Construction	45	9%	2%	2%	1%
Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods	50,51,52	13%	13%	12%	6%
Hotels and restaurants	55	7%	0%	0%	0%
Transport, storage and communication	60,61,62,63,64	9%	0%	0%	0%
Financial intermediation	65,66,67	5%	0%	0%	0%
Non-market services	75,80,90,95,99	12%	0%	0%	0%
Health and social work	85	3%	0%	0%	0%
Other community, social and personal service activities	91,92,93	3%	0%	6%	4%
Production-intensive sectors					
Food products, beverages and tobacco	15,16	2%	1%	1%	5%
Other non-metallic mineral products	26	4%	5%	4%	6%
Basic metals and fabricated metal products	27,28	2%	6%	5%	2%
Machinery and equipment n.e.c.	29	1%	12%	13%	5%
Transport equipment	34,35	2%	0%	0%	0%
Science-based sectors					
Mining and quarrying; coke, refined petroleum products and nuclear fuel	10-14, 23	0%	1%	1%	2%
Chemicals and chemical products	24	1%	11%	10%	37%
Electrical and optical equipment	31,32,33	1%	7%	6%	5%
Electricity, gas and water supply	40,41	2%	0%	1%	1%
Real estate, renting and business activities	70,71,72,73,74	13%	19%	16%	20%
Total		100%	100%	100%	100%

TABLE 2. Frequency of citations in patents from the Valencian Community with firm applicants – full sample.

Variable	Category	Nº of patents	Average number of examiner citations (mode; frequency of mode)	Average number of applicant citations (mode; frequency of mode)	Significance of intergroup variation
Year of application	1999	86	5.87 (5; 23%)	1.26 (0; 81%)	*
	2000	114	4.82 (4; 25%)	1.82 (0; 72%)	*
	2001	129	4.64 (4; 24%)	0.95 (0; 74%)	*
	2002	133	4.71 (3 and 4; 21%)	1.05 (0: 65%)	*
	2003	109	4.40 (4; 24%)	0.83 (0; 69%)	*
	Significance of the growth rate		*	n.s.	
Route of protection	National	459	4.81 (4; 24%)	1.02 (0; 74%)	*
	International	112	4.91 (5; 22%)	1.81 (0; 63%)	*
	Significance of intragroup variation		n.s.	*	
Ownership	No co-applicants	545	4.80 (4; 22%)	1.10 (0; 73%)	*
	With co-applicants	26	5.38 (4; 27%)	2.62 (0; 46%)	*
	Significance of intragroup variation		n.s.	*	
NUTS 3 region	Alicante (1)	148	5.32 (4; 21%)	1.53 (0; 70%)	*
	Castellon (2)	99	4.65 (5; 25%)	1.81 (0; 66%)	*
	Valencia (3)	324	4.66 (4; 23%)	0.81 (0; 74%)	*
	Significance of the ratio (3)/(1)		*	*	
	Significance of the ratio (3)/(2)		n.s.	*	
Technology class (IPC section)	A. Human Necessities	152	4.48 (3; 17%)	1.18 (0; 72%)	*
	B. Performing Operations; Transporting	160	4.06 (3; 18%)	0.55 (0; 83%)	*
	C. Chemistry; Metallurgy	52	3.16 (3; 24%)	2.41 (0; 54%)	*
	D. Textiles; Paper	18	4.24 (4; 21%)	0.43 (0; 75%)	*
	E. Fixed Constructions	74	4.02 (4; 19%)	1.26 (0; 67%)	*
	F. Mechanical Engineering; Lighting; Heating; Weapons; Blasting	32	3.27 (2 and 3; 18%)	0.19 (0; 87%)	*
	G. Physics	25	3.20 (3; 24%)	0.17 (0; 82%)	*
	H. Electricity	59	3.48 (4; 22%)	0.82 (0; 61%)	*
	Significance of intragroup variation		*	*	

n.s. = not significant

TABLE 3. Frequency of citations in patents from the Valencian Community with firm applicants – sample of firms assigned to economic sectors.

Variable	Category	Nº of patents	Average number of examiner citations (mode; frequency of mode)	Average number of applicant citations (mode; frequency of mode)	Significance of intergroup variation
Economic sector	Supplier-dominated sectors				
	Manufacture of wood, paper, publishing, media (NACE 20, 21, 22)	44	4.80 (4; 29%)	0.16 (0; 86%)	*
	Manufacture of rubber and plastic products (NACE 25)	52	5.06 (5; 21%)	0.69 (0; 71%)	*
	Trade, maintenance and repair (NACE 50, 51, 52)	70	4.33 (4; 27%)	0.38 (0; 78%)	*
	Other business activities (NACE 70, 71, 74)	83	4.23 (4; 23%)	1.24 (0; 59%)	*
	Other supplier-dominated sectors (NACE 01, 17, 18, 19, 36, 45, 63, 85)	41	5.40 (4; 28%)	0.28 (0; 79%)	*
	Production-intensive sectors				
	Manufacture of other non-metallic mineral products (NACE 26)	27	4.22 (3 and 4; 21%)	1.29 (0; 75%)	*
	Manufacture of fabricated metal products, except machinery and equipment (NACE 28)	30	4.60 (4; 23%)	0.53 (0; 80%)	*
	Manufacture of machinery and equipment n.e.c. (NACE 29)	66	5.05 (5; 22%)	0.35 (0; 81%)	*
	Other production-intensive sectors (NACE 15, 33, 34)	26	4.62 (4; 19%)	1.88 (0; 62%)	*
	Science-based sectors				
	Manufacture of chemicals and chemical products (NACE 24)	58	4.69 (5 and 3; 22%)	4.47 (0; 38%)	n.s.
	Manufacture of electrical machinery and equipment n.e.c. (NACE 31)	10	4.50 (4; 30%)	0.20 (0; 90%)	*
	R&D and computer activities (NACE 72, 73)	19	5.47 (5; 26%)	2.00 (0; 74%)	*
	Other science-based sectors (NACE 14, 23, 32, 40)	15	4.57 (4; 29%)	1.74 (0; 65%)	*
	Significance of intragroup variation		*	*	

TABLE 4. Frequency of citations in patents from the Valencian Community with firm applicants – sample of patents with at least one applicant citation.

Variable	Category	Nº of patents	Average number of examiner citations (mode; frequency of mode)	Average number of applicant citations (mode; frequency of mode)	Significance of intergroup variation
Year of application	1999	16	6.00 (5; 31%)	6.75 (3; 31%)	n.s.
	2000	32	5.03 (4; 34%)	6.50 (1; 22%)	n.s.
	2001	34	4.56 (5; 23%)	3.62 (1; 35%)	n.s.
	2002	46	4.72 (3; 24%)	3.04 (1; 33%)	n.s.
	2003	34	4.18 (4; 38%)	2.65 (1; 44%)	*
	Significance of the growth rate		*	*	
Route of protection	National	120	4.68 (3; 22%)	3.88 (1; 34%)	n.s.
	International	42	4.98 (4; 33%)	4.83 (1; 26%)	n.s.
	Significance of intragroup variation		n.s.	n.s.	
Ownership	No co-applicants	148	4.64 (4; 24%)	4.06 (1; 32%)	n.s.
	With co-applicants	14	6.00 (4; 29%)	4.86 (1; 29%)	n.s.
	Significance of intragroup variation		*	n.s.	
NUTS 3 region	Alicante (1)	45	5.18 (4 and 6; 22%)	5.04 (2; 31%)	n.s.
	Castellon (2)	34	5.06 (4 and 5; 26%)	5.26 (1; 32%)	n.s.
	Valencia (3)	83	4.41 (3; 25%)	3.17 (1; 36%)	*
	Significance of the ratio (3)/(1)		*	*	
	Significance of the ratio (3)/(2)		n.s.	*	
Technology class (IPC section)	A. Human Necessities	42	4.52 (4; 17%)	4.34 (1; 27%)	n.s.
	B. Performing Operations; Transporting	37	3.58 (3;31%)	2.18 (1; 33%)	*
	C. Chemistry; Metallurgy	25	3.49 (3; 20%)	5.29 (1; 23%)	*
	E. Fixed Constructions	23	4.20 (4; 23%)	3.82 (2; 23%)	n.s.
	H. Electricity	23	3.25 (4; 31%)	2.16 (1; 35%)	*
	Significance of intragroup variation		*	*	

n.s. = not significant

TABLE 5. Frequency of citations in patents from the Valencian Community with firm applicants – sample of patents with at least one applicant citation and firms assigned to economic sectors.

Variable	Category	Nº of patents	Average number of examiner citations (mode; frequency of mode)	Average number of applicant citations (mode; frequency of mode)	Significance of intergroup variation
Economic sector	Supplier-dominated sectors				
	Manufacture of rubber and plastic products (NACE 25)	15	5.13 (5; 27%)	2.40 (1 or 2; 33%)	*
	Trade, maintenance and repair (NACE 50, 51, 52)	14	4.66 (4; 37%)	1.74 (1; 44%)	*
	Other business activities (NACE 70, 71, 74)	34	3.79 (4; 26%)	3.03 (2; 35%)	n.s.
	Production-intensive sectors				
	Manufacture of machinery and equipment n.e.c. (NACE 29)	11	4.04 (4 and 5; 23%)	1.83 (1; 38%)	*
	Other production-intensive sectors (NACE 15, 33, 34)	10	4.50 (4; 30%)	4.90 (3; 40%)	n.s.
	Science-based sectors				
	Manufacture of chemicals and chemical products (NACE 24)	36	4.69 (4; 28%)	7.19 (1; 19%)	*
	Significance of intragroup variation		*	*	

n.s. = not significant