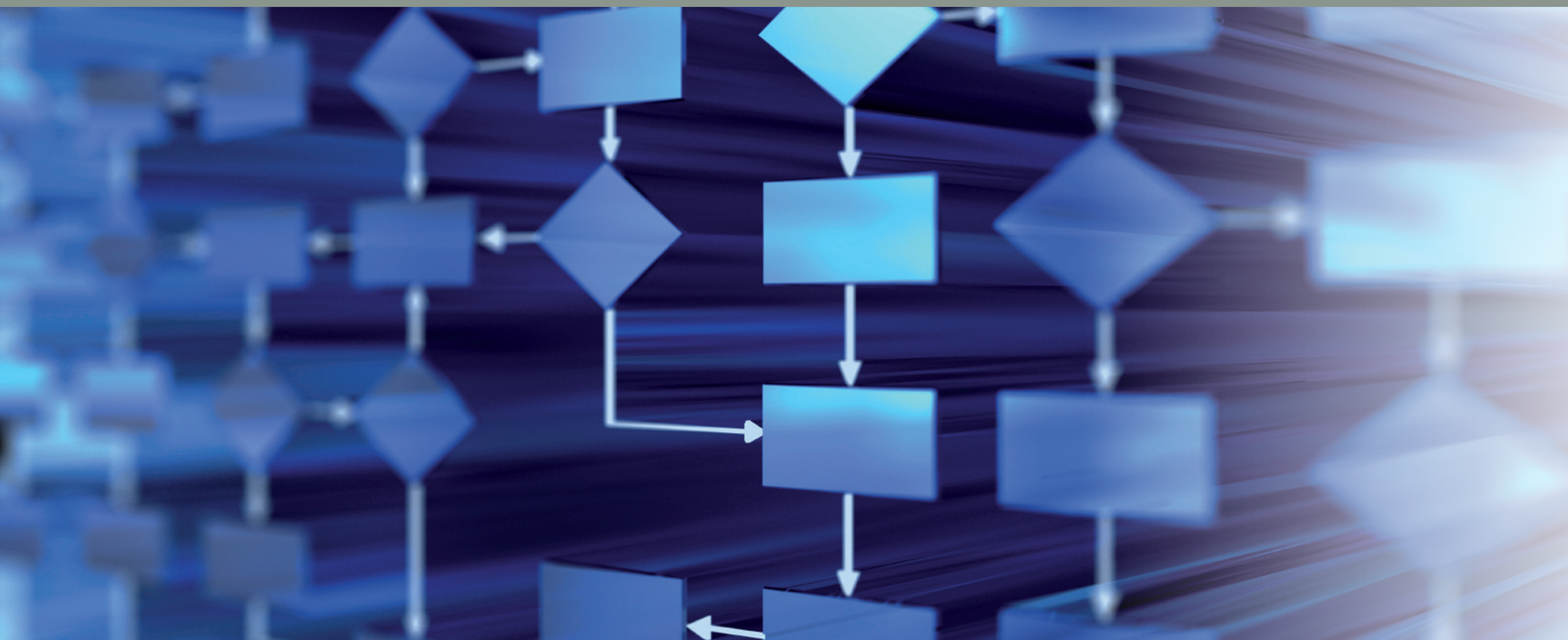
Economic Research Working Paper No. 33

Economics & Statistics Series

Identifying the gender of PCT inventors

Gema Lax Martinez
Julio Raffo
Kaori Saito



November 2016

WIPO
WORLD INTELLECTUAL PROPERTY ORGANIZATION

Identifying the Gender of PCT inventors

Gema Lax Martínez*, Julio Raffo**, Kaori Saito**

Abstract

This paper analyzes the gender of inventors in international patent applications. We compile a worldwide gender-name dictionary, which includes 6.2 million names for 182 different countries to disambiguate the gender of PCT inventors. Our results suggest that there is a gender imbalance in PCT applications, but the proportion of women inventors is improving over time. We also find that the rates of women participation differ substantially across countries, technological fields and sectors.

JEL Classification: J16, O31, O32, O34

Key Words: Patent gender gap; Gender innovation metrics; Patents; Gender-name dictionary.

Disclaimer

The views expressed in this paper are those of the authors, and do not necessarily reflect the views of the World Intellectual Property Organization or its member states.

Acknowledgements

The authors are thankful and indebted to the inputs, support and kind review of the following list of people (alphabetically sorted): Carsten Fink, Cornelia Moussa, Ernest Miguelez, Ernesto Fernandez-Polcuch, Grégory Bos, Hanna Kang, Hao Zhou, Intan Hamdan-Livramento, Juan Yang, Kritika Saxena, Kyle Bergquist, Lili Cheng, Maite Aragonés Lumeras, Matt Bryan, Mosahid Khan, Young-Woo Yun, Yukio Yanagimoto, and, Zhilong Yu. We are also thankful to all WIPO Gender focal points, and the participants of two WIPO internal seminars discussing preliminary results of this paper.

1 Introduction

Women contribute to all fields of creativity and intellectual endeavors, yet they remain underrepresented in many of these areas. Women are less often employed than men, including in advanced economies such as the US, Japan and in all of the EU-28 Member Countries (Eurostat, 2015). On average, women also earn less and are employed fewer hours than men in every EU-28 country (Eurostat, 2016). Moreover, women are less likely to appear as authors in scientific publications and inventors in patents than men (Frietsch et al., 2009).

In recognition of the prevalence of this gender imbalance, the 193-members of the United Nations General Assembly adopted the 2030 Agenda for Sustainable Development, which came into force on January 1, 2016.[1] The Agenda underscores that gender equality and the empowerment of women and girls would contribute to progress across all the Sustainable Development Goals and targets.

As a specialized agency of the United Nations, the World Intellectual Property Organization (WIPO) is committed to promoting gender equality in the field of intellectual property and has taken steps to raise the profile of gender equality within the day-to-day running of the Organization. One of these steps includes the disaggregation of intellectual property (IP) data by gender as a key performance indicator for policies aimed at promoting innovation and creativity and spurring economic, social and cultural development.

The role of gender in economic outcomes is not a new revelation. For some time, evidence has clearly indicated the existence of gender gaps in pay scale, educational attainment and labor participation rates. In addition, in some areas this gender gap only appears to be worsening throughout the career of an individual.

The reasons for why such gender gap exists have been difficult to isolate (Mickelson, 1989; Jacobs, 1996; Huyer, 2016). In particular, studies pointed to the need for more gender disambiguated data (Mauleón and Bordons, 2006; Frietsch et al., 2009; Naldi and Parenti, 2002b). Our paper tries to fill this gap by analyzing the gender of approximately 9 million inventors and individual applicants of patent applications filed through the Patent Cooperation Treaty (PCT) System. In order to accomplish this, we compile a World Gender-Name Dictionary (WGND), which includes 6.2 million names for 182 different countries.

Our analysis suggests that the participation of women in PCT applications is improving over time, although it is still far from a balanced distribution. We also find substantial differences across countries, technological fields and sectors.

This paper is organized as follows. In the following section, we review the existing literature related to patent use and gender; and, we explore the possible methodological approaches to be used for gender disambiguation. In section 3, we describe how we built and applied the WGND to the PCT data. In section 4, we

---

* University of Lausanne
** World Intellectual Property Organization

[1] This is part of the UN commitments expressed in the Convention on the Elimination of All Forms of Discrimination against Women (1979), the Beijing Declaration and Platform for Action (1995), Economic and Social Council Agreed Conclusions 1997/2 (A/52/3), Millennium Development Goals (2000) and the UN System-Wide Policy on Gender Equality and the Empowerment of Women (CEB/2006/2).

analyze and discuss the statistical results. Finally, in section 5, we conclude by making final remarks on the scope of the results and potential future steps.

2 Literature review

In this section, we first outline the results from previous studies on gender differences in the scientific and technological (S&T) field, with special focus on those addressing scientific publications and patents. Next, we review the main approaches these studies have used to disambiguate the gender of inventors.

Gender inequality is not necessarily distributed in the same way along women's career path. In an average high income country, both women and men may have equal access to opportunities that would propel their careers forward. For instance, there are more women graduating from bachelor and master programs than men. However this gender distribution changes after this stage. There are less women graduating from PhD programs and even less working as researchers than men (Huyer, 2015). And even if they do, women earn, publish or patent less than men. This gender gap worsens over time. For example, the proportion of women using the patent system remains low compared to the proportion of scientific papers they publish each year (Frietsch et al., 2009). Scholars refer to this pattern of change in gender distribution as the "leaky pipeline" where the gender gap tends to be increasingly worse (Huyer, 2015).

A major part of the gender studies in the S&T field have looked at the scientific and technological productivity gap (Cole and Cole, 1973; Cole and Zuckerman, 1984; Fox, 2005; Levin and Stephan, 1998; Zuckerman, 1987, 2001). In particular, their lower position in the academic hierarchy seems to explain in part the observed lower productivity (Long, 2001; Xie and Shauman, 1998). Whittington and Smith-Doerr (2005) believe that an overall lack of encouragement for women in academia and industry explains some of the productivity gap and the lower women's position in the academic hierarchy. Some studies have suggested that although their total S&T productivity might be lower, women's marginal productivity tends to be higher. When using patent citations as a measure of S&T impact, empirical studies have found that even though women patent less than men, the quality and impact of their patents are equal or better than those of men (McMillan, 2009; Whittington and Smith-Doerr, 2005).

Working environments and structure also appear to explain some of the S&T productivity gap observed for women. Women working in firms with hierarchical organizational structures are less likely to participate in patenting activities than men. Studies show that women in flatter and more flexible firms – such as those in the biotechnology or life sciences industry – are more likely to patent than in more hierarchical kind of firms (Whittington and Smith-Doerr, 2008; Jung and Ejermo, 2014; Eaton, 1999). Studies that look at the composition of inventors within patent applications find that women are more likely to be part of larger research groups when inventing than men (Naldi et al, 2015; Busolt et al., 2008; Moody and Light, 2006; Moody, 2004). In the same vein, women are less likely to be single authors or inventors (Naldi et al, 2015). Furthermore, recent studies have explored the role of women in research groups, finding that they may have given up patent inventorship in order to secure scientific authorship, even when entitled to both (Lissoni et al, 2013).

Another explanation for the S&T production gap may be due to the industries women participate more in. Eaton (1999) shows that the biotechnology field offers more opportunities to women. In contrast, some of these studies find that women's representation in electrical and mechanical engineering tend to be significantly lower relative to life sciences and higher in academia than industry (Naldi et al, 2015; Jung and Ejermo, 2014; Hunt et al, 2012; Whittington and Smith-Doerr, 2005). This may impact the productivity gap observed overall as a given sector has different productivity rates and different patenting rates.

At the country level, most gender patenting studies that provide aggregate statistical analyses indicate rising women participation rates in patenting (Ding et al., 2006; Frietsch et al., 2009; Jung and Ejermo, 2014).[2] Frietsch et al. (2009) also find progress in countries where women participation in scientific publications has stagnated. Moreover, many of these studies also show substantial differences in women participation across countries (Frietsch et al, 2009; Naldi et al, 2015; UKIPO, 2016a, 2016b). Interestingly, these differences do not appear to be directly correlated with typical socioeconomic indicators, such as GDP per capita or women participation in the labor market.

2.1 How can we obtain IP data with gender breakdown?

Several gender studies point to the lack of tools to ascertain the gender of the subjects under study (Mauleón and Bordons, 2006). Frietsch et al. (2009) emphasized that one of the main disadvantages they faced was the inability to identify the gender of some inventors and authors. Similarly, Naldi and Parenti (2002b) faced the limitation of considering only names from six European languages which, given the increase in migration, made names originating from other languages hard to identify.

In principle, there are two main approaches to obtain data with gender breakdown. The most direct one is through primary data collection of asking for the respondent's gender (see Walsh and Nagaoka, 2009). In the case of patent data, this would imply that the patent application form contains gender fields for inventors and applicants. The main advantage of the primary collection is precisely capturing the information directly at the source.[3] Another advantage is that self-declaration can capture more gender diverse categories than just women and men. However, a very important limitation is that the direct approach cannot be applied retroactively. Moreover, it may take considerable time to implement a change in the primary collection method – e.g. the patent application form – which is compounded if needed to be changed in multiple countries.

The second and most commonly used approach is to attribute gender after the primary data collection. There are three main methods for gender attribution.

The first attribution method relates to using a secondary source with gender information and linking it to clearly identified individuals. Women patenting studies have matched inventors with a national register of individuals for which gender has been collected already (e.g. Jung and Ejermo, 2014). If the quality and coverage of the secondary source are good enough, gender attribution can be as good as primary data collection. However, databases which are so easily linked to patents are rare and having access to this type of data for multiple years and countries may be daunting task.

[2] Whittington and Smith-Doerr (2005) find that the gender disparity remains constant for a sample of US life science PhDs.
[3] In the case of patents, it is worth noting that each inventor may not fill the application directly, which is done more often by the applicant or patent attorney.

The second attribution method relates to the semantics of how individuals are named (Park and Yoon, 2007; Tripathi and Faruqui, 2011). For example, name honorific titles – such as "Mrs", "Ms" or "Mr." in English – or names structures – such as ending in "o" for men and "a" for women in Spanish and Portuguese – can unambiguously refer to a given gender in some naming conventions. Unfortunately, many languages do not have clear gender distinction based on semantic rules and databases may not collect and store title information. In addition, migration and changing trends in naming conventions can affect the coverage of this method.

Table 1: Summary of methodological approaches to attribute gender

Method	Type	Advantages	Disadvantages	Examples
Primary data collection	Direct	+ Captures information directly at source + Self-declaration allows more gender diverse categories	– Cannot be applied retroactively – Implementation time – Difficult in multiple countries	Walsh & Nagaoka (2009)
Attribution based on secondary source on individuals' data	Indirect	+ Can be as reliable as primary data if based on unique identifiers. + Self-declaration in secondary source may also allow more gender diverse categories. + Can be applied retroactively if secondary source permits.	– Depends heavily on secondary source coverage. – May be difficult to collect secondary source in multiple countries and years.	Jung & Ejermo (2014)
Attribution based on name gender semantics	Indirect	+ Can be applied retroactively if language or customs permit. + Can be applied to countries sharing the same language conventions	– Depends heavily on quality and coverage of naming rules. – Difficult for languages without clear-cut rules. – Affected by migration and naming trends	Park & Yoon (2007), Tripathi & Faruqui (2011)
Attribution based on name-gender dictionary	Indirect	+ Can be applied retroactively + Can be applied to countries sharing the same naming conventions	– Depends heavily on the dictionary coverage. – Affected by migration and naming trends	Frietsch et al. (2009), Naldi & Parenti (2002a, 2002b), UKIPO (2016a, 2016b)

The third method concerns the use of a list of names with their most commonly associated gender. Such lists are referred to as gender-name dictionaries. Most women patenting studies covering multiple countries have made use of this method (Frietsch et al., 2009; Kugele, 2010; Naldi and Parenti, 2002a, 2002b; Naldi et al, 2005; UKIPO, 2016a, 2016b, amongst other). Indeed, one of the main advantages of attributing gender through the means of a gender-name dictionary is that it can be applied to several countries and retroactively. Of course, as with the other indirect methods, the quality of the gender attribution depends heavily on the quality and coverage of the gender-name dictionary. In addition, migration and changing naming conventions can also affect the coverage of gender-name dictionaries, an issue highlighted already for using language semantics and titles.

All in all, there are advantages and disadvantages of each approach, which makes them complementary rather than strict alternatives. **Table *1*** summarizes the discussed methods. For the needs of this study, the utilization of a gender dictionary was the most appropriate as it allows for the analysis of the most countries and with the longest time periods. With that said, we also utilized the honorific titles method to add an additional source, as we will discuss in the methodological annex.

3 Data and Methods

In this section, we describe how we identified the gender of the applicants and inventors named in patent applications filed under the PCT system.

Using patent data for gender inequality analysis is not new. Previous studies have made use of rich datasets like the one provided by the United States Patent and Trademark Office (Whittington and Smith-Doerr, 2008) or by the European Patent Office (Naldi and Parenti, 2002a, 2002b). Until recently (see UKIPO, 2016b), there has been no worldwide attempt to map gender inequality in patent statistics.

The first challenge we faced in order to attribute gender was to get a gender-name dictionary with worldwide coverage. For this reason, we compiled 13 different sources of gender-name dictionaries, which combined, cover 173 different countries. Most of the sources used for this study come from national public institutions. These are the US Social Security Administration and Census Bureau, the Alberta government, the UK Office for National Statistics, Statistics Sweden, Spain's *Instituto Nacional de Estadística*, France's *Institut National de la Statistique*, and Denmark Statistics. We also relied on lists compiled by previous gender studies, including those by Michael (2007), Tang et al. (2011) and Yu et al. (2014). Likewise, following Sugimoto et al. (2015), we make use of popular names lists by country available through Wikipedia. Finally, we used information extracted from the publically available list of participants in the Assemblies of the Member States of WIPO. In addition to these public sources, we also made use of an ad-hoc list, which was created by Chinese, Indian, Japanese, and Korean WIPO staff native speakers. The final version of our world gender-name dictionary (WGND) contains 6,247,039 unique pairs of names and countries and can be found on the WIPO website.[4]

Patent information was obtained from our PCT system database. For the purpose of our worldwide gender analysis the PCT has a series of advantages. The first advantage concerns its global coverage. The PCT is an international treaty administered by WIPO offering patent applicants an advantageous route for seeking patent protection internationally since it came into force in 1978. To date, there are 151 PCT contracting states. Accordingly, applicants have opted for the PCT route for a significant share of international patent applications. In 2012, nearly 60% of patents seeking international protection were filed through the PCT system (WIPO, 2016).

The PCT patent dataset provides another advantage in comparison to those studies using only national collections as it avoids the concern of home bias. Typically, US residents are more likely to file at the USPTO while Japanese are more likely to file at the JPO, comparing these two countries using only one national collection may bias the results.[5] In this sense, we follow a similar approach as Frietsch et al. (2009), who use EPO data to avoid national collections home bias.

A third and related advantage is that the PCT system applies only one set of procedural rules to applications worldwide, which also implies collecting the applicants and inventors information in a uniform standard. This standard makes for easier cross-country analysis when recording the names of individuals. The quality of the names of inventors and individual applicants is instrumental for the overall quality of the gender attribution exercise. In addition, the PCT system also collects data about inventors and

[4] See http://www.wipo.int/edocs/pubdocs/en/wipo_pub_econstat_wp_33-tech1.zip
[5] For a discussion of home bias in national patent collections see Dernis and Khan (2004).

applicants in non-Romanized spellings, which is critical for gender attribution of Chinese and Korean names (Yu et al, 2014; Park and Yoon, 2007). Last, the PCT data contain information about both the nationality and country of residence of inventors and applicants, which increases the likelihood of attributing a gender to a given observation.[6]

The dataset used for this study contains information on 8,788,617 names of individuals, which refer mostly to inventors, individual applicants or both. Within these data, there are 394,422 unique names[7], which are associated with one country of residence and one nationality. It is also worth noting that 18% of total names are composite names, such as “Mary Jane” or “Jean Pierre”. In addition, our PCT data have 249,795 names in original Chinese or Korean characters.

We applied the WGND to our PCT dataset in different ways in order to maximize attribution. [8] We applied it to full names, to first names only and to second names only. We also made use of the country of residence and the nationality as reference for the WGND. We combined all these different ways in order to obtain 89% of attribution in our PCT data. In addition, we exploited the gender information obtained for composite names in our PCT data in a recursive way (see methodological annex). This resulted in an increase in attribution to 96% of names. Most top PCT filing countries have relatively high attribution scores. Among these, the least complete countries are China (12% unattributed), India (11%), Korea (8%) and Japan (6%). All the other top 20 countries each have less than 5% unattributed observations. As whole, the remaining 198 countries also observe unattributed observations below 5%.

It is worth noting that attribution does not necessarily imply correct attribution. Indeed, our results are likely to contain both false negatives and false positives. Therefore, results are to be interpreted as reporting the most likely gender of each name. We checked manually the results for the most frequent attributed and unattributed names for the main languages.[9]

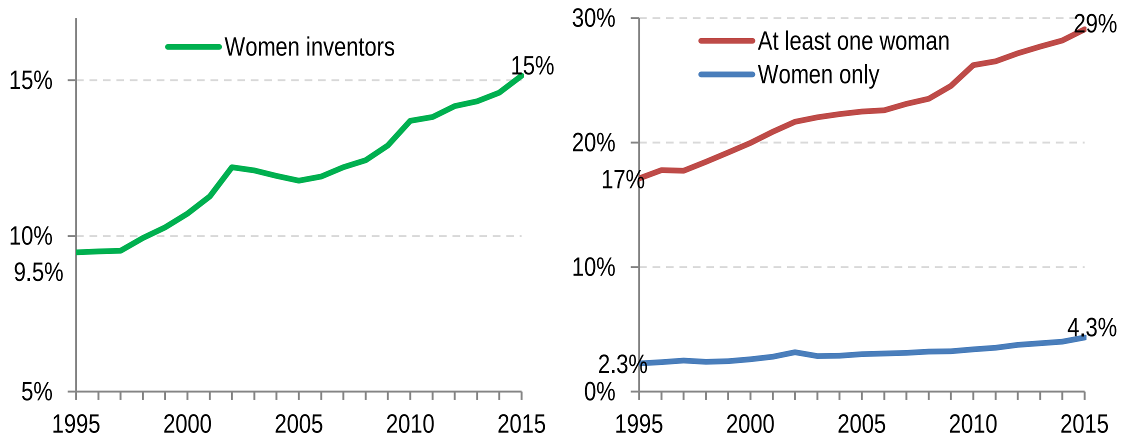
[6] See Miguelez and Fink (2013) for details about the coverage and usefulness of nationality data in PCT applications.
[7] We refer as names only to given names throughout this paper, which implies that we are not considering the family names.
[8] See more details about the PCT data, the WGND and the method in the methodological annex.
[9] The results of these verifications were introduced in the ad hoc list mentioned earlier.

4 Is gender balance improving in the patent system?

All indicators related to gender balance in the PCT system show some degree of progress from 1995 to 2015. Despite the increase, in 2015, less than one-third of all international patent applications include women inventors and only one out seven inventors is a woman. These encouraging trends but still low proportions are consistent with the literature discussed previously (Ding et al., 2006; Frietsch et al., 2009; Jung and Ejermo, 2014).

Figure *1* shows the trend for three different indicators of women participation. The first indicator is the share of women inventors in the total of all listed inventors in PCT international applications. In 2015, this share stood at 15.1%, up from 9.5% in 1995. In terms of absolute volume, the total number of women inventors in all international patent applications increased from 7,780 in 1995 to 81,316 in 2015 – representing 12.5% average annual growth. The number of male inventors in all international applications grew by 9.5% annually – from 74,394 in 1995 to 455,624 in 2015. At these rates, it will take no less than 64 years to reach a balanced gender distribution of inventors.

Figure 1: Upward trend in women participation in international patenting



(a) Share of women in total listed inventors

(b) Share of PCT applications with at least one woman inventor and with women inventors only

The right panel of **Figure *1*** shows the two other indicators: PCT applications with women only inventors and those with at least one woman as inventor. Very few PCT applications have women only listed as inventors (4.3%), but the proportion has almost doubled from 1995. Similarly, PCT applications with at least one woman – i.e. PCT applications with women only inventors and those with both men and women – have increased from 17% to 29% in the same period.

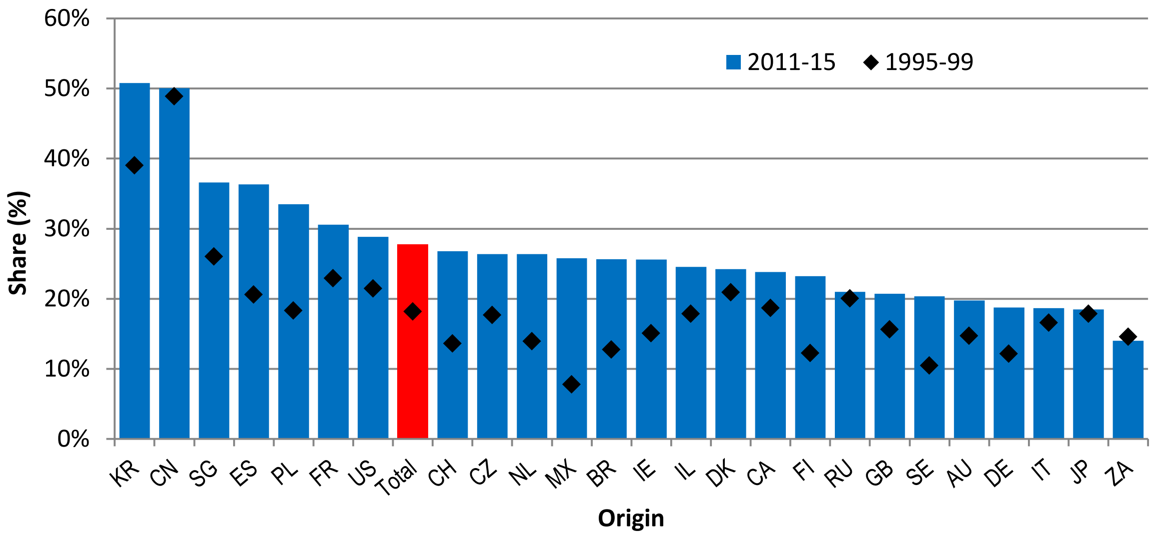
These indicators also suggest women are less likely to file without the opposite gender than men. In 2015, less than 15% of PCT applications with women participation were filed amongst women only, while roughly three-quarters of those with men participation were filed amongst men only. In addition, women are more likely to be in larger groups of inventors – in 2015, the average group was 4.8 for women and 4.2 for men; and, women were less likely to be the only inventor in a patent (7% in 2015) than men (11%). These findings are in line with the existing gender literature (Naldi et al, 2015; Busolt et al., 2008; Moody and Light, 2006; Moody, 2004).

4.1 How does the gender gap vary across countries?

The global average as shown in figure 1 masks the cross country variation in participation of women in international patenting. But how is the geographic distribution to be measured? Borrowing from the patent literature, we can think of at least three different ways to report gender statistics data geographically. First, we could make use of the country of residence of the applicants. Second, we could employ the country of residence of the inventors. Last, we could also make use of the nationality of inventors (or applicants).

Figure *2* reports the share of PCT applications with at least one woman inventor by country of origin of the first applicant.[10] Typically, China and the Republic of Korea have a higher gender balance in international patenting, as 50% of all international applications included women inventors for the period of 2011-2015. In contrast, Germany, Italy, Japan and South Africa have the greatest gender gap among the listed origins – less than one-fifth of all international patent applications included women inventors. The US – the largest user of the PCT System – has around 29% of all international applications with women inventors for the period of 2011-2015. Middle-income countries, such as Brazil and Mexico have marginally better gender balance in international patenting compared to some of the high-income countries such as Canada, Denmark and Finland.

Figure 2: Share of international patent application with at least one woman inventor by selected origins



Qualitatively similar results are found when considering the country of residence or the nationality of inventors (see **Table *2***). We observe little differences in the women shares when reporting country of residence or nationality of inventors for most of the top PCT filing countries. Similarly to women participation, women inventors resident in or nationals of China (29% of Chinese inventors) and the republic of Korean (27%) observe the highest balance among top filing PCT countries for the period of 2011-2015. Women inventors from middle-income countries – such as Brazil (19%) and Mexico (17-18%) – also score comparatively better with these indicators than women

[10] This is WIPO's statistical reporting standard.

inventors in more industrialized economies, such as Italy (14-16%), the US (12-14%), Canada (13%), Denmark (11-13%), Finland (11-12%), Germany (8-9%) or Japan (8%).

Table 2: Women inventors share by residence country and nationality, selected countries

Country	2011-15		1995-99	
	Nationality	Residence	Nationality	Residence
AT	7.2	7.9	4.3	4.9
AU	11.8	12.0	11.2	11.3
BE	14.2	15.3	10.9	12.2
BR	18.8	19.2	13.9	12.4
CA	13.2	12.5	12.0	10.6
CH	7.9	12.3	4.7	5.4
CN	29.2	29.2	37.5	35.5
DE	8.4	9.1	5.6	5.6
DK	11.1	13.0	12.4	12.6
ES	23.0	23.1	18.8	18.2
FI	10.8	11.7	10.2	9.7
FR	17.2	17.0	14.4	14.0
GB	9.4	11.6	8.4	9.0
IE	10.9	10.8	15.4	9.2
IL	12.8	13.5	11.1	11.1
IN	12.9	12.9	20.8	18.1
IT	16.2	13.9	13.4	13.1
JP	7.9	8.2	7.0	7.1
KR	27.2	27.2	23.8	23.4
MX	16.7	18.0	11.6	11.5
NL	9.4	11.8	7.0	6.4
NO	9.0	9.5	6.3	6.7
NZ	11.5	11.2	9.2	8.7
PL	22.1	23.7	17.2	11.5
RU	12.3	12.1	11.5	11.4
SE	10.5	9.9	7.6	7.4
SG	26.0	23.5	26.7	23.6
TR	11.6	11.8	9.0	9.1
US	12.3	13.9	11.8	11.4
ZA	9.7	10.9	9.9	12.4

For most reported top origin countries, there has been an improvement in gender balance between the periods of 1995-99 and 2011-15 (see **Figure *2*** and **Table *2***). We observe the fastest improvement in gender participation for Mexico followed by Spain, Poland and Switzerland (**Figure *2***). Mexico's share of international patent applications with women inventors increased from 7.8% in 1995-99 to 25.8% in 2011-15. Spain, Poland and Switzerland saw a similar magnitude of increase in the share of international patent applications with women inventors – around 15 percentage points. South Africa saw a small decline in the share of patent applications with women inventors, while the share for Japan and the Russian Federation has remained stagnant over the two periods. Again, we find qualitatively similar results when considering the country of residence or the nationality of inventors (**Table *2***). Two notable exceptions concern China and India, for which we observe a substantial reduction of the gender shares both for nationals and residents. However, these two countries had very few PCT applications in the period 1995-1999, which limits the interpretation of these results.

All in all, our results deepen the findings of the existing literature. We also find substantial differences in women participation across countries and a general trend of improvement for the vast majority of these (Frietsch et al, 2009; UKIPO, 2016b).

4.2 Could technology specialization explain the variation in women participation shares?

Our results suggest that women participation varies across technological fields, which explains in part the disparity across countries. However, some countries – notably China and the Republic of Korea – observe higher participation rates across all technologies.

Figure *3* shows the share of international patent applications with women inventors by field of technology in 2015.[11] There is substantial variation across technological fields. The technological field with the highest women participation is biotechnology (57.6%) and the lowest is mechanical elements (10.9%). Both are substantially far from the average 29% and they are not alone among the 35 technological fields. Other fields follow biotechnology closely, such as pharmaceuticals (55.5%), organic fine chemistry (54.1%) and food chemistry (50.7%). In contrast, technological fields related to civil engineering (14.7%); engines, pumps, turbines (14.5%); machine tools (14.1%); and, transports (13.2%) join the mechanical elements field in observing the lowest female participation rates. These results are in line with the literature suggesting life science industries are more prone to be gender balanced and engineering related fields less so (Jung and Ejermo, 2014; Hunt et al, 2012; Whittington and Smith-Doerr, 2008; Eaton, 1999).

This observed disparity of women participation across technological fields is likely to explain at least in part the variation found across countries. Countries specialized in technologies with higher women participation may observe better scores and vice-versa. As shown in **Figure *2***, Germany, Japan, South Africa and the UK observe a low gender balance score, which could be partly due to the large number of patent filings in technological fields where female participation rates tend to be low. For example, only 13% of all transport related PCT applications have women inventors. Germany has a high share of applications in this sector (around one-tenth of all PCT applications), while China – which had a high female participation rate in all PCT applications – filed only 2% of its total PCT applications in this sector. This finding substantiates the findings of previous gender patenting studies affirming that the gender gap is related to women's underrepresentation in patent intensive fields of study (Hunt el al, 2012).

[11] For details of the IPC technology concordance table see www.wipo.int/ipstats.

Figure 3: Share of PCT international patent application with women inventors by field of technology

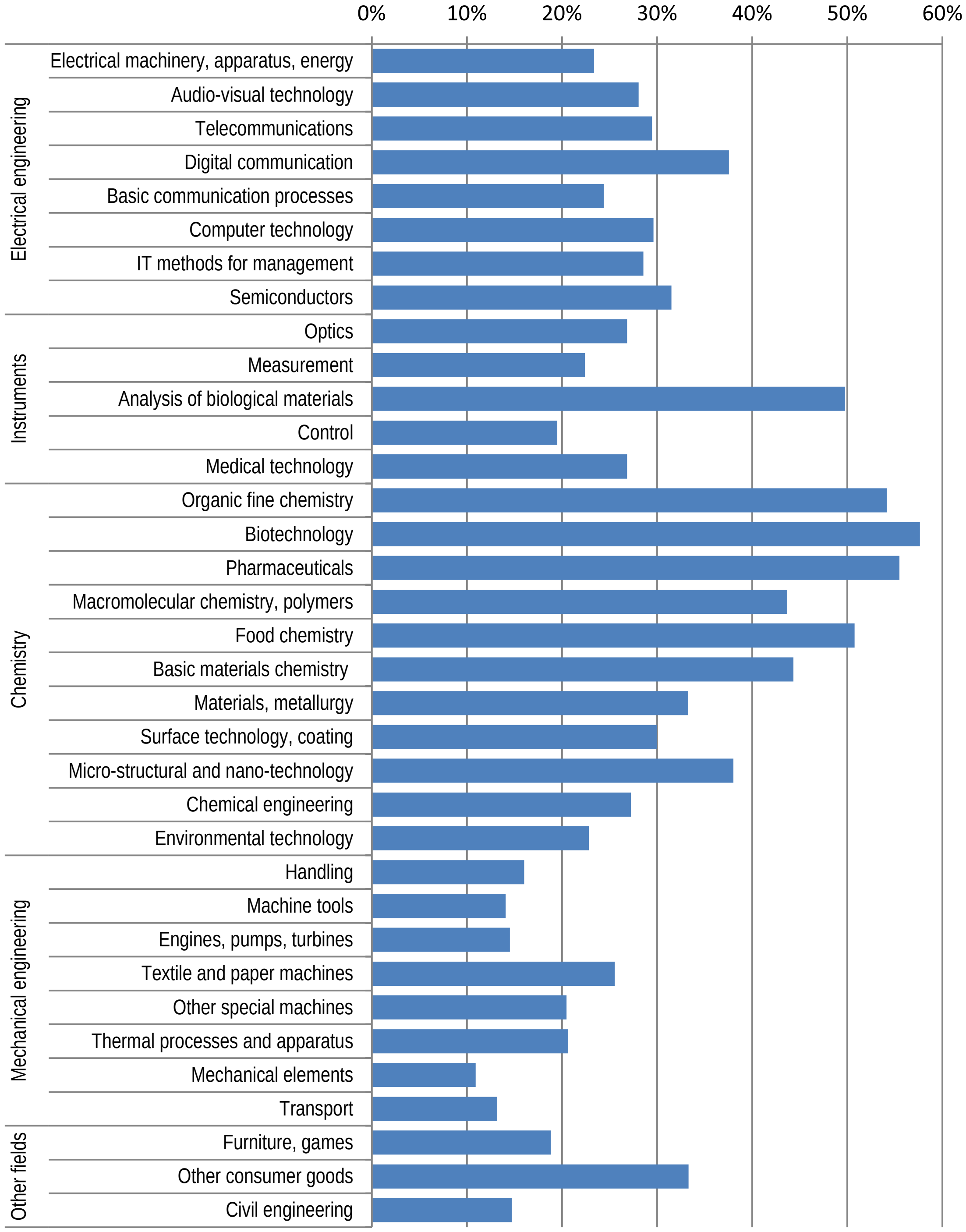
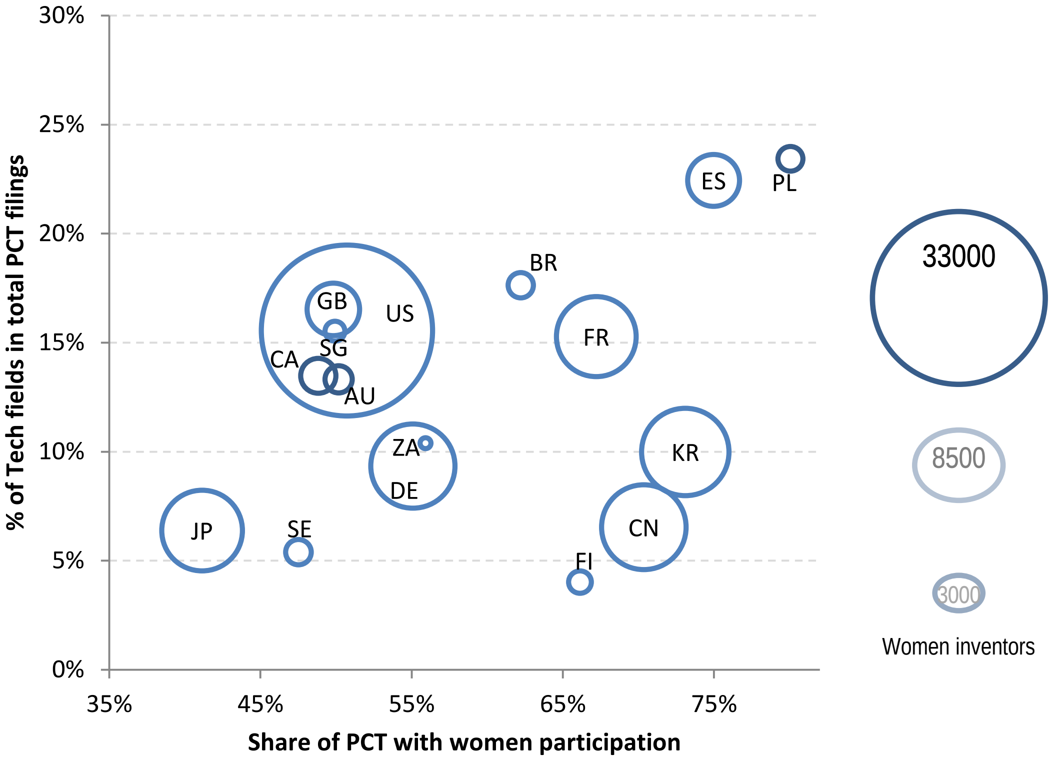
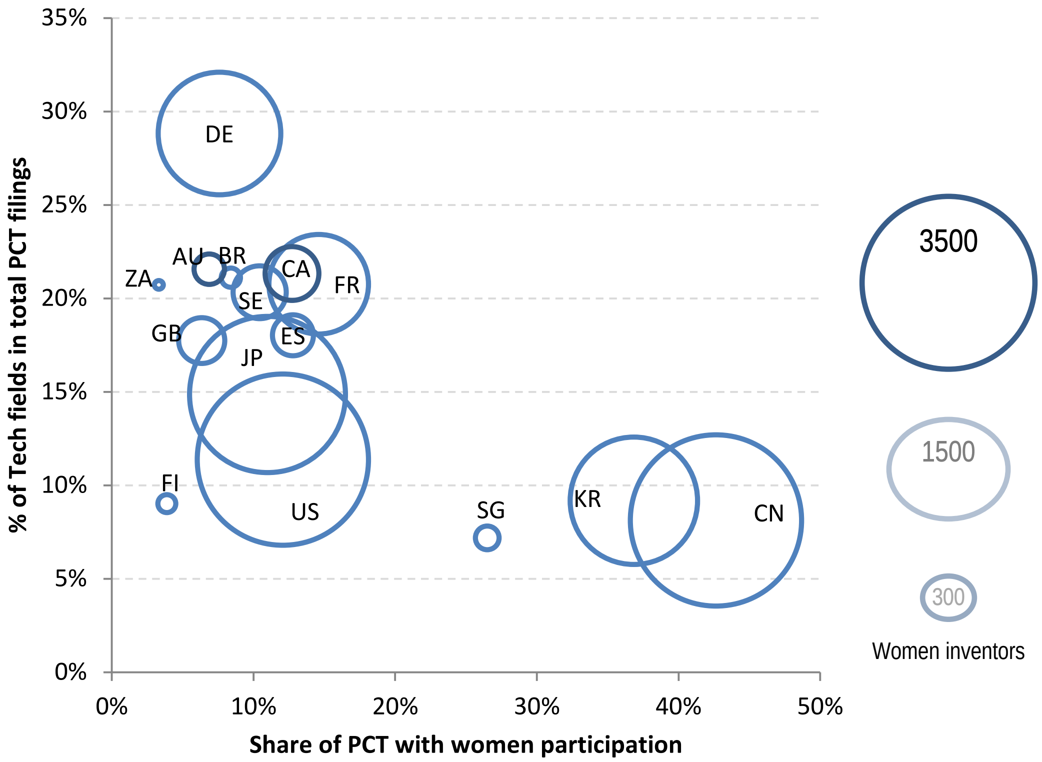


Figure 4: Women participation rates by field of technology and origin



(a) Top 5 gender balanced technological fields



(b) Bottom 5 gender balanced technological fields

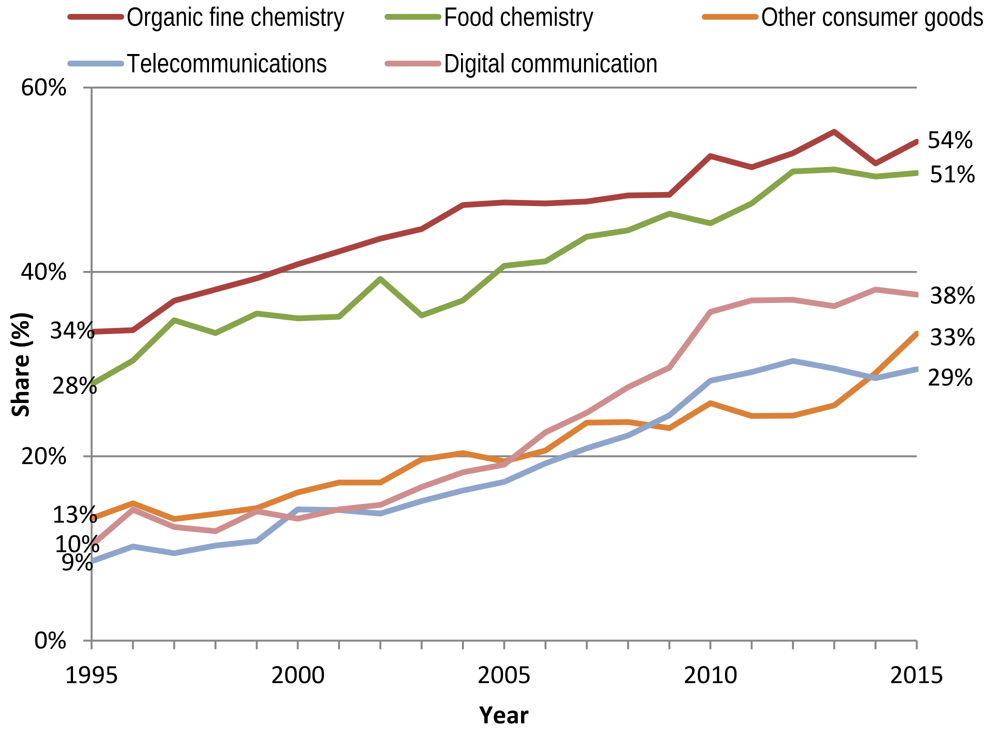
We further illustrate this point by showing women participation rates for the five technological fields with the highest women participation (**Figure *4***a) and those five with the lowest one (**Figure *4***b) for a selection of countries. We compare these participation rates with the share these groups of fields represent in each country's total PCT applications and the amount of women inventors (bubbles' size).

In **Figure *4***a, we observe that Poland and Spain – and to a lesser extent France and Brazil – have high women participation in fields that account for a relatively higher proportion of their total PCT filings. This explains, at least partially, why these countries score a higher than average overall women participation rate. In contrast, we observe that Japan and Sweden have both lower patent activity and women participation rates in fields where women are usually more represented. The Republic of Korea and China are interesting cases with high women participation in fields that represent a relatively lower proportion of their total PCT filings. This indicates that their overall high women participation rate is not necessarily explained by these five top gender-participation fields.

Figure 4b offers the corresponding picture for the five technological fields with the largest gender gap. In this figure, Germany stands out as being more specialized than the other countries in these typically less balanced technological fields. On top of which, Germany has relatively low gender balance for these fields. In contrast, China and the Republic of Korea show the opposite pattern: they are not particularly specialized in these low women participation fields but they still have relatively high women participation.

Figure *5* reports the five technological fields that saw the fastest improvement in gender balance. There has been a sizable increase in the share of PCT applications with women inventors for all the fields of technologies reported in this figure. For example, digital communication and telecommunication narrowed the gender gap between 1995 and 2015 by 27 and 21 percent points, respectively. A large proportion of PCT applications in these two fields originate from China and the Republic of Korea which have a good overall gender balance. In 2015, China accounted for 33.7% and the Republic of Korea for 9.3% of all digital communication filings. In the case of computer technology, China accounted for 17.4% and the Republic of Korea for 7%. Food chemistry and organic fine chemistry and other consumer good categories have also recorded an impressive improvement in narrowing the gender disparity.

Figure 5: Trends in women participation rates by field of technology

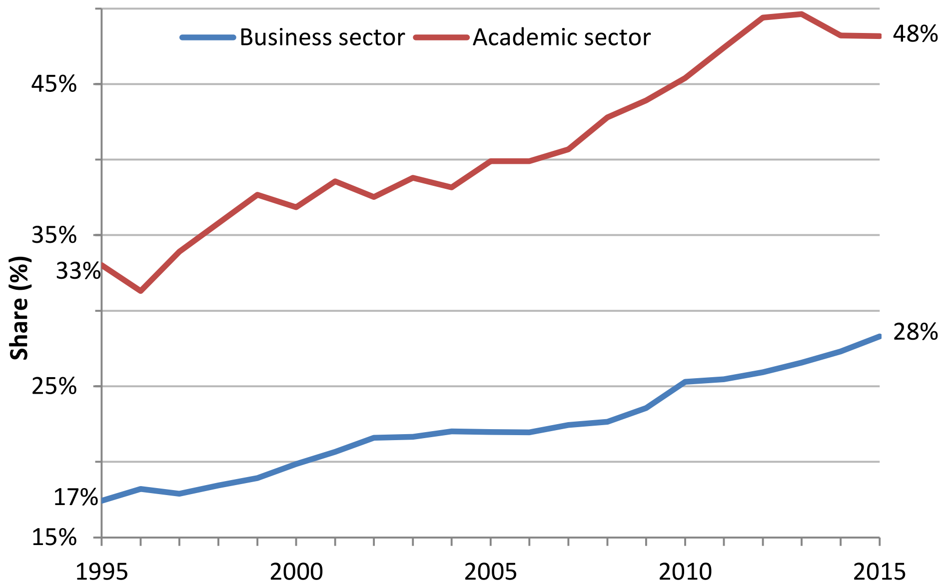


4.3 How does gender disparity vary across and within institutional sectors?

Universities and public research organizations tend to have a high share of PCT applications with women inventors compared to the business sector. In 2015, around 48% of all PCT applications filed by the academic sector included women inventors, compared to only 28% for the business sector. These findings are in agreement with the existing literature (Whittington and Smith-Doerr, 2005).

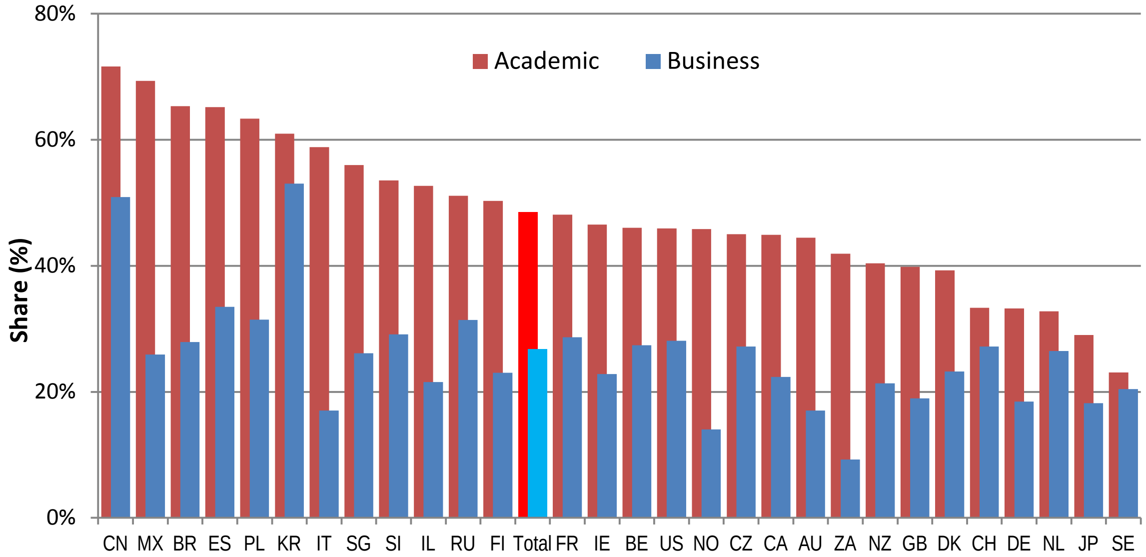
Furthermore, the share women inventors for both sectors have followed an upward trend between 1995 and 2015 (see **Figure *6***). Although the academic sector has the highest women participation rate, the business sector accounts for the largest absolute number of women inventors – by a factor of five. The total number of women inventors recorded in PCT applications between 1995 and 2015 amounted to 702,764 for the business sector and 121,087 for the academic sector. This is expected considering that, in 2015, the business sector accounted for 85% of all PCT applications, compared to 7% by the academic sector.

Figure 6: Share of international patent applications with women inventors by institutional sector



China, Mexico, Brazil and Spain have the highest shares of PCT applications with women inventors for the academic sector – for each of these origins around two-thirds of all PCT applications included women inventors. In contrast, less than 30% of academic PCT applications originating from Japan and Sweden included women inventors (**Figure *7***). The largest gender disparity between the academic and the business sectors is observed for Brazil, Italy, Mexico and South Africa. For example, Mexico has a women participation rate of 69% for the academic sector and 26% for the business sector. In contrast, the Netherlands, the Republic of Korea, Sweden and Switzerland have the least gender disparity between the two sectors.

Figure 7: Shares of international patent applications with women inventors by institutional sector and origin, 2011-15



The majority of the top PCT company applicants saw an increase in the share of PCT applications with women inventors between the periods of 1995-1999 and 2011-2015. **Table *3*** lists companies with higher women participation among the top 100 PCT company applicants. Amongst these, LG Chem Limited of the Republic of Korea has the highest share of women participation for the period 2011-2015. It is followed by L'Oréal of France, Henkel of Germany, Novartis and F. Hoffmann-LA Roche, both of Switzerland, and Merck Patent GMBH of Germany. For each of these companies, around three-fifths of total PCT applications include women inventors. Three of these companies specialize in pharmaceuticals products, while one in chemical and two in beauty products. ZTE Corporation and Huawei Technologies of China are the two top PCT applicants. For both of them, around 50% of PCT applications include women inventors and they ranked in 14th and 15th position respectively in terms of gender balance. However, in absolute numbers, ZTE had the largest number of women inventors (9,298) in PCT applications for the period of 2011-2015, followed by Huawei Technologies (8,531).

Correspondingly, **Table *4*** lists academic institutions with higher women participation among the top 100 PCT applicants of the academic sector. Amongst these, Korea Research Institute of Bioscience and Biotechnology (the Republic of Korea), *Consejo Superior de Investigaciones Cientificas* (CSIC, Spain) and Electronics & Telecommunications Research Institute of Korea (the Republic of Korea) have at least 80% of PCT applications with women inventors. Eight of the top 10 academic applicants with the highest share of PCT applications with women inventors are located either in China or the Republic of Korea. The two exceptions are the CSIC of Spain and the *Institut National de la Sante et de la Recherche Medicale* of France.

Table 3: Top gender balanced PCT applicants for the business sector

Business applicants	Women participation[a] (%)		Women inventors[b]	PCT applications[c]
	2011-15	1995-99	2011-15	2011-15
LG Chem Ltd	71.3	73.2	2,849	2,288
L'Orèal	69.4	63.8	1,737	1,530
Henkel KGaA	65.8	37.8	1,346	1,174
Novartis AG	61.4	35.1	1,168	1,019
F. Hoffmann-La Roche AG	60.7	32.2	1,024	935
Merck Patent GMBH	59.8	43.6	858	935
Samsung Electronics Co Ltd	59.3	38.7	5,007	5,689
BOE Technology Group	56.2	n.a.	1,543	2,045
LG Electronics Inc	56.2	42.9	4,387	5,642
Dow Global Technologies Inc	54.7	n.a.	1,576	1,993
Tencent Technology (Shenzhen) Co Ltd	52.0	n.a.	1,984	2,419
Procter & Gamble Company	51.4	37.0	1,909	2,288
BASE SE	51.2	31.1	3,005	3,646
ZTE Corporation	51.1	n.a.	9,298	13,076
Huawei Technologies Co Ltd	50.5	n.a.	8,531	12,770
NESTEC SA	49.9	n.a.	1,084	1,208
Huawei Device Co Ltd	46.8	n.a.	980	1,615
DSM IP Assets	46.8	n.a.	615	949
Corning Inc	40.7	24.0	807	1,423
Qualcomm Incorporated	40.3	15.6	5,003	9,721
Shenzhen China Star Optoelectronics Tech. Co Ltd	38.6	n.a.	1,274	2,651
Applied Materials Inc	38.0	28.9	888	1,689
E.I. Du Pont de Nemours and Company	37.9	25.3	863	1,693
Microsoft Corporation	36.9	21.8	1,969	3,602
Intel Corporation	36.0	15.2	2,682	5,556
International Business Machines Corporation	35.9	15.6	1,243	2,624
Nitto Denko Corporation	35.2	29.5	812	1,604
3M Innovative Properties Company	34.9	26.7	1,580	3,139
Uni-Charm Corporation	32.9	25.0	365	923
Hitachi High-Technologies Corporation	32.9	n.a.	398	979
Nokia Siemens Networks	31.4	n.a.	416	1,203
Alcatel Lucent	30.6	18.5	941	2,467
General Electric Company	30.1	15.7	885	2,222
Koninklijke Philips Electronics	28.9	9.1	2,403	6,502
Hewlett-Pckard Development Company	28.9	18.9	1,514	4,089
Toray Industries Inc	28.6	19.6	392	1,166
Mitsubishi Chemical Corporation	28.3	45.4	302	884
Compagnie Generale des Etablissements Michelin	28.0	5.5	395	1,039
Thomson Licensing	27.5	24.6	552	1,461
Soc. Nat. Etude et Const. Moteurs Aviation	26.2	25.0	296	916
Kabushiki Kaisha Toshiba	26.0	16.8	859	2,766
Telefonaktiebolaget LM Ericsson	26.0	7.6	2,076	6,703
Asahi Glass Company Ltd	25.7	27.7	440	1,537
Google Inc	25.4	n.a.	935	2,892
Terumo Kabushiki Kaisha	25.0	17.6	314	1,132
Apple Computer Inc	24.6	19.1	649	2,146
Nokia Corporation	24.1	25.0	846	2,885
Daikin Industries Ltd	23.5	9.0	352	1,077
Hitachi Ltd	22.7	21.9	1,102	4,293
Bosch-Siemens Hausgerate GMBH	22.6	7.6	454	1,471

Notes: (a) Share of PCT applications with women inventors (%), (b) Number of women inventors, (c) Number of PCT applications

Table 4: Top gender balanced PCT applicants for the academic sector

Applicants	Women participation[a] (%)		Women inventors[b]	PCT applications[c]
	2011-15	1995-99	2011-15	2011-15
Korea Research Institute of Bioscience and Biotechnology	83.1	100.0	639	261
Consejo Superior de Investigaciones Cientificas	81.2	55.6	734	426
Electronics & Telecommunications Research Institute of Korea	80.5	75.0	606	395
Korea Research Institute of Chemical Technology	77.9	59.2	284	181
China Academy of Telecommunications Technology	75.2	n.a.	1,152	875
Tsinghua University	74.5	50.0	522	329
Korea Institute of Science and Technology	74.5	78.9	194	141
Peking University	74.1	50.0	416	351
Institut National de la Sante et de la Recherche Medicale	70.4	73.1	701	595
Korea Institute of Energy Research	66.5	n.a.	361	245
Yeda Research and Development Co Ltd	64.9	39.7	192	188
Kyunghee University	64.4	n.a.	132	132
Korea Institute of Industrial Technology	63.0	n.a.	302	276
Chonbuk National University	62.9	n.a.	133	132
Korea Advanced Institute of Science and Technology	62.0	33.3	399	408
Korea Research Institute of Standards and Science	61.9	n.a.	110	134
Korea Electronics Technonlogy Institute	61.8	n.a.	175	199
Nanyang Technological University	60.4	n.a.	227	298
Kyungpook National University	60.2	n.a.	147	166
Hanyang University	60.2	n.a.	218	246
Seoul National University	59.5	n.a.	467	462
Ajou University	59.4	n.a.	123	133
USA as repr. by The Secr. Dept. of Health and Human Services	59.4	40.4	445	453
Centre National de la Recherche Scientifique	58.7	46.0	825	846
Agency of Science Technology and Research	58.6	n.a.	618	681
Yonsei University	57.9	n.a.	259	278
Gwangju Institute of Science and Technology	57.7	n.a.	114	142
Korea Institute of Machinery & Materials	53.9	n.a.	150	167
Leland Stanford Junior University	53.6	35.1	414	491
Duke University	53.5	36.0	159	228
New York University	53.2	37.1	203	267
University of Rochester	53.1	35.0	108	147
Hebrew University of Jerusalem	53.0	45.5	152	198
State University of New Jersey	53.0	42.0	104	151
Yale University	52.7	35.1	128	182
Postech Foundation	52.7	n.a.	173	245
Tel Aviv University	52.5	32.4	177	179
Korea University	52.1	n.a.	228	292
Massachusetts Institute of Technology	51.5	28.4	720	1,010
Johns Hopkins University	51.2	42.1	504	664
Council of Scientific and Industrial Research	50.6	n.a.	370	443
University of Pennsylvania	50.3	33.5	246	346
University of California	50.1	33.6	1,305	1,800
Wisconsin Alumni Research Foundation	50.0	38.9	156	192
Sloan-Kettering Institute for Cancer Research	50.0	50.6	125	166
Purdue University	49.0	45.0	120	196
Northeastern University	48.8	40.9	112	172
Columbia University	48.6	38.1	338	521
Max-Planck-Gesellschaft Zur Forderung Der Wissenschaften	48.3	34.3	159	232
University of Colorado	48.1	39.1	138	208

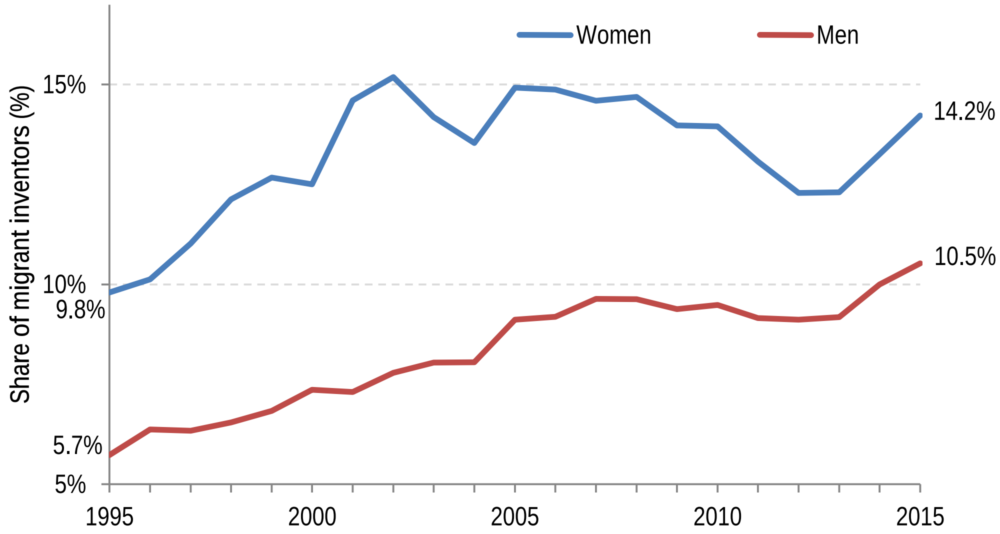
Notes: (a) Share of PCT applications with women inventors (%), (b) Number of women inventors, (c) Number of PCT applications

4.4 Other gender indicators based on patent data

There are many other gender related indicators that could be explored using gender attributed patent data. For instance, **Figure *8*** illustrates the trend of migrant inventors by gender. We observe that women inventors are proportionally more internationally mobile than men, although men have been closing the gap in recent years. In any case, this exploratory statistic indicates that the work done by Miguelez and Fink (2013) could be extended by adding the gender breakdown.

Gender attributed PCT data can also be used to contribute to studies analyzing the gender productivity gap (Cole and Cole, 1973; Cole and Zuckerman, 1984; Fox, 2005; Levin and Stephan, 1998; Zuckerman, 1987, 2001). This will require name disambiguation of inventors and the use of patent citations, as performed in previous studies (see Raffo and Lhuillery, 2009; and, Hall et al, 2001).

Figure 8: Migrant inventors by gender, 1995-2015



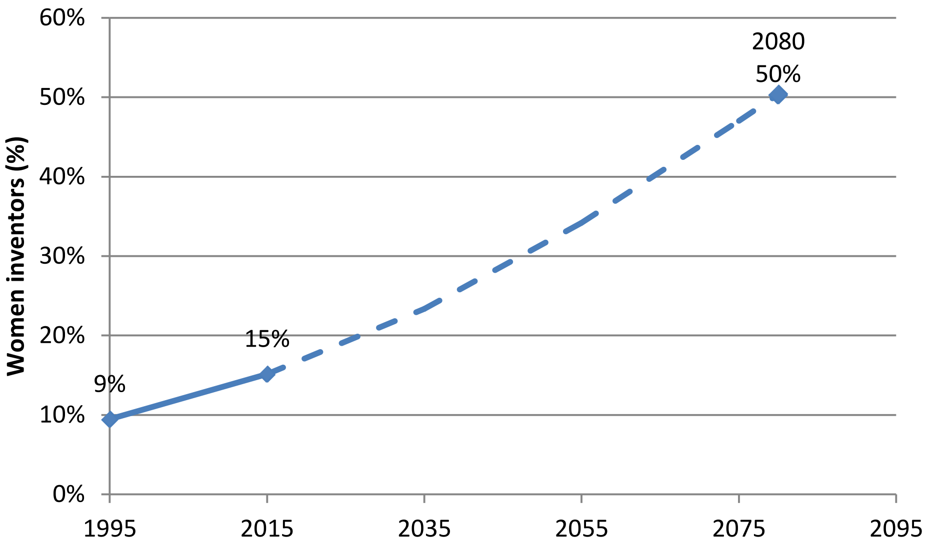
As mentioned above, our aggregate analysis has indicated that women are less likely to file only among the same gender than men. But gender disambiguated PCT data can do much more fine grained analyses. For example, our data also suggest that women inventors are more likely to participate in patents with larger numbers of inventors than men; and, that women participation is even lower for PCT filings by individual applicants than those filed by companies or academia in virtually all top PCT filing countries. All of these basic gender statistics can be extended through network or geographic analysis based on the rich underlying patent data.

5 Conclusions

The main encouraging message of our study is that gender participation in the IP system is getting better. Virtually all indicators related to gender balance in the PCT system show some degree of progress from 1995 to 2015. Overall, women have increased their participation from 17 to 29%. This kind of progress is observed in most countries, in all technical fields and in both academic institutions and companies, although at different rates.

With that said, the proportion of inventing women relative to men remains far from balanced. Assuming that the current progression rates were maintained, we would observe gender balance not before 2080 (see **Figure *9***). This estimate, however, oversimplifies several patterns in gender participation that we observe currently in the PCT filing activity.

Figure 9: When will we achieve gender balance?



First, women participation in patenting is not equally distributed across countries. Some of the most active countries – like China and the Republic of Korea – have contributed significantly to the growth of PCT applications in the recent years; but other gender balanced countries – like Singapore, Spain and Poland – are unlikely to have a large impact on the overall number of PCT applications in the near future. In this sense, women patenting in the US, Germany and Japan are expected to determine to a great extent the progression of gender balance in the following decades.

Second, some technological fields have seen more progress than others, whereby life science related fields – like biotechnology and pharmaceuticals – are among those with higher gender balance scores. In addition, technological fields related to ICTs – e.g. digital communications and telecommunications – have observed woman participation to grow faster than average. The weight these fields will have in overall patenting activity will influence the gender balance in future years.

Third, gender balance progression is higher in academic institutions – i.e. universities and public research organizations – than in the business sector. However, the latter contribute significantly more patent applications to the PCT system. Therefore, successful policies promoting gender balance in the business sector may have greater impact on the overall gender balance than those for the academic one.

Last, it is worth acknowledging that gender balance in the patent system is likely the result of a long social process that accumulates balances and imbalances from previous institutional settings. How the gender balance will evolve in the different scientific fields, higher education institutions and most innovative industries around the world will shape the future gender balance of the PCT system.

One severe challenge for any kind of gender analysis – including ours – is insufficient data. We believe the WGND is a valuable contribution to the gender studies community, although there is always room for further improvement. We intend to update the dictionary to increase even more its international coverage.

In addition, we still know very little about women's contributions in other areas of intellectual property, such as trademarks, industrial designs, copyright and utility models. To date, only limited data are available for these other forms of IP and even less has been explored with a gender breakdown.

In sum, more empirical work is needed to better understand how both women and men could equally access and use the IP system and profit fully from their creative and innovative assets for economic, social, and cultural development.

References

Busolt, U., Kugele, K., and Tinsel, I. (2008). European Studies on Gender Aspects of Inventions-Statistical Survey and Analysis of Gender Impact on Inventions (ESGI).

Cole, J., and Cole, S. (1973). Social stratification in science (Chicago: University of Chicago Press).

Cole, J., and Zuckerman, H. (1984). The productivity puzzle: Persistence and change in patterns of publication of men and women scientists. In Advances in Motivation and Achievement, P. Maehr, and M.W. Steinkamp, eds. (Greenwich, CT: JAI Press), pp. 217–258.

Dernis, H., and Khan, M. (2004). Triadic Patent Families Methodology. OECD Science, Technology and Industry Working Papers.

Ding, W.W., Stuart, T.E., and Murray, F. (2006). Gender Differences in Patenting in the Academic Life Sciences. Science 313, 665–667.

Eaton, S.C. (1999). Surprising Opportunities: Gender and the Structure of Work in Biotechnology Firms. Ann. N. Y. Acad. Sci. 869, 175–188.

Eurostat (2015). Employment statistics, Eurostat Statistics Explained, August 2015.

Eurostat (2016). Gender statistics, Eurostat Statistics Explained, March 2016.

Fox, M.F. (2005). Gender, Family Characteristics, and Publication Productivity among Scientists. Soc. Stud. Sci. 35, 131–150.

Frietsch, R., Haller, I., Funken-Vrohlings, M., and Grupp, H. (2009). Gender-specific patterns in patenting and publishing. Res. Policy 38, 590–599.

Guellec, D. and van Pottelsberghe de la Potterie, B. (2002) “The Value of Patents and Patenting Strategies: Countries and Technology Areas Patterns” Economics of Innovation and New Technology, 11(2): 133-148.

Hall, B., Jaffe, A., and Trajtenberg, M. (2001). The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools.

Hunt, J., Garant, J.P., Herman, H., and Munroe, D. J. (2012). Why Don’t Women Patent? National Bureau of Economic Research, Working Paper No. 17888.

Huyer, S. (2015). Is the gender gap narrowing in science and engineering?, Chapter 3 in UNESCO Science Report: towards 2030.

Jacobs, J. A. (1996). Gender Inequality and Higher Education. Annual Review of Sociology, 22, 153–185.

Jung, T., and Ejermo, O. (2014). Demographic patterns and trends in patenting: Gender, age, and education of inventors. Technol. Forecast. Soc. Change 86, 110–124.

Kugele, K. (2010). Analysis of women’s participation in high-technology patenting, Innovating Women: Contributions to technological advancement. Emerald Group Publ. Ltd. 123–151.

Levin, S.G., and Stephan, P.E. (1998). Gender Differences in the Rewards to Publishing in Academe: Science in the 1970s. Sex Roles 38, 1049–1064.

Lissoni, F., Montobbio, F., and Zirulia, L. (2013). Inventorship and authorship as attribution rights: An enquiry into the economics of scientific credit. Journal of Economic Behavior & Organization, 95, 49–69.

Long, J.S. (1992). Measures of Sex Differences in Scientific Productivity. Soc. Forces 71, 159–178.

Long, J.S. (2001). From Scarcity to Visibility: Gender Differences in the Careers of Doctoral Scientists and Engineers (Washington, DC: National Academies Press).

Mauleón, E., and Bordons, M. (2006). Productivity, impact and publication habits by gender in the area of Materials Science. Scientometrics 66, 199–218.

McMillan, G. (2009). Gender differences in patenting activity: An examination of the US biotechnology industry. Scientometrics 80, 683–691.

Michael, J. (2007). 40000 Namen, Anredebestimmung anhand des Vornamens, c't, issue 17/2007, pp. 182-183

Mickelson, R. A. (1989). Why does Jane read and write so well? The anomaly of women's achievement. Sociol. Educ. 62:47-63.

Miguelez, E., and Fink, C. (2013). Measuring the International Mobility of Inventors: A New Database, WIPO Economic Research Working Paper No.8, May, 2013.

Moody, J. (2004). The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999. Am. Sociol. Rev. 69, 213–238.

Moody, J., and Light, R. (2006). A view from above: The evolving sociological landscape. Am. Sociol. 37, 67–86.

Naldi, F., and Parenti, I.V. (2002a). Scientific and Technological Performance by Gender A feasibility study on Patent and Bibliometric Indicators Vol. I : Statistical Analysis.

Naldi, F., and Parenti, I.V. (2002b). Scientific and Technological Performance by Gender A feasibility study on Patent and Bibliometric Indicators Vol. II : Methodological Report.

Naldi, F., Luzi, D., Valente, A., and Parenti, I. V. (2005). Scientific and Technological Performance by Gender. In Handbook of Quantitative Science and Technology Research, H.F. Moed, W. Glänzel, and U. Schmoch, eds. (Springer Netherlands), pp. 299–314.

Park, S.-B., and Yoon, H.-G. (2007). Determining the Gender of Korean Names for Pronoun Generation. International Journal of Computer Science and Engineering, 1(4).

Raffo, J., and Lhuillery, S. (2009). How to Play the 'Names Game': Patent Retrieval Comparing Different Heuristics. Research Policy 38, no. 10: 1617–27.

Sugimoto, C.R., Ni, C., West, J.D., and Larivière, V. (2015). The Academic Advantage: Gender Disparities in Patenting. PLOS ONE 10, e0128000.

Tang, C., Ross, K., Saxena, N., and Chen, R. (2011). What's in a Name: A Study of Names, Gender Inference, and Gender Behavior in Facebook. In J. Xu, G. Yu, S. Zhou, and R. Unland (Eds.), Database Systems for Adanced Applications (pp. 344–356). Springer Berlin Heidelberg.

Tripathi, A., and Faruqui, M. (2011). Gender prediction of Indian names (pp. 137–141). Students' Technology Symposium (TechSym), IEEE.

UKIPO (2016a) Gender Profiles in UK Patenting: An analysis of female inventorship, UK Intellectual Property Office Informatics Team, March, 2016

UKIPO (2016b) Gender Profiles in Worldwide Patenting: An analysis of female inventorship, UK Intellectual Property Office Informatics Team, September, 2016

Walsh, J., and Nagaoka, S. (2009). Who invents?: Evidence from the Japan–US inventor survey. RIETI Discuss. Pap.

Whittington, K.B., and Smith-Doerr, L. (2005). Gender and Commercial Science: Women's Patenting in the Life Sciences. J. Technol. Transf. 30, 355–370.

Whittington, K.B., and Smith-Doerr, L. (2008). Women Inventors in Context: Disparities in Patenting across Academia and Industry. Gend. Soc.

WIPO (2016). PCT Yearly Review: The International Patent System, 2016 Edition. WIPO Economics & Statistics Series, 2016:901.

Xie, Y., and Shauman, K.A. (1998). Sex Differences in Research Productivity: New Evidence about an Old Puzzle. Am. Sociol. Rev. 63, 847–870.

Yu, J., Zhao, H., Zheng, B., and Yu, Z. (2014). A method of gender discrimination based on character feature of Chinese names. J. Shandong Univ.

Zuckerman, H. (1987). Persistence and change in the careers of men and women scientists and engineers: a review of current research. In Women: Their Underrepresentation and Career Differentials in Science and Engineering: Proceedings of a Conference, (Washington, D.C.: National Academies Press), pp. 127–162.

Zuckerman, H. (2001). The careers of men and women scientists: Gender differences in career attainments. In Women, Science and Technology: A Reader in Feminist Science Studies, M. Wyer, ed. (Routledge), pp. 69–78.

Methodological annex

Building a World Gender-Name Dictionary

This section describes how we consolidated our world gender-name dictionary (WGND). We compiled the information from 13 different sources, which combined, cover 173 different countries to build our WGND. Most of the sources we used come from national public institutions or previous gender studies (see discussion in section 3). Whenever available, we have used the country reported by the source; and, if not available, we have used the origin of the source. We applied name gender semantics – based on honorific titles in English, French and Spanish – in order to attribute the gender to the list of participants in the Assemblies of the Member States of WIPO information. **Table A - *1*** lists these sources.

Table A - 1 – Sources of information

Source	Observations
Social Security Administration (US)	91,320
Alberta government [a]	87,573
Michael (2007)	72,670
Office for National Statistics of United Kingdom (ONS) [a]	34,214
Tang et al. (2011)	21,512
US Census Bureau (2000)	5,164
Wikipedia [a]	2,358
WIPO (Assemblies list)	980
Statistics Sweden [a]	965
Instituto Nacional de Estadística (Spain) [a]	200
Institut National de la Statistique (France) [a]	183
Yu et al. (2014)	155
Denmark Statistics [b]	46
WIPO (Manual check) [c]	2,445
TOTAL	319,785

Notes: Some observations were dropped due to text cleaning or duplications; (a) Accessed in December, 2015; (b) Accessed in May, 2016; (c) ad-hoc list.

In addition to these public sources, we also made use of an ad-hoc list of names, which was compiled after a preliminary round of results. Chinese, Indian, Japanese, and Korean WIPO staff native speakers created this list after manually checking the results of a first round of gender attribution. This step was instrumental in signaling that Chinese and Korean names are hard to disambiguate by gender in their Romanized form. In this respect, our final WGND includes 184 Chinese and 380 Korean unique names in their original characters. We compiled these following the indications in Yu et al. (2014) and those from WIPO staff native speakers. It is important to note that our final WGND has a less complete coverage for these two languages than most western ones.

These 14 sources – 13 public and one ad-hoc list – totaled 319,785 observations, which pair names and countries. Out of these, 174,418 (54.5%) are attributed to female names and 123,374 (38.6%) to male ones. The remaining 21,993 (6.9%) have been regrouped as unisex or ambiguous cases. However, there is an approximately 10% overlap between sources, and a few listed names are combinations of single letters, which reduces the final list of unique name-country pairs to 290,020. It is worth noting that sources may conflict about the gender of certain names. However, we found only 1,927 cases (0.7%) of all name-country pairs with conflicting gender across sources. We also observe a certain redundancy of names across countries, as only 185,924 (64%) names are unique in our data. Out of these, 9,299 (5%) names have conflicting gender across countries. Nevertheless, such divergence is to be expected due to different customs in different languages – e.g. "Andrea" being a male name in Italian but a female name in Spanish.

We provide four different versions of our WGND dictionary. The first one – labeled *WGND_source* – contains all information from the original sources, including the conflicting cases. Given that there is low conflicting gender attribution across sources, we also provide a version where only the most frequent name-gender pairs across sources is reported, which is named *WGND_country*. Given that there is relatively low conflicting gender attribution across countries, we also provide two additional versions of our WGND final dictionary. In the first one, we expand the name-country pairs based on common language for the 12 most frequent languages: Arabic, Dutch, English, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish, and Chinese.[12] The resulting dataset – labeled *WGND_langcountry* – contains 6,247,039 unique name-country pairs. The second addition – labeled *WGND_nocountry* – is a dataset containing the unique 177,042 names which were non-conflicting across sources and countries.

All four versions of the WGND share the following common traits. The variable labeled *name* includes all available names. In the case of Romanized names, we have capitalized, removed any punctuation marks and replaced any accentuated characters by the non-accentuated version in the list. Instead, we left the non-Romanized names unchanged. All versions of the WGND include composite names, such as "MARIA TERESA" or "JESSE JAMES". The variable labeled *gender* refers to expected gender for each name-country pair (or any country in the case of *WGND_nocountry*). The gender variable is coded as female ("F"), male ("M") and unknown ("?"). The latter groups ambiguous and unisex names. In the case of names in original Chinese or Korean characters, variables *gchar1*, *gchar2* and *gchar12* inform if the gender information refers respectively to a first character, a second one or both (see Yu et al, 2014). The variable labeled *code* refers to the country or region code (omitted in the case of *WGND_nocountry*).[13] In addition, the file *WGND_source* includes 14 variables (headed by "src_") detailing the gender specified by each original source.

We provide all versions of the WGND on our website.[14] The large amount of sources that comprises our dictionary allows users to get reliable results in the attribution of gender for a wide range of countries and nationalities. Moreover, users of the dictionary are able to choose the most relevant sources for the purposes of their studies.

---

[12] Based on the CIA's World Factbook (Accessed in December, 2015).
13 Please note that in some cases codes may not refer to countries but to other geographical units such as provinces or overseas territories. The codes follow the recommended standard on two-letter codes by the Committee on WIPO Standards.
14 See http://www.wipo.int/edocs/pubdocs/en/wipo_pub_econstat_wp_33-tech1.zip

Applying the WGND to PCT patent data

The dataset used for this study contains information on 8,788,617 names of individuals, which refer mostly to inventors, individual applicants, and to a lesser extent includes information about agents and common representatives. In this dataset, there are 394,422 unique names, with an average of 22 repetitions per name. [15] There is substantial variation across names. Roughly half of these unique names appear only once while the most frequent name – "MICHAEL" – is repeated 120,294 times accounting for 1.4% of the total observations.

Most names (98%) can be associated with one of the 218 different countries of residence, but they are concentrated among a small set of countries: the top 20 countries account for 96% of the names, and only six countries – namely the US, Japan, Germany, China, Republic of Korea and France – account for more than three quarters of the observations. Likewise, 59% of the observations can be associated with one of the 196 different nationalities and they evidence a very similar concentration pattern to that of countries of residence.

It is also worth noting that 70% of the unique names are composite names. These are, however, repeated less often than single names – 6.7 times on average – making them only 18% of total observations. Similar to single names, 52% of composite names appear only once, although the most frequent one – "JEAN PIERRE" – is repeated 4,847 times. Geographic concentration is also high, whereby 62% of individuals with composites names in our data are residents of only three countries: the US, the Republic of Korea and the UK. In addition, the PCT data have 249,795 observations with names in original Chinese or Korean characters, out of which 3,846 are unique names.[16]

The PCT dataset offers different ways to approach gender attribution. First, the WGND can be applied to full names or to either the first or second names separately.[17] It also can be applied to the names available in Chinese or Korean characters (Yu et al., 2014). Second, the country of reference for the WGND – either the original or the language expansion – can be either the country of residence or the nationality.

Table A - *2* displays 18 possible permutations of how the WGND could be applied. Analyzing these different combinations offers some useful insights. First, we obtain more gender attribution when employing only the first name than full name, although other results are close. Second, the language expanded version – *WGND_langcountry* – generates more gender attributions than the original dictionary. But as the countries already present in the original version – *WGND_country* – dominate the underlying PCT data, the improvements are rather smaller. Third, we observe systematic improvements in the attribution scores when using nationality instead of country of residence. This is probably related to the migratory background of some of the individuals in our dataset. Still, the fact that some observations lack information on nationality suggests that a combination of country of origin and nationality is a better approach. And lastly, the use of middle names and the original Chinese and Korean spellings score lower than the other methods, although they do add information where

[15] We also take into consideration the permutation of given names and family names in some Asian countries like Japan.
[16] The cases of foreign names transliterated to either Chinese or Korean characters were excluded from the data.
[17] Only 6,045 (1.5%) unique names had more than two names, most of which appear only once.

the other methods may fail. In general, we find that the proportion of observations attributed as women is relatively consistent across the different approaches. The main exceptions are those names based on Korean and Chinese original names, which by definition are not representative of the entire sample.

Table A - 2 – Gender attribution scores

Method	Gender attributed				Female
	Yes	%	No	%	%
Full name + Resident	5,886,591	67.0%	2,902,026	33.0%	11.2%
Full name + Resident + Language	6,158,284	70.1%	2,630,333	29.9%	10.5%
Full name + Nationality	3,879,313	74.5%	1,329,066	25.5%	10.2%
Full name + Nationality + Language	4,042,678	77.6%	1,165,701	22.4%	8.9%
Full name without country	5,051,300	57.5%	3,737,317	42.5%	11.0%
1st name + Resident	6,796,899	77.3%	1,991,718	22.7%	11.6%
1st name + Resident + Language	7,096,912	80.8%	1,691,705	19.2%	10.8%
1st name + Nationality	4,506,650	86.5%	701,729	13.5%	10.5%
1st name + Nationality + Language	4,695,724	90.2%	512,655	9.8%	9.1%
1st name without country	5,559,678	63.3%	3,228,939	36.7%	10.8%
2nd name + Resident	991,833	63.5%	570,735	36.5%	14.9%
2nd name + Resident + Language	1,088,806	69.7%	473,762	30.3%	14.0%
2nd name + Nationality	660,830	68.6%	302,349	31.4%	13.1%
2nd name + Nationality + Language	729,418	75.7%	233,761	24.3%	12.1%
2nd name without country	735,949	47.1%	826,619	52.9%	10.3%
Full Chinese or Korean name	34,807	13.9%	214,988	86.1%	23.7%
1st Chinese or Korean name character	63,442	25.4%	186,353	74.6%	33.0%
2nd Chinese or Korean name character	70,489	28.2%	179,306	71.8%	22.9%
Consolidation without any conflict	7,394,641	84.1%	1,393,976	15.9%	12.0%
Consolidation based on majority	7,844,216	89.3%	944,401	10.7%	12.3%
Consolidation with recursive extension	8,449,937	96.1%	338,680	3.9%	13.2%

Notes: Percentages are based on the availability of data for each combination.

As displayed in the three last rows of **Table A - *2***, the combination of all 18 approaches allows us to improve the completeness of the gender attribution exercise. The best we can do using only one approach is to attribute gender to 7,096,912 observations (80.8%) by using *WGND_langcountry* on the first name only and country of residence. If we consolidate all non-conflicting results, we increase the attribution for 297,729 observations reaching 84.1% of attribution. However, most observations with contradictory gender attribution across approaches show a clear pattern in favor of a given gender. Therefore, if we consolidate results based on the most frequent imputed gender, we can increase attribution for 449,575 additional observations and reach 89.3% attribution.

Table A - 3 – Women names share with and without recursive extension

Country	Without extension		With recursive extension	
	Nationals	Residents	Nationals	Residents
China	30.65%	29.46%	30.50%	29.65%
Korea (Rep. of)	28.01%	28.11%	25.40%	25.43%

In addition, we exploited the gender information obtained for composite names in our PCT data in a recursive way. By attributing gender using full, first and middle names, we produced new information for 43,470 names which were not included in the WGND. For instance, we observe the Korean name "JONG" 11,534 times as first name in our PCT dataset and we attributed 6,538 (57%) of these as male names and only 370 (3%) as female ones. Similarly, the Chinese name "MING" appears 6,287 as first name in our PCT dataset and we attributed 2,763 (44%) of these as female names and only 153 (2%) as male. Therefore, based on such results, we decided to attribute the remaining unattributed observations to their most frequent attributed gender. This recursive extension allows us to recover 605,721 additional observations and reach 96.1% of name attribution. As a result, we observe a slight increase of the proportion of female names from 12.3 to 13.2%. Given that this extension increases particularly the coverage of Chinese and Korean names, we analyzed the results of these countries with and without the recursive extension (**Table A - *3*** summarizes these). Chinese residents and nationals observe virtually the same proportions with and without the extension. In the case of Korean residents and nationals, we observe a diminution of the female names proportion – of less than 3 percent points – when applying the recursive approach.

Table A - 4 – Gender attribution scores by country of residence

Country	Observations %	Gender attributed yes	Gender attributed no	Female %
US	33.5%	97.0%	3.0%	13.6%
Japan	15.9%	94.0%	6.0%	7.4%
Germany	11.0%	99.2%	0.7%	7.9%
China	6.2%	88.3%	11.7%	29.6%
Korea (Rep. of)	4.7%	92.1%	7.9%	25.4%
France	3.9%	99.0%	1.0%	16.3%
UK	3.8%	98.9%	1.1%	11.0%
Netherlands	1.9%	98.4%	1.6%	9.7%
Canada	1.7%	96.5%	3.5%	13.2%
Sweden	1.7%	98.6%	1.4%	9.5%
Italy	1.6%	98.9%	1.1%	14.6%
Switzerland	1.3%	98.8%	1.2%	11.0%
Israel	1.0%	96.6%	3.4%	14.3%
Australia	1.0%	98.3%	1.7%	12.0%
India	1.0%	88.9%	11.1%	15.3%
Finland	0.9%	98.8%	1.2%	10.8%
Spain	0.9%	99.0%	1.0%	21.5%
Belgium	0.7%	99.0%	1.0%	15.2%
Denmark	0.7%	99.1%	0.9%	13.0%
Russia	0.6%	98.3%	1.7%	12.2%
Other	6.1%	95.3%	4.7%	12.1%

Notes: the consolidated with recursive extension method applied.

Table A - *4* reports the attribution scores for the top 20 countries of residence when consolidating approaches and applying the recursive extension. Most countries have relatively high attribution scores. The least complete countries are China (12% unattributed), India (11%), Korea (8%) and Japan (6%). All the other top 20 countries each have less than 5% unattributed observations. As whole, the remaining 198 countries also observe unattributed observations below 5%.

Table A - 5 – Comparison with previous studies

Country	Study	Years [a]	% female	
			Ref	WGND
Austria	Kugele (2010)	2001-2003	5	6.3
Austria	Frietsch et al. (2009)	2003-2005	3.2	5.7
Canada	UKIPO (2016a)	2000-2015	7.2	13.4
France	Naldi and Parenti (2002a)	1998	11.1	13.7
France	Frietsch et al. (2009)	2003-2005	10.2	15.6
France	UKIPO (2016a)	2000-2015	12.8	16.3
Germany	Naldi and Parenti (2002a)	1998	4.6	6.2
Germany	Frietsch et al. (2009)	2003-2005	4.7	7.6
Germany	UKIPO (2016a)	2000-2015	5.0	8.4
Ireland	UKIPO (2016a)	2000-2015	8.6	10.9
Italy	Naldi and Parenti (2002a)	1998	8.8	12.4
Japan	Walsh and Nagaoka (2009)	1995-2001	1.7	6.9
Japan	UKIPO (2016a)	2000-2015	5.8	7.7
Korea, Rep. of	UKIPO (2016a)	2000-2015	18.1	26.1
Lithuania	Kugele (2010)	2001-2003	23	17.4
Spain	Naldi and Parenti (2002a)	1998	15.8	18.2
Spain	Frietsch et al. (2009)	2003-2005	12.3	20.6
Sweden	Naldi and Parenti (2002a)	1998	6.3	8.5
Sweden	Jung and Ejermo (2014)	2007	9.1	10.9
Switzerland	Frietsch et al. (2009)	2003-2005	5.9	11.1
United Kingdom	UKIPO (2016a)	2000-2015	7.0	11.4
United Kingdom	Naldi and Parenti (2002a)	1998	7.6	10.3
United States	Walsh and Nagaoka (2009)	2000-2003	5.2	14.3
United States	Frietsch et al. (2009)	2003-2005	8.3	13.1
United States	UKIPO (2016a)	2000-2015	6.4	13.8
World	UKIPO (2016b)	2001	7.1	11.2
World	UKIPO (2016b)	2015	11.5	15.1

Notes: (a) the attribution of patents to years may vary according to studies. Some use national application dates and other use publication year. WGND column replicates these by using priority and publication year, respectively.

Several of the studies mentioned in the literature review section provide aggregated descriptive statistics which can be used to benchmark our methodological strategy. **Table A - *5*** summarizes the percentage of female inventors by country and time period for some of these relevant papers. We have also included the equivalent figure obtained when applying the WGND to the PCT data. Our results always exceed in proportion those from previous studies, with the only exception being the proportion found by Kugele (2010) for Lithuania in the period 2001-2003. The difference between the previous studies and ours could be accounted by two main factors: (a) these studies do not use PCT data but mostly national patent collections; and, (b) the different time periods may not align due to methodological differences in assigning patent applications to years.

Overall, our results align with the patterns identified in these studies. First, we observe similar country ranks that have been found by those studies which compare more than one country, such as Naldi and Parenti (2002a), Frietsch et al. (2009), Walsh and Nagaoka (2009), or Kugele (2010). We also observe a qualitatively similar pattern to the recent study by the UKIPO (2016a); although, in this study, Irish inventors are relatively more gender balanced than UK, Canadian and US inventors, whereas our results show US inventors are more gender balanced than UK and Canadian inventors, while the Irish ones are less. Second, we also observe a similar pattern when comparing results from the same country observed in different periods by different studies. For instance, we observe the same pattern of decline in Austria from the period 2001-2003 (Kugele, 2010) to the period 2003-2005 (Frietsch et al., 2009). This is also the case for France (Naldi and Parenti, 2002a; Frietsch et al., 2009; UKIPO, 2016a), Germany (Naldi and Parenti, 2002a; Frietsch et al., 2009; UKIPO, 2016a), Japan (Walsh and Nagaoka, 2009; UKIPO, 2016a), Sweden (Naldi and Parenti, 2002a; Jung and Ejermo, 2014), and the World (UKIPO, 2016b).

We do not, however, observe the exact same patterns for the UK (UKIPO, 2016a; Naldi and Parenti, 2002a) and the US (Walsh and Nagaoka, 2009; Frietsch et al., 2009; UKIPO, 2016a). However, these make use of different patent collections which may limit the comparability.