Intellectual Property Office

Gender Profiles in UK Patenting

An analysis of female inventorship

Intellectual Property Office is an operating name of the Patent Office

This report was prepared by the
UK Intellectual Property Office Informatics Team
March 2016

e-mail: *informatics@ipo.gov.uk*

Intellectual Property Office
Concept House
Cardiff Road
Newport
NP10 8QQ
United Kingdom

www.ipo.gov.uk/informatics

1 Introduction

The Government Office for Science[1], supported by organisations such as Innovate UK[2], the Research Councils[3], the Royal Academy of Engineering[4], the IET[5] and campaigns such as WISE[6] and WiSET[7] have, for many years, been inspiring girls and women to study and build careers in the STEM fields – science, technology, engineering and mathematics. Statistical surveys from these bodies[8] highlight the impact of their work within the education sector with the number of females attaining STEM vocational qualifications increasing from 8% in 2011 to 24% in 2013. The same surveys also highlight the gender demographic transition to the workplace with women making up only 13% of the STEM workforce and women accounting for only 5.5% of engineering professionals.

Diversity statistics regarding the number of women studying STEM subjects in the education sector, up to and including degree level, are quite comprehensive because gender data is readily available regarding the number of women studying these subjects ('inputs') and those receiving qualifications ('outputs'). When looking at industry however the statistical research in this field primarily relies on 'inputs', such as the number of women employed in a given industry. Very little data is available on the 'outputs' of work undertaken by women within STEM industries.

For this reason, a recurring question that has been asked of the IPO over the past few years – by Government colleagues, journalists and diversity and equality groups amongst others – relates to patent statistics about female inventors. Patents are well known as a measurable 'output' of STEM industries and being able to determine the level of invention of female inventors is highly desirable. It is a legal requirement of the Patents Act 1977 that each inventor is named on the patent application. Disclosing an inventor's gender, or any other protected (diversity) characteristics, is obviously not a legal requirement. This means that it has not been possible to provide statistical information about the gender of inventors named on patent applications.

However, recent gender inference work by several academic researchers has changed this. It is now possible, with a high degree of confidence, to infer the gender of inventors and therefore provide some statistical analysis about the patenting activity of female inventors.

This report outlines the approach undertaken by the Informatics Team at the IPO and provides a preliminary study looking at the type of patent analysis that can be undertaken using inferred gender data.

---

[1] GO-Science (Government Office for Science) - *https://www.gov.uk/government/organisations/government-office-for-science*

[2] Innovate UK - *https://www.gov.uk/government/organisations/innovate-uk*

[3] Research Councils UK - *http://www.rcuk.ac.uk/*

[4] Royal Academy of Engineering - *http://www.raeng.org.uk/*

[5] IET (The Institution of Engineering and Technology) - *http://www.theiet.org/*

[6] WISE (Women in Science, Technology and Engineering) - *https://www.wisecampaign.org.uk/*

[7] WiSET (Women in Science, Engineering and Technology) - *http://www.wiset.org.uk/*

[8] For example, WISE: UK Statistics 2014 (*https://www.wisecampaign.org.uk/resources/2015/07/wise-statistics-2014*) and IET, Women in STEM: Statistics and facts (*https://communities.theiet.org/files/7976#.VbTQ7fkbJ_8*)

2 Data sources

A small number of recent academic research projects looking at inferring gender from name data have made it possible to infer inventor gender on patent applications. This study focuses on two different methodologies, one originating from Peking University/NYU Polytechnic School of Engineering/Max Planck Institute for Software Systems[9] (hereinafter known as the Tang methodology) and the other from Massachusetts Institute of Technology (MIT)[10] (hereinafter known as the Matias methodology).

A combination of both the Tang and Matias methodologies has been used and inventor gender inferred by joining this research data with inventor names stored in internal IPO patent databases for all GB patent applications filed between 1978 and 2015. Analysis of the proportion of female inventors (the inventor gender ratio) was subsequently undertaken on the joined up dataset. Further details on both methodologies and how they were used can be found in Appendix A.

[9] Tang, C. et al (2011) What's in a Name: A Study of Names, Gender Inference, and Gender Behavior in Facebook. *Database Systems for Advanced Applications*. 6637. p. 344-356

[10] *https://github.com/OpenGenderTracking/globalnamedata* - MIT PhD research (Matias, N.) undertaken in collaboration with Bocoup and funded by the Knight Foundation

3 Analysis of female inventors

3.1 Historical profiling

Figure 1 shows the annual percentage of female inventors on all GB patent applications between 1980 and 2015. There is a clear increase in the proportion of female inventors from less than 4% in the early 1980s to over 8% in recent years, with a 5.6% average over this time period. Although absolute numbers remain relatively low, the last 10 years has seen an increase of over 15% in the proportion of female inventors (7.2% in 2006; 8.3% in 2015).

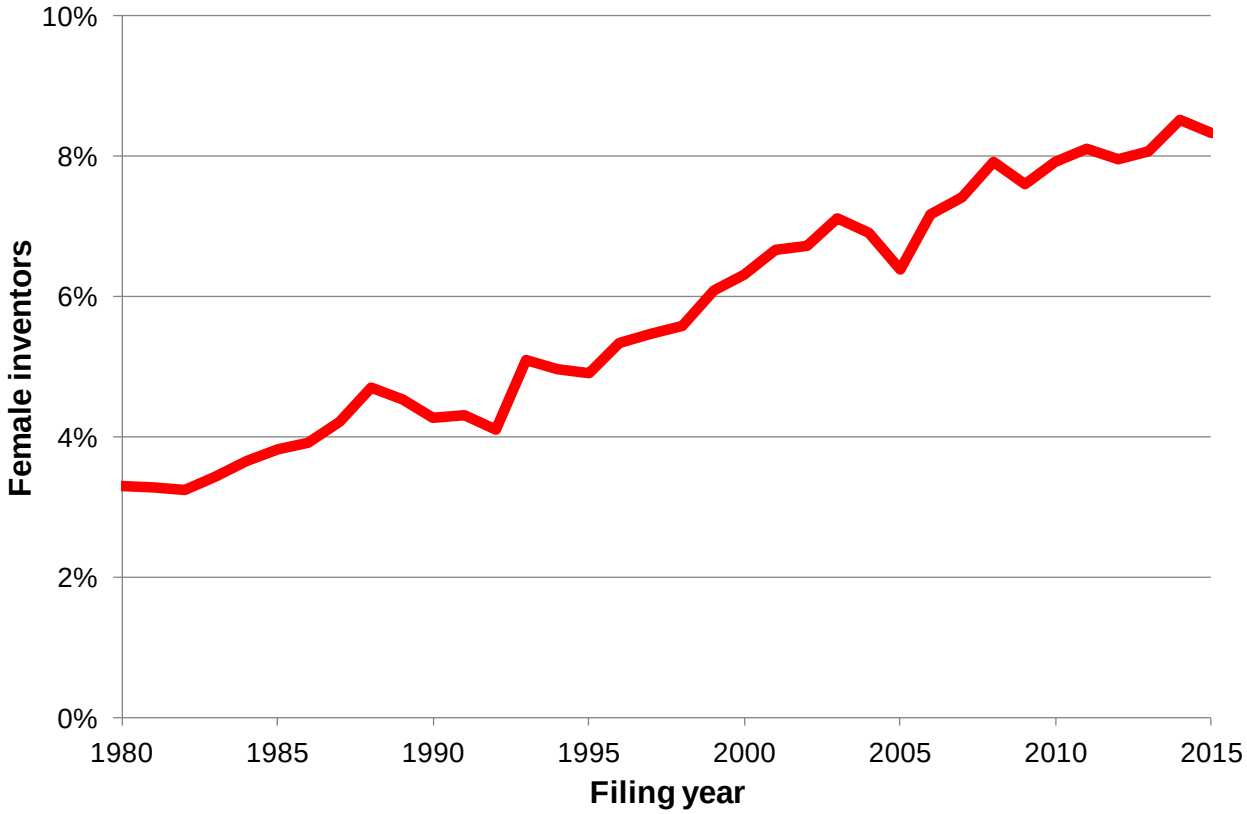

Figure 1: Female inventors on GB patent applications

It is clear from Figure 1 that the proportion of female inventors is increasing at a steady rate. This is not surprising because the gender demographic within industry is very different to what it was in the 1980s. It would therefore be misrepresentative and potentially misleading to perform a more in-depth patent analysis, such as identifying the technology areas with the highest proportion of female inventors, on almost 40 years of patent data; for this reason, most of the subsequent analysis has been performed on a data subset comprising inferred name/gender data on GB patent applications filed between 2000 and 2015, with the proportion of female inventors averaging 7.4% over this time period.

3.2 Geoanalytics

Table 1 shows the inventor gender ratio for the top 20 UK postcode areas with the most inventors for GB patent applications filed between 2000 and 2015. Over this time period the percentage of British female inventors on GB patent applications averages 7.0% and Table 1 shows that only 8 of the top 20 UK postcode areas with the most inventors have an above-average proportion of female inventors.

Table 1: Top 20 highest filing postcode areas on GB patent applications (2000-2015)

Postcode area		Male inventors	Female inventors
CB	Cambridge	92.0%	8.0%
OX	Oxford	91.4%	8.6%
BS	Bristol	95.2%	4.8%
RG	Reading	92.9%	7.1%
DE	Derby	96.5%	3.5%
SO	Southampton	94.5%	5.5%
GU	Guildford	94.1%	5.9%
CV	Coventry	95.2%	4.8%
B	Birmingham	92.8%	7.2%
LE	Leicester	93.0%	7.0%
CM	Chelmsford	95.1%	4.9%
SN	Swindon	95.1%	4.9%
S	Sheffield	94.8%	5.2%
NG	Nottingham	92.0%	8.0%
GL	Gloucester	95.4%	4.6%
AB	Aberdeen	95.4%	4.6%
SW	South West London	87.6%	12.4%
W	West London	88.5%	11.5%
SK	Stockport	94.8%	5.2%
EH	Edinburgh	93.6%	6.4%

Table 2 shows the top 10 and bottom 10 postcode areas for British female inventors on GB patent applications filed between 2000 and 2015, with nationwide data graphically represented in Figure 2.

Table 2: Female inventors on GB patent applications (2000-2015) by postcode area

Postcode area			Male inventors	Female inventors
Top 10	DG	Dumfries	84.7%	15.3%
	IG	Ilford	84.9%	15.1%
	CR	Croydon	85.7%	14.3%
	PA	Paisley	86.4%	13.6%
	BD	Bradford	86.9%	13.1%
	SM	Sutton	87.0%	13.0%
	N	North London	87.3%	12.7%
	SW	South West London	87.6%	12.4%
	SE	South East London	88.0%	12.0%
	BR	Bromley	88.1%	11.9%
Bottom 10	KY	Kirkcaldy	95.2%	4.8%
	LN	Lincoln	95.4%	4.6%
	GL	Gloucester	95.4%	4.6%
	AB	Aberdeen	95.4%	4.6%
	WV	Wolverhampton	95.8%	4.2%
	DY	Dudley	96.0%	4.0%
	DE	Derby	96.5%	3.5%
	BB	Blackburn	97.1%	2.9%
	HR	Hereford	97.2%	2.8%
	TD	Galashiels	98.0%	2.0%

Table 3 shows the inventor gender ratio in each region of the UK and, with seven of the top 10 postcode areas in Table 2 being in the Greater London area, it is not surprising that London is the region with the highest proportion of female inventors. It is interesting, but perhaps not unsurprising, that the proportion of female inventors in London is higher than the other ten regions of the UK, all of which have relatively similar inventor gender ratios.

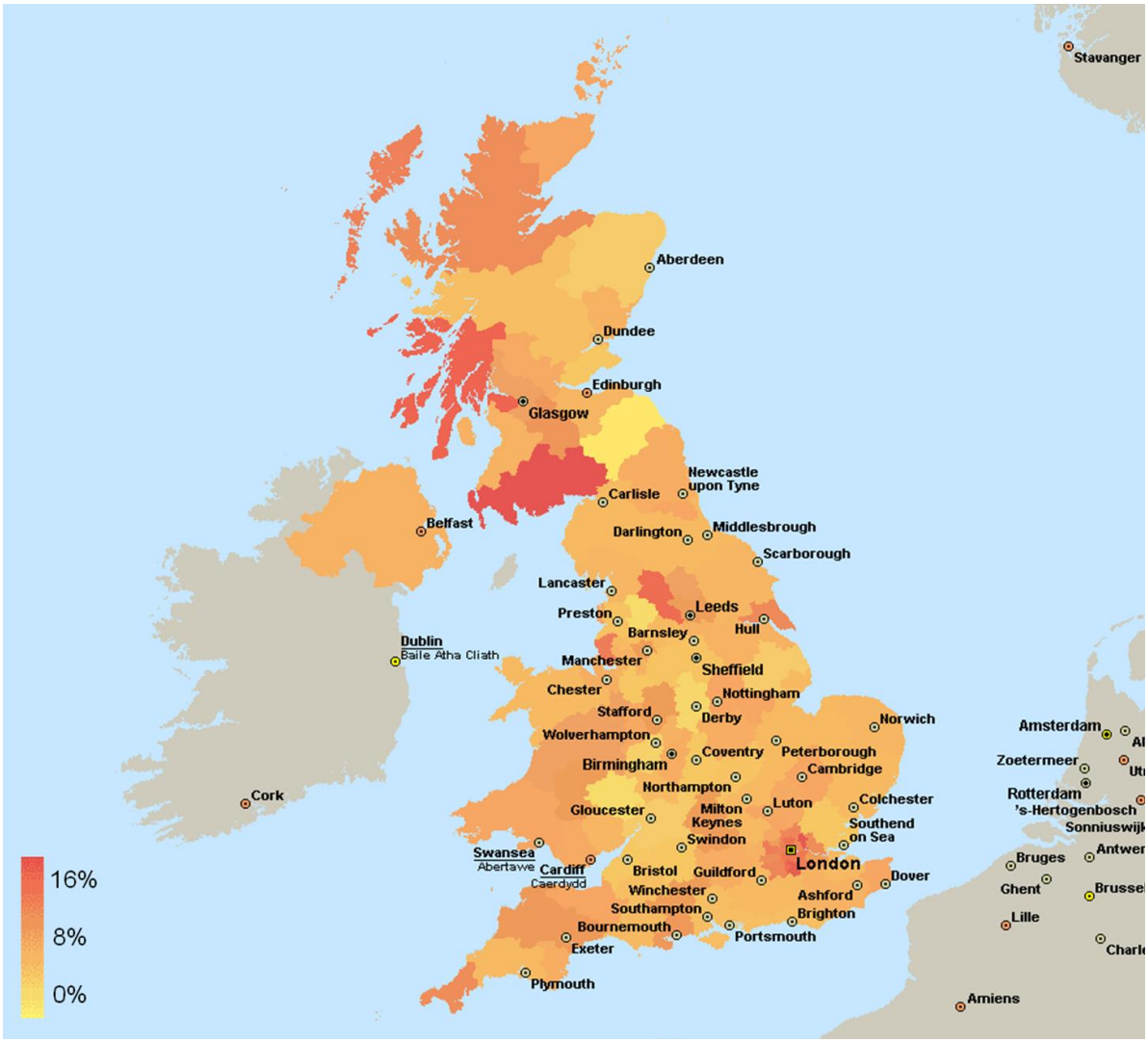

Figure 2: Female inventors on GB patent applications (2000-2015) by postcode area

Table 3: Female inventors on GB patent applications (2000-2015) by region

Region	Male inventors	Female inventors
London	88.5%	11.5%
Yorkshire	92.6%	7.4%
Wales	92.8%	7.2%
South East	93.2%	6.8%
East of England	93.2%	6.8%
Northern Ireland	93.3%	6.7%
Scotland	93.4%	6.6%
North West	93.6%	6.4%
South West	94.0%	6.0%
West Midlands	94.1%	5.9%
East Midlands	94.3%	5.7%

British inventors only account for 40% of all inventors listed on GB patent applications filed between 2000 and 2015. Table 4 shows the top inventor countries on GB patent applications within this time period and the corresponding inventor gender ratio for each country. Data coverage limitations within the inventor country field mean that almost 17% of inventors have an unknown country. The data in Table 4 suggests that the UK has a higher percentage of female inventors than the USA and Germany[11], but is some way behind the proportion of female inventors in France.

Table 4: Top 10 inventor countries on GB patent applications (2000-2015)

Inventor country	Proportion of dataset	Gender known	Inventor gender ratio	
			Male	Female
United Kingdom	39.38%	96.0%	93.0%	7.0%
USA	21.13%	91.1%	93.6%	6.4%
Unknown	*16.63%*	*85.5%*	*92.6%*	*7.4%*
Germany	4.20%	89.6%	95.0%	5.0%
Japan	2.12%	69.6%	94.2%	5.8%
Taiwan	1.86%	14.7%	91.9%	8.1%
France	1.49%	89.3%	87.2%	12.8%
Korea	1.32%	15.9%	81.9%	18.1%
Ireland	1.07%	97.4%	91.4%	8.6%
Canada	1.03%	93.9%	92.8%	7.2%

The combined Tang and Matias methodologies used to infer inventor gender for this analysis give very good data coverage with 86.5% of all inventors on GB patent applications since 1978 having an inferred gender with confidence score of 95% or above[12]. However one of the limitations of the methodologies used for this study is visible in Table 4; the techniques used within the Tang and Matias methodologies clearly struggle with non-Latin alphabet names (and machine-transliterations of non-Latin characters) because Table 4 shows very low gender data coverage for Taiwanese and Korean inventors.

[11] Assuming the same or similar inventor gender ratios apply across other national patents (*e.g.* US patents) and regional patents (*e.g.* EP patents), analysing both of which was beyond the scope of this study

[12] See Appendix A and Appendix B for further details

3.3 Technology focus

All patents are classified according to the International Patent Classification (IPC)[13]. Patents are classified based upon the technical features of the invention and therefore provide an insight into the area of technology for which protection is sought. Table 5 shows the inventor gender ratio for the top 20 IPC subclasses with the most inventors for GB patent applications filed between 2000 and 2015. Over this time period the percentage of female inventors on GB patent applications averages 7.4% and Table 5 shows that 12 of the top 20 IPC subclasses with the most inventors have an above-average proportion of female inventors.

Table 5: Top 20 highest filing IPC subclasses on GB patent applications (2000-2015)

IPC subclass		Male inventors	Female inventors
G06F	Computing; Calculating; Counting -> Electric Digital Data Processing	94.0%	6.0%
E21B	Earth Or Rock Drilling; Mining -> Earth Or Rock Drilling; Obtaining Oil, Gas, Water, Soluble Or Meltable Materials Or A Slurry Of Minerals From Wells	95.9%	4.1%
H04L	Electric Communication Technique -> Transmission Of Digital Information, e.g. Telegraphic Communication	93.6%	6.4%
G01N	Measuring; Testing -> Investigating Or Analysing Materials By Determining Their Chemical Or Physical Properties	88.7%	11.3%
H01L	Basic Electric Elements -> Semiconductor Devices; Electric Solid State Devices Not Otherwise Provided For	91.1%	8.9%
H04N	Electric Communication Technique -> Pictorial Communication, e.g. Television	92.6%	7.4%
A61K	Medical Or Veterinary Science; Hygiene -> Preparations For Medical, Dental, Or Toilet Purposes	74.9%	25.1%
H04W	Electric Communication Technique -> Wireless Communication Networks	90.7%	9.3%
G06Q	Computing; Calculating; Counting -> Data Processing Systems Or Methods, Specially Adapted For Administrative, Commercial, Financial, Managerial, Supervisory Or Forecasting Purposes	91.0%	9.0%
B65D	Conveying; Packing; Storing; Handling Thin Or Filamentary Material -> Containers For Storage Or Transport Of Articles Or Materials, e.g. Bags, Barrels, Bottles, Boxes, Cans, Cartons, Crates, Drums, Jars, Tanks, Hoppers, Forwarding Containers; Accessories, Closures, Or Fittings Therefor; Packages	89.8%	10.2%
A61B	Medical Or Veterinary Science; Hygiene -> Diagnosis; Surgery; Identification	90.6%	9.4%
G01V	Measuring; Testing -> Geophysics; Gravitational Measurements; Detecting Masses Or Objects; Tags	95.0%	5.0%
A61P	Medical Or Veterinary Science; Hygiene -> Therapeutic Activity Of Chemical Compounds Or Medicinal Preparations	74.7%	25.3%
H04B	Electric Communication Technique -> Transmission	94.3%	5.7%
B01D	Physical Or Chemical Processes Or Apparatus In General -> Separation	92.5%	7.5%
G02B	Optics -> Optical Elements, Systems, Or Apparatus	94.4%	5.6%
G06T	Computing; Calculating; Counting -> Image Data Processing Or Generation, In General	93.9%	6.1%
B60R	Vehicles In General -> Vehicles, Vehicle Fittings, Or Vehicle Parts, Not Otherwise Provided For	96.0%	4.0%
A61M	Medical Or Veterinary Science; Hygiene -> Devices For Introducing Media Into, Or Onto, The Body; Devices For Transducing Body Media Or For Taking Media From The Body; Devices For Producing Or Ending Sleep Or Stupor	91.5%	8.5%
A61F	Medical Or Veterinary Science; Hygiene -> Filters Implantable Into Blood Vessels; Prostheses; Devices Providing Patency To, Or Preventing Collapsing Of, Tubular Structures Of The Body, e.g. Stents; Bandages, Dressings; First-Aid Kits	83.9%	16.1%

[13] https://www.gov.uk/government/publications/patent-classification/patent-classification

Table 6 shows the top 10 and bottom 10 IPC subclasses[14] for female inventors on GB patent applications filed between 2000 and 2015. There are clear, and somewhat gender stereotypical, differences between the top and bottom IPC subclasses.

Table 6: Female inventors on GB patent applications (2000-2015) by IPC subclass

	IPC subclass		Male inventors	Female inventors
Top 10	A41C	Wearing Apparel -> Corsets; Brassieres	44.5%	55.5%
	A41B	Wearing Apparel -> Shirts; Underwear; Baby Linen; Handkerchiefs	49.2%	50.8%
	C12R	Biochemistry; Beer; Spirits; Wine; Vinegar; Microbiology; Enzymology; Mutation Or Genetic Engineering -> Indexing Scheme Relating To Micro-Organisms	61.5%	38.5%
	A61Q	Medical Or Veterinary Science; Hygiene -> Use Of Cosmetics Or Similar Toilet Preparations	63.1%	36.9%
	A41F	Wearing Apparel -> Garment Fastenings; Suspenders	66.1%	33.9%
	A47D	Furniture; Domestic Articles Or Appliances; Coffee Mills; Spice Mills; Suction Cleaners In General -> Furniture Specially Adapted For Children	68.8%	31.2%
	A41D	Wearing Apparel -> Outerwear; Protective Garments; Accessories	69.6%	30.4%
	C40B	Combinatorial Technology -> Combinatorial Chemistry; Libraries, E.G. Chemical Libraries, In Silico Libraries	70.2%	29.8%
	A23C	Foods Or Foodstuffs; Their Treatment, Not Covered By Other Classes -> Dairy Products, e.g. Milk, Butter, Cheese; Milk Or Cheese Substitutes; Making Thereof	70.2%	29.8%
	C12N	Biochemistry; Beer; Spirits; Wine; Vinegar; Microbiology; Enzymology; Mutation Or Genetic Engineering -> Micro-Organisms Or Enzymes; Compositions Thereof; Propagating, Preserving, Or Maintaining Micro-Organisms; Mutation Or Genetic Engineering; Culture Media	73.1%	26.9%
Bottom 10	B25D	Hand Tools; Portable Power-Driven Tools; Handles For Hand Implements; Workshop Equipment; Manipulators -> Percussive Tools	98.9%	1.1%
	B21J	Mechanical Metal-Working Without Essentially Removing Material; Punching Metal -> Forging; Hammering; Pressing; Riveting; Forge Furnaces	99.0%	1.0%
	F41G	Weapons -> Weapon Sights; Aiming	99.0%	1.0%
	F41B	Weapons -> Weapons For Projecting Missiles Without Use Of Explosive Or Combustible Propellant Charge; Weapons Not Otherwise Provided For	99.1%	0.9%
	F42B	Ammunition; Blasting -> Explosive Charges, e.g. For Blasting; Fireworks; Ammunition	99.1%	0.9%
	F22B	Steam Generation -> Methods Of Steam Generation; Steam Boilers	99.1%	0.9%
	F02N	Combustion Engines; Hot-Gas Or Combustion-Product Engine Plants -> Starting Of Combustion Engines; Starting Aids For Such Engines, Not Otherwise Provided For	99.2%	0.8%
	F16D	Engineering Elements Or Units; General Measures For Producing And Maintaining Effective Functioning Of Machines Or Installations; Thermal Insulation In General -> Couplings For Transmitting Rotation; Clutches; Brakes	99.4%	0.6%
	F04F	Positive-Displacement Machines For Liquids; Pumps For Liquids Or Elastic Fluids -> Pumping Of Fluid By Direct Contact Of Another Fluid Or By Using Inertia Of Fluid To Be Pumped; Siphons	99.6%	0.4%
	F23G	Combustion Apparatus; Combustion Processes -> Cremation Furnaces; Consuming Waste By Combustion	100.0%	0.0%

The granularity of IPC subclasses means that outliers may be exaggerated, especially as each IPC subclass is not equal in size. A higher level overview of the IPC sections is considered to be too crude a metric because the sections are so broad; for example the inventor gender ratio for female inventors varies from 15.7% in section C (Chemistry; Metallurgy) and 15.4% in section A (Human Necessities) to 3.9% in section E (Fixed Constructions) and 3.6% in section F (Mechanical Engineering; Lighting; Heating etc).

[14] With over 600 IPC subclasses, analysis was limited to subclasses comprising over 100 inventors (98% coverage)

Previous work by WIPO has produced a technology concordance table[15] that links the IPC symbols with 35 fields of technology. This provides a good level of granularity for this study, as shown in Table 7. There is a clear divide between the chemistry technology fields and the mechanical and electrical engineering fields; eight of the 11 chemistry fields fall in the top 12 fields for female inventors, and five of the eight mechanical engineering technology areas fall within the bottom six fields.

Table 7: Female inventors on GB patent applications (2000-2015) by WIPO technology concordance

Technology area	Male inventors	Female inventors
Chemistry: Biotechnology	74.5%	25.5%
Chemistry: Pharmaceuticals	75.6%	24.4%
Chemistry: Organic fine chemistry	76.6%	23.4%
Chemistry: Food chemistry	80.8%	19.2%
Other fields: Other consumer goods	82.3%	17.7%
Instruments: Analysis of biological materials	82.7%	17.3%
Chemistry: Macromolecular chemistry, polymers	84.7%	15.3%
Chemistry: Basic materials chemistry	85.9%	14.1%
Instruments: Medical technology	87.9%	12.1%
Other fields: Furniture, games	89.9%	10.1%
Chemistry: Materials, metallurgy	90.4%	9.6%
Chemistry: Micro-structural and nano-technology	90.9%	9.1%
Mechanical engineering: Textile and paper machines	90.9%	9.1%
Electrical engineering: IT methods for management	91.0%	9.0%
Electrical engineering: Semiconductors	91.1%	8.9%
Mechanical engineering: Other special machines	92.1%	7.9%
Chemistry: Surface technology, coating	92.1%	7.9%
Electrical engineering: Digital communication	92.6%	7.4%
Mechanical engineering: Handling	92.6%	7.4%
Chemistry: Environmental technology	92.7%	7.3%
Chemistry: Chemical engineering	92.8%	7.2%
Instruments: Optics	93.2%	6.8%
Electrical engineering: Telecommunications	93.7%	6.3%
Electrical engineering: Computer technology	93.8%	6.2%
Electrical engineering: Audio-visual technology	94.0%	6.0%
Instruments: Control	94.1%	5.9%
Instruments: Measurement	94.5%	5.5%
Electrical engineering: Electrical machinery, apparatus,	95.5%	4.5%
Electrical engineering: Basic communication processes	95.6%	4.4%
Mechanical engineering: Transport	95.6%	4.4%
Mechanical engineering: Engines, pumps, turbines	95.9%	4.1%
Other fields: Civil engineering	96.1%	3.9%
Mechanical engineering: Machine tools	96.7%	3.3%
Mechanical engineering: Thermal processes and apparatus	96.7%	3.3%
Mechanical engineering: Mechanical elements	97.1%	2.9%

[15] http://www.wipo.int/ipstats/en/statistics/technology_concordance.html

3.4 Teamworking

One area of particular interest when looking at inventor gender is in the analysis of the number of inventors on each patent application. It is a legal requirement to list every inventor involved in the patent seeking protection. Each patent application can range from having one named inventor (a lone/individual inventor) to multiple inventors (working collaboratively as part of a team). By linking the inferred gender of each named inventor and the number of inventors listed on each patent application, analysis could then be undertaken on whether female inventors are more likely to work on their own, as part of an all-female team, or as part of a mixed team. The results in Figure 3 show the percentage of female inventors on GB patent applications between 1980 and 2015 split by inventor type.

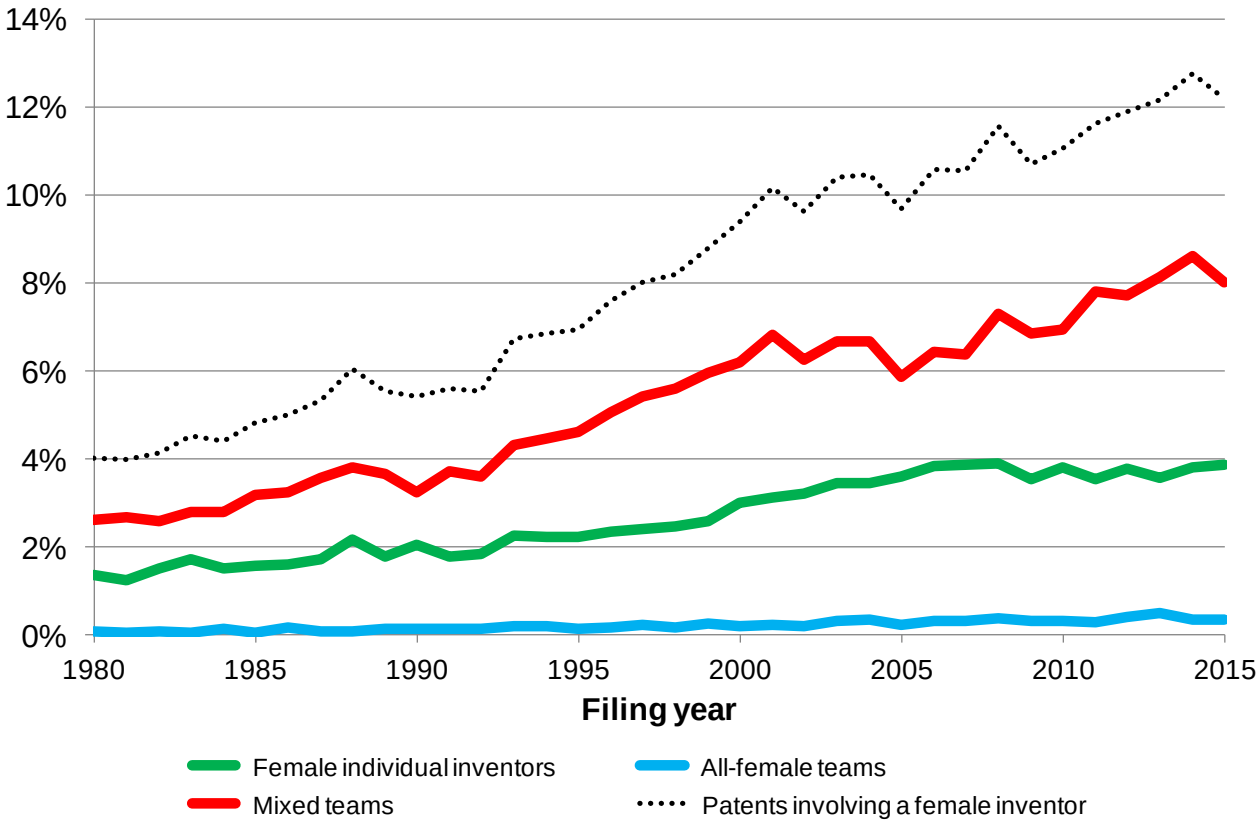

Figure 3: Female inventors on GB patent applications by inventor type

The overall proportion of patents involving a female inventor has more than tripled from 4% in 1980 to over 12% in 2015. A steady rise has also been seen in the number of female inventors working as part of mixed teams. Individual female inventors accounted for less than 1.5% of patents in 1980 and this rose to 3.5% by 2005. Interestingly, however, for the last 10 years the proportion of individual female inventors has remained fairly constant at between 3.5% and 4%. The number of all-female teams has increased over five times since 1980 but the absolute numbers are very low with 0.06% of patents coming from all-female teams in 1980 and only 0.33% in 2015.

The pie charts in Figure 4 show the clear gender disparity and compare the split of inventor types on GB patent applications filed in 1980 and 2015. Although the positive trends shown in Figure 3 appear promising, the increasing proportion of female inventors is slow, and the absolute numbers are still very low. Figure 4 shows the bigger picture and is a more accurate reflection of the gender disparity within UK patenting; 88% of all GB patent applications in 2015 are still from all-male inventors and this rises to almost 96% when mixed teams (with at least one male inventor) are considered.

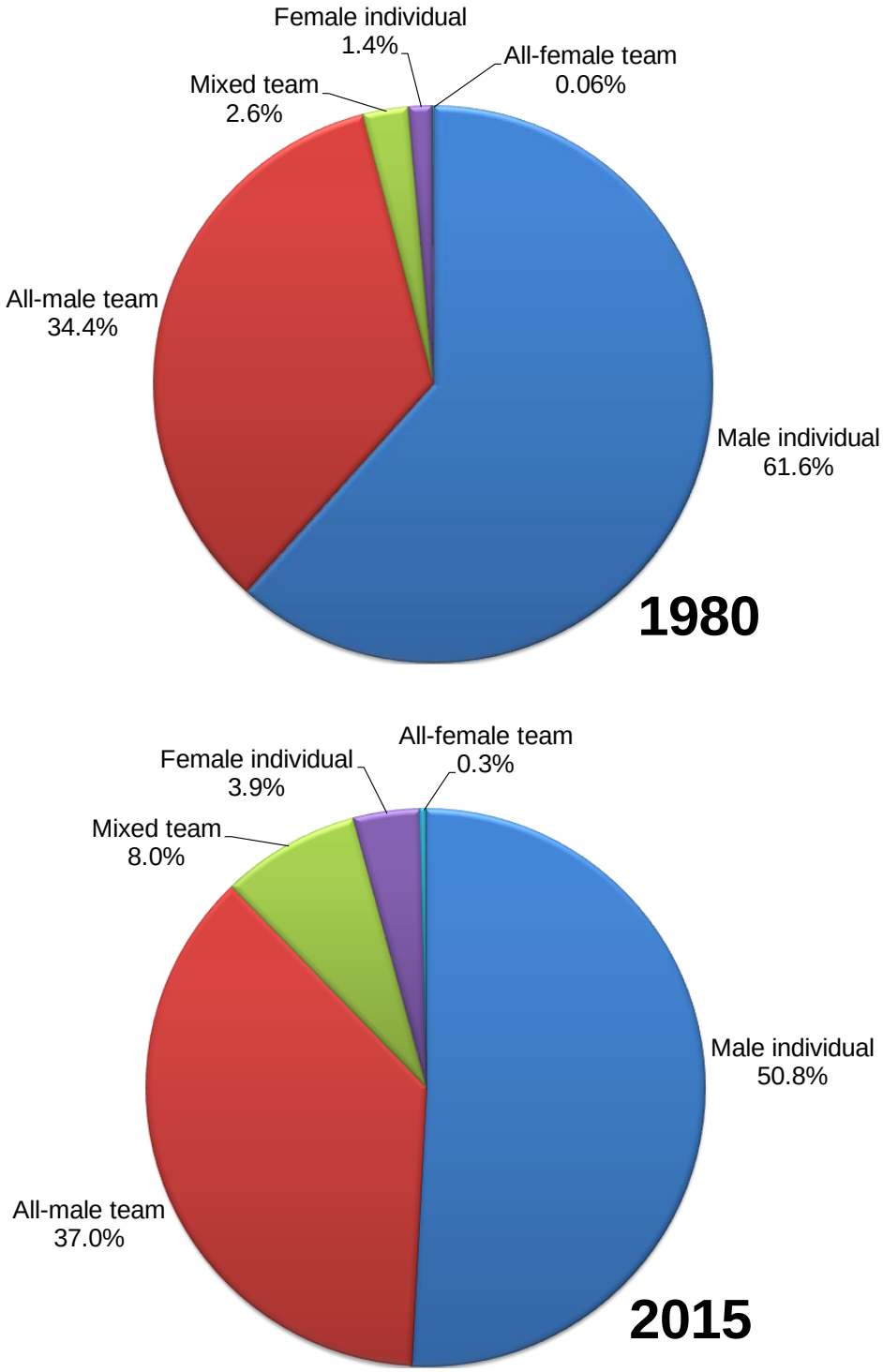

Figure 4: Comparing the inventor types on GB patent applications filed in 1980 and 2015

4 Conclusions

In the 1980s women represented less than 4% of inventors on GB patent applications but this has steadily risen to over 8% in recent years. Although absolute numbers remain relatively low, the last 10 years has seen a 16% increase in the proportion of female inventors.

By analysing postcode areas it was determined that Scotland has both the highest and lowest proportion of female British inventors, with the highest proportion residing in the western Scottish Borders in Dumfries and the lowest in the eastern Scottish Borders in Galashiels, although absolute levels of patenting in these areas are lower than in other parts of the UK. Seven of the top ten postcode areas are within Greater London and some well-known historic industrial towns feature in the bottom ten postcode areas, including Blackburn, Aberdeen and the Black Country (Dudley and Wolverhampton). Regional analysis showed London as the clear leader for female inventor representation with very little separating the other regions of the UK.

British inventors however only account for 40% of all inventors listed on GB patent applications and the inventor gender ratio of female French inventors is 80% higher than it is for female British inventors. The UK does however have a higher proportion of female inventors than the USA, Germany and Japan.

Analysing the granular classification codes applied to each patent application revealed a number of traditional associations with the highest proportions of female inventors listed on patents relating to brassieres, clothing, footwear, cosmetics, furniture and food, and the lowest proportions on patents relating to combustion engines, tools and weapons. A higher level analysis using WIPO's technology concordance table smoothed out the niche technical fields with lower absolute level of patenting and revealed a more accurate reflection of the STEM industries in which the most women are employed, with several chemistry areas including biotechnology and pharmaceuticals having the highest proportion of female inventors.

The overall proportion of patents involving a female inventor (either working alone or as part of a team) has more than tripled from 4% in 1980 to over 12% in 2015 but the last 10 years has seen the proportion of individual female inventors plateau at around 3.75%. The number of all-female teams has increased over five times since 1980 but the absolute numbers are very low with only 0.33% of patents coming from all-female teams in 2015. Although historical analysis reveals ever-increasing levels of female patenting, the growth rate is slow and the absolute numbers are still very low. The world of patenting remains male-dominated and even in 2015 there is a clear gender disparity with 88% of all GB patent applications coming from all-male inventors, rising to almost 96% when mixed teams are considered.

This study has revealed some interesting results whilst also providing some quantitative data to back up the anecdotal evidence within the IP industry about the representation of female inventors. Inventor gender analysis was limited to GB patent applications but further work could be undertaken to investigate wider patent data worldwide. To make this possible, some further work would need to be undertaken to improve the data coverage and gender inference on inventor names with non-Latin alphabets.

Appendix A Data sources

A.1 Tang methodology

The Tang methodology comprises collaborative research by Peking University, NYU Polytechnic School of Engineering and the Max Planck Institute for Software Systems[16]. It involved crawling Facebook® public profile pages for millions of users to generate an annotated name list. The research goes on to use this name list to infer gender information for users who do not explicitly specify their gender and then provides some analysis of gender characteristics and gender behaviour in Facebook®. For the purposes of this study the IPO was only interested in using the annotated name list that was populated using web-crawling to extract the user-disclosed name and gender data from these Facebook® public profiles. This name list comprises the number of sampled Facebook® users having each name (all one-letter names, names without a vowel, and names referenced only once were removed), the number of times it is labelled as male and the number of times it is labelled as female.

A.2 Matias methodology

The Matias methodology originated from research at Massachusetts Institute of Technology (MIT)[17] that was undertaken in collaboration with Bocoup and funded by the Knight Foundation. It involved collecting open source annual birth data from the US Social Security Administration and the UK Office for National Statistics (ONS) into a single database. US data from the US Social Security Administration provides records for name and gender by year for births between 1880 and 2011. UK ONS data records births for England and Wales between 1996 and 2011, with Scotland (2009 and 2010 only) and Northern Ireland (1997-2011) recorded separately. The resulting US and UK name lists each comprise the number of male and female entries and the number of years in which each name appears. For the purposes of this study the IPO combined both the US and UK name lists for further analysis.

[16] Tang, C. et al (2011) What's in a Name: A Study of Names, Gender Inference, and Gender Behavior in Facebook. *Database Systems for Advanced Applications*. 6637. p. 344-356

[17] https://github.com/OpenGenderTracking/globalnamedata - MIT PhD research by Matias, N.

A.3 Comparing methodologies

Both the Tang and Matias methodologies provide open source datasets listing names alongside a count of how many entries are male and female, as shown in Table 8.

Table 8: Example of name list entry format

Name	Entries	Male	Female
Samantha	11906	0	11906
Matty	116	109	7

The IPO used all of the names in these two open source databases to infer a gender with an assigned confidence score based on the number of male/female entries compared to the total number of entries. For example, in Table 8 the gender of Samantha would be inferred to be female with a 100% confidence and Matty would be inferred to be male with a 94% confidence score.

Table 9 provides a summary of the two methodologies and shows that 85% of the names in the Matias list have a gender with a known confidence score of 95% or above, compared to 78% in the Tang name list.

Table 9: Comparison of gender inference between methodologies

Methodology	Confidence Score	Gender Known	Gender Unknown
Tang	≥80%	78.8%	21.2%
	≥90%	78.6%	21.4%
	≥95%	78.3%	21.7%
Matias	≥80%	85.2%	14.8%
	≥90%	85.1%	14.9%
	≥95%	84.9%	84.9%

Names with a gender confidence score of 95% or above were then matched to inventor names held in internal IPO patent databases. Table 10 provides a summary of the gender match for all inventors on GB patent applications (with confidence scores of at least 95%) and shows a difference of almost 7% between the two methodologies. Table 11 shows a comparison of the male/female split of the inventors with an inferred gender from Table 10.

Table 10: Methodology comparison when name/gender lists are matched to IPO inventor names

Methodology	Gender Inferred	Gender Unknown
Tang	78.3%	21.7%
Matias	84.9%	15.1%

Table 11: Inventor gender ratio with gender inferred using each methodology

Methodology	Male	Female
Tang	95.0%	5.0%
Matias	94.6%	5.4%

A.4 Combination dataset

Table 9 and Table 10 clearly show that the Matias methodology provides better results than the Tang methodology. Further exploration of the differences between the name/gender matches in each dataset revealed a small proportion of name/gender matches in the Tang dataset that were not present in the Matias dataset. To increase the name/gender match coverage for this study, these names were appended to the Matias dataset to create a single merged file comprising data from both sources. Table 9 shows that using the Matias methodology 84.9% of inventors on GB patent applications have a gender inferred with a confidence score of 95% or above; this coverage increases to 86.5% in the merged dataset.

The merged name/gender dataset was the chosen source for this study of female inventors. The patent analysis subsequently undertaken focused on the proportion of male and female inventors and with 86.5% coverage it provided a very good indication of the inventor gender ratio on GB patent applications.

A.5 Patent databases and coverage

Internal IPO databases comprise full inventor names (first and last names) for all GB patent applications from 1978 onwards. The inventor gender analysis undertaken in this study was limited to GB patent applications (extracted on 2 February 2016). Wider international coverage could be obtained in the future by using a worldwide patent database such as PATSTAT[18].

[18] https://www.epo.org/searching-for-patents/business/patstat.html

Appendix B Dataset analysed

All patent analysis[19] undertaken in this study used the combination name/gender dataset discussed in Appendix A.4 with a gender confidence score of 95% or above. Table 12 provides a summary of the dataset analysed.

Table 12: Summary of joined up name/gender dataset used for patent analysis

	Male	Female	Gender Unknown	Total	Gender match
All GB patent applications[20]	835,577	47,752	137,814	1,021,143	
	81.8%	4.7%	13.5%		86.5%
	94.6%	5.4%			
All GB patent publications	770,117	41,888	131,749	943,754	
	81.6%	4.4%	14.0%		86.0%
	94.8%	5.2%			
All GB granted patents	532,890	25,583	100,883	659,356	
	80.8%	3.9%	15.3%		84.7%
	95.4%	4.6%			

[19] Further details on how to interpret and analyse patent data can be found in *The Patent Guide* - https://www.gov.uk/government/publications/the-patent-guide

[20] Filed between 1978 and 2015

Concept House
Cardiff Road
Newport
NP10 8QQ

Tel: 0300 300 2000
Fax: 01633 817 777

www.gov.uk/ipo

For copies in alternative formats please contact our Information Centre.

When you no longer need this booklet, please recycle it.

DPS-003379