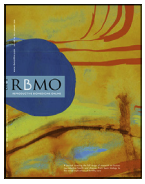
RBMO

ELSEVIER



ARTICLE

Mapping research in assisted reproduction worldwide

BIOGRAPHY

Dr Désirée García obtained her PhD at Barcelona University in 2017 and her BSc in Pharmacy in 2005. She also has a Master's Degree in Research Methodology and Statistics. In 2009, she joined Clínica EUGIN in Barcelona, where her current research focuses on reproductive medicine.

Désirée García[1], Francesco Alessandro Massucci[2], Alessandro Mosca[2], Ismael Ràfols[3], Amelia Rodríguez[1], Rita Vassena[1,*]

KEY MESSAGE

This article identifies the current hot topics in human assisted reproduction worldwide and the temporal trends for the period 2005–2016. It provides an innovative picture of the current research that could help to explore the areas where further research is needed.

ABSTRACT

Research question: What are the current research trends in human assisted reproduction around the world?

Design: An analysis of 26,000+ scientific publications (articles, letters and reviews) produced worldwide between 2005 and 2016. The corpus of publications indexed in PubMed was obtained by combining the Medical Subject Heading (MeSH) terms: 'Reproductive techniques', 'Reproductive medicine', 'Reproductive health', 'Fertility', 'Infertility' and 'Germ cells'. An analysis was then carried out using text mining algorithms to obtain the main topics of interest.

Results: A total of 44 main topics were identified, which were then further grouped into 11 categories: 'Laboratory techniques', 'Male factor', 'Quality of ART, ethics and law', 'Female factor', 'Public health and infectious diseases', 'Basic research and genetics', 'Pregnancy complications and risks', 'General – infertility & ART', 'Psychosocial aspects', 'Cancer' and 'Research methodology'. The USA was the leading country in terms of number of publications, followed by the UK, China and France. Research content in high-income countries is fairly homogeneous across categories and it is dominated by 'Laboratory techniques' in Western-Southern Europe, and by 'Quality of ART, ethics and law' in North America, Australia and New Zealand. 'Laboratory techniques' is also the most abundant category on a yearly basis.

Conclusions: This study identifies the current hot topics on human assisted reproduction worldwide and their temporal trends for 2005–2016. This provides an innovative picture of the current research that could help explore the areas where further research is needed.

[1] Eugin, Barcelona 08006, Spain
[2] SIRIS Lab, Research Division of SIRIS Academic, Barcelona, Spain
[3] Ingenio (CSIC-UPV), Universitat Politècnica de València, València, Spain and Centre for Science and Technology Studies (CWTS), University of Leiden, the Netherlands

*Corresponding author. E-mail address: rvassena@eugin.es (R Vassena). https://doi.org/10.1016/j.rbmo.2019.10.013 1472-6483/
Declaration: The authors report no financial or commercial conflicts of interest.

KEYWORDS

Assisted reproduction
IVF
Latent Dirichlet allocation
Temporal trends
Text mining
Topic modelling

INTRODUCTION

Scientific research, especially in health and medical sciences, aims to address specific needs that society, and especially patients, perceive as pressing. One of the main challenges for policy-makers and research funders alike is therefore to align research priorities to societal needs (*Sarewitz and Pielke, 2007*). Research agendas can therefore be thought of in terms of a demand side (societal needs) and a supply side (research outputs).

Research output in human assisted reproduction has expanded in recent years, as indicated by the increasing number of high-quality scientific publications in indexed journals in this area. A previous analysis of publication metrics in reproductive medicine concluded that research in this field has attained scientific growth and maturity in the first decade of the century (2003–2012) (*Aleixandre-Benavent et al., 2015*). This report explores the most cited investigators, institutions and keywords in the field, with a clear focus on clinical aspects of human reproduction. However, a broader analysis of the literature, including basic and psychosocial aspects of research, has not yet been performed.

The description of the current research landscape enables an alignment of policies and funding, which can be achieved by comparing the research conducted with societal needs (*Cassi et al., 2017*; *Ciarli, 2019*). This is especially relevant to the field of human assisted reproduction research, where public funding is scarce, most research development follows local interests, and research priorities are usually not discussed at the national or international level.

Many techniques have been proposed to analyse, in a coarse-grained fashion, the outputs of scientific production (*Börner et al., 2003*; *Griffiths and Steyvers, 2004*; *van Eck and Waltman, 2007*). A fairly widespread approach is that of topic modelling (*Griffiths and Steyvers, 2004*; *Hofmann, 1999*; *Steyvers and Griffiths, 2007*), a field of machine learning that, given a textual corpus, seeks to infer the set of topics that generated the observed word sequences. This work focuses on the study of the current research supply in the field of human assisted reproduction and the use of assisted reproductive technologies (ART). Specifically, topic modelling is applied to infer the topics underlying the scientific publications produced by research on human assisted reproduction in the last decade.

MATERIALS AND METHODS

Corpus identification

This work aimed to determine the main research topics in human assisted reproduction tackled by scientific publications produced worldwide in the period 2005–2016.

The first step consisted of identifying the appropriate textual corpus to analyse. To properly limit the perimeter of the analysis, the bibliographic database PubMed of the US National Library of Medicine was used, together with the indexation of Medical Subject Headings (MeSH terms) to identify relevant publications. Scientific publications on PubMed were searched via a query that combined via a logical 'OR' the following MeSH terms: (i) reproductive techniques, (ii) reproductive medicine, (iii) reproductive health, (iv) fertility, (v) infertility, (vi) germ cells. To capture the core of the relevant literature, also included in the query were all publications that, albeit not indexed by the above MeSH terms, appeared in the following journals: *Human Reproduction*, *Human Reproduction Update*, *Molecular Human Reproduction*, *Fertility and Sterility* and *Reproductive Biomedicine Online*.

In order to obtain publications associated with human reproduction, but not contraception or research on animals for farming or veterinary purposes, specific journals out of the scope of this work were removed (*Theriogenology*, *Journal of Animal Sciences*, *Biology of Reproduction*, *Animal Reproduction Science*) and specific MeSH terms (contraception and contraceptive behaviour) via a NOT logical cause in the query.

A set of 108,377 was identified, all tagged with their respective PubMed identifier and all their associated MeSH terms.

Only documents of type 'article', 'letter' and 'review' were considered in the analysis. To filter out these types of records, all identified publications were retrieved in the bibliometric database Scopus, via their PubMed identifier, and retained their metadata, which included affiliations, keywords and references. The corpus was then enriched by including potential relevant publications that were not indexed in PubMed, by adopting the following strategy: first, all references of type 'article', 'letter' and 'review' cited by more than 50 records of the initial corpus and published after 2003 were included. Second, groups of 20 common keywords were extracted across the original corpus via matrix factorization (*Lin, 2007*). Third, a search of Scopus was carried out for records published after 2003, which featured combinations of four of those keywords plus the term 'assisted reproduction'.

A total of 106,101 publications potentially dealing with the research themes of interest for this work was collected. This corpus was cleaned by computing statistics of records for journals, excluding those journals that were not expected to be publishing research within the scope of analysis, extracting clusters of common MeSH terms (matrix factorization) and removing records associated with groups of MeSH outside the perimeter of the analysis.

Ultimately, by considering only those records for which an abstract was available, a final corpus was constructed of 22,133 documents, which were then studied by means of text mining algorithms.

Topic modelling

Before applying any text mining algorithms, all texts were pre-processed (*Bird et al., 2009*) by harmonizing to British spelling, identifying as many synonyms as possible and converting all variants into the same canonical form (e.g. converting into 'sperm index' all its variations) and by applying standard techniques of textual stemming (i.e. removing plurals) and textual chunking (i.e. recognizing 'human assisted reproduction' as a single word).

Topic modelling was then applied via latent Dirichlet allocation (LDA) (*Blei et al., 2003*) to all the selected abstracts. LDA assumes that the observed distribution of words in a textual corpus is determined by a statistical model that fixes both a word topic and a document topic association. Within LDA, one tries to infer document/topic and word/topic associations by maximizing the likelihood that the observed corpus is indeed generated by the theoretical model. Here, the Gibbs sampling strategy was adopted (*Bishop, 2006*), carried out via the Mallet software (*McCallum, 2002*).

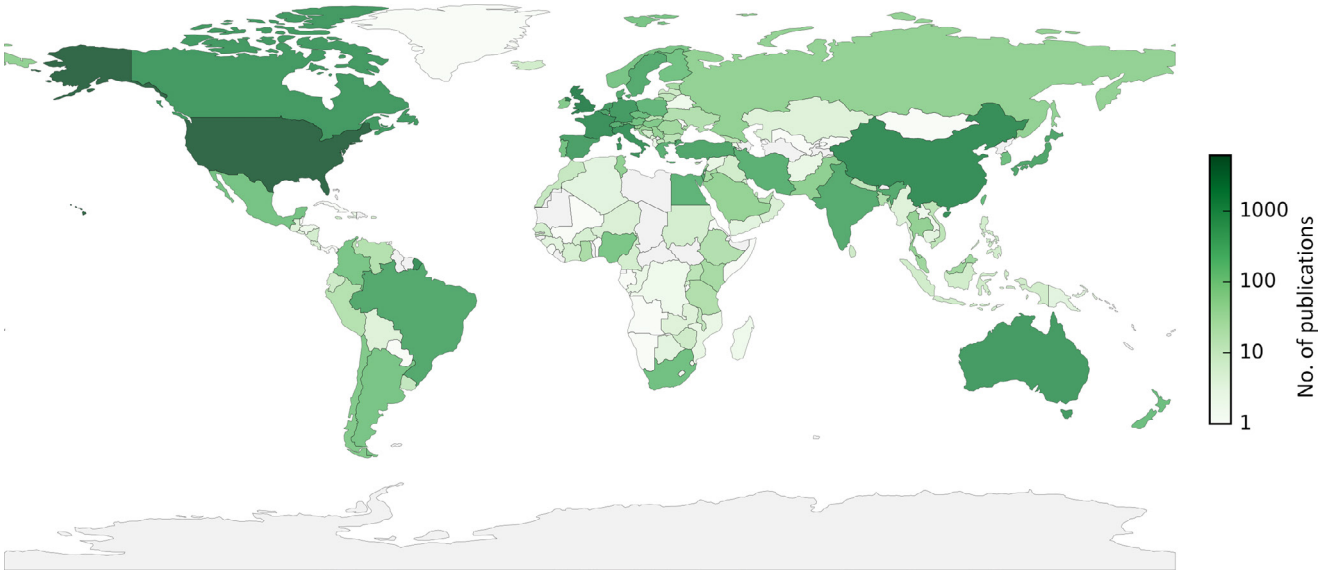


FIGURE 1 World distribution of publications on human assisted reproduction research. The USA, the UK, China, France and Italy dominate the ranking.

The results of the LDA algorithm consist of a list of topics, and in the weighted relations between each document in the corpus and each topic. In practice, every topic is a list of characterizing words and each document may be connected to more than one single topic. In LDA, the number of topics is left as a free parameter, and there are several possible ways to fix this number (*Arun et al., 2010*). In this work, the number of topics was selected by first evaluating the maximum log-likelihood as a function of the number of topics and then the number of topics for which the log-likelihood reached a plateau was chosen. This procedure allowed the selection of 44 topics.

Graphical representation of topics

All the topics are further represented on a map, arranged by the software ldavis (*Sievert and Shirley, 2014*) in a two-dimensional plane via multidimensional scaling (*Borg and Groenen, 2006*). On this map, each circle represents a topic, and closer topics share a broader vocabulary (i.e. they are more similar). The area of the circle is proportional to the frequency of the topic in the corpus, while the colour denotes which macro-category the topic belongs to. The 44 topics were grouped into 11 macro-categories via a qualitative inspection of the conceptual content of each topic. This grouping is consistent with the spatial arrangement provided in the map: topics belonging to the same macro-category are located in neighbouring regions.

Because the affiliation of all authors of the collected publications was known, the topics of each document can be transitively associated with the respective affiliations and countries. This linkage allows topics and topic group distributions to be derived at the geographical level. Here, two analyses were performed looking at distributions of categories by geographical regions: one by using the world regions defined by the United Nations Statistics Division (*United Nations Statistics Division, 2017*), and another by using the income levels defined by the World Bank (*WorldBankData, 2017*).

RESULTS

Human assisted reproduction research output per country and per year

The number of publications finally considered in this study, sorted per country of production, is displayed on a map in FIGURE 1. Not surprisingly, most of the records retrieved are produced in the USA, followed by the UK and China. France and Italy complete the top 5 of the most active countries in research on assisted reproduction in the time period analysed.

FIGURE 2 shows the trend in the number of documents retrieved per year between 2005 and 2016, FIGURE 2A and the variation with respect to the production in 2005, FIGURE 2B. A growth in scientific production of about 120% between 2005 and 2015 was observed. As a reference, also shown is the number of documents returned by PubMed by simply searching for the MeSH term 'IVF' and the combination of MeSH terms 'IVF' and 'Humans'. As can be seen, this actually covers a much wider perimeter of analysis (the number of documents per year of the dataset in FIGURE 2A is much bigger than those baselines), but we are analysing a production whose growth rate is very similar to that of research on IVF only, thus suggesting we are not enlarging our perimeter in an inconsistent manner.

Human assisted reproduction research production by topic

The list of all 44 topics and the 10 main terms included in each is provided in TABLE 1, sorted in decreasing size: the first topic is the most frequent, while the last one is the smallest in the corpus.

The set of words of a topic allows a grasp of what the topic is about and an understanding of whether it is related to clinical, basic or social research. For instance, the first topic deals with general themes regarding ART, the second topic with ovarian stimulation and the third with pregnancy risks and complications. Notably, it was possible to extract topics specifically related to either the female side of assisted reproduction research (e.g. topics 5, 28 and 29) or the male (e.g. topics 6, 12 and 25). This result allows quantification of the amount of research efforts devoted to either male or female related research and to further drill down into the specific

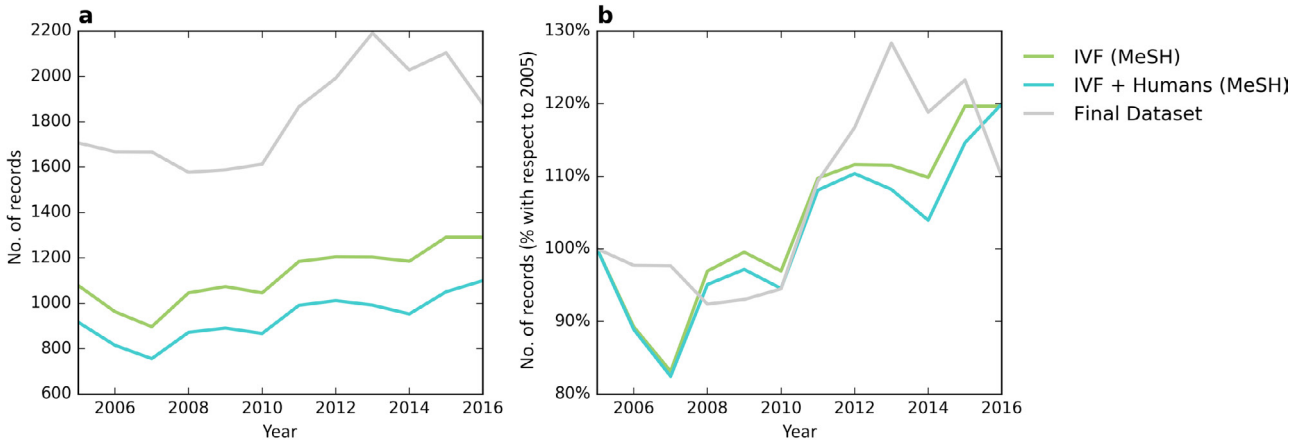


FIGURE 2 Comparison of temporal trends between the dataset analysed in the present paper (Final Dataset) and documents yielded by PubMed by searching by MeSH term 'IVF' and the combination of the MeSH terms 'IVF' and 'Humans'. Documents per year (a) and growth rate (b). Publications were collected publications well beyond research on IVF, but in a research field which features a similar growth rate to that of research on IVF.

themes. It should be mentioned that the terms preimplantation genetic screening (PGS) and preimplantation genetic diagnosis (PGD), which appear in topic 32, are no longer used, having been recently replaced by the term preimplantation genetic testing (PGT) (*Zegers-Hochschild et al., 2017*).

The 11 macro-categories are shown in the second column of TABLE 1, and these deal with basic research, clinical and psychosocial aspects of research on assisted reproduction.

Regions of topics

FIGURE 3 shows the map of all the 44 topics represented, grouped by colour into the 11 macro-categories. The upper part of the graph can be identified as the region of 'Laboratory techniques', the lower left part mostly as the region focused on 'Basic research' and the lower right as the one concerned with the 'Psychosocial aspects of assisted reproduction research'. Female factor research is almost homogeneously distributed along the x-axis. Interestingly, a few topics dealing with 'Cancer' and 'Pregnancy complications' lie in between the 'Female factor' and the 'Psychosocial aspects' region. In a similar fashion, 'Male factor' research seems to be an intermediary between 'Female factor' and 'Basic research'.

The interactive version of this figure is available online at http://sirislab.com/lab/EUGIN/ARTtopicLDA.html. In this, the reader can explore all the individual terms included in each topic defining a macro-category, and their relative importance in each.

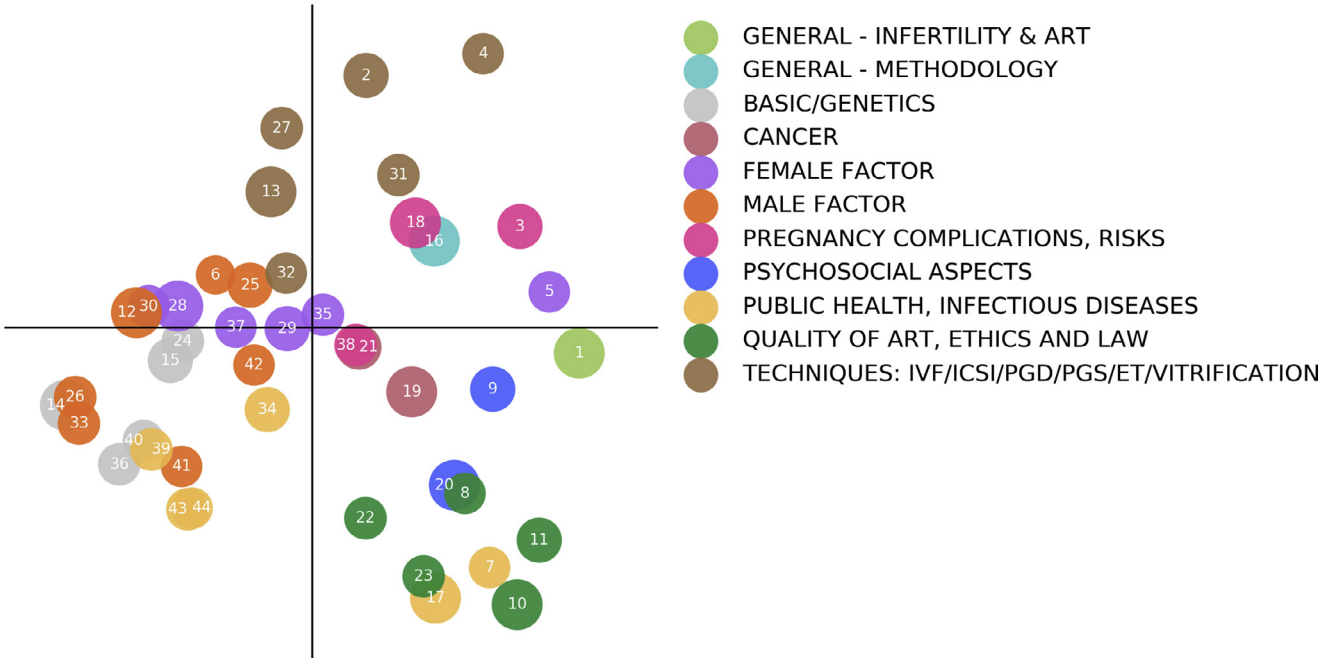


FIGURE 3 A map of the 44 topics extracted via latent Dirichlet allocation (LDA) from the corpus analysed in the present work. Coordinates are calculated via a dimensional reduction algorithm: each circle is a topic; closer circles denote similar topics. The size of the circles is proportional to the weight of the topic in the whole corpus, while colours denote the macro-categories to which each topic was manually mapped.

TABLE 1 LIST OF TOPICS, TERMS INCLUDED IN EACH TOPIC AND TOPIC GROUP

Topic	Topic group	Term 1	Term 2	Term 3	Term 4	Term 5	Term 6	Term 7	Term 8	Term 9	Term 10
1	GENERAL – INFERTILITY & ART	infertility	ART	treatment	couple	pregnancy	risk	technique	method	management	fertility
2	LABORATORY TECHNIQUES	day	HCG	ovarian stimulation	oocyte	cycle	controlled ovarian hyperstimulation	stimulation	GnRH agonist	follicle	oocyte retrieval
3	PREGNANCY COMPLICATIONS AND RISKS	ART	child	pregnancy	risk	birth	twin	singleton	birth weight	increase risk	ICSI
4	LABORATORY TECHNIQUES	embryo transfer	day	embryo	cycle	transfer	live birth rate	live birth	pregnancy	clinical pregnancy rate	pregnancy rate
5	FEMALE FACTOR	age	infertility	pregnancy	risk	risk factor	time	BMI	exposure	men	year
6	MALE FACTOR	men	semen quality	varicocele	semen concentration	semen analysis	seminal parameter	male infertility	infertile men	semen motility	semen count
7	PUBLIC HEALTH AND INFECTIOUS DISEASES	adolescent	sexual health	men	reproductive health	young people	sex	sexually transmitted infection	sexuality	HIV	intervention
8	QUALITY, ETHICS AND LAW	embryo	law	ART	child	decision	regulation	couple	debate	embryo donation	donation
9	PSYCHOSOCIAL ASPECTS	infertility	couple	men	treatment	depression	infertile woman	anxiety	partner	infertile couple	quality
10	QUALITY, ETHICS AND LAW	care	guideline	quality	clinic	health care	doctor	nurse	recommendation	health professional	health care provider
11	QUALITY, ETHICS AND LAW	physician	respondent	gynaecologist	survey	questionnaire	attitude	obstetrician gynaecologists	abortion	recommendation	response rate
12	MALE FACTOR	DNA fragmentation	spermatozoon	DNA damage	semen sample	DFI	DNA integrity	TUNEL	fragmentation	integrity	men
13	LABORATORY TECHNIQUES	oocyte	embryo	day	vitrification	blastocyst	culture medium	human oocyte	cryopreservation	fertilization	IVM
14	BASIC RESEARCH AND GENETICS	spermatozoon	protein	human semen	human spermatozoon	acrosome reaction	capacitation	expression	semen motility	progesterone	cell
15	BASIC RESEARCH AND GENETICS	male infertility	gene	deletion	polymorphism	frequency	mutation	azoospermia	infertile men	microdeletions	SNP
16	METHODOLOGY RESEARCH	meta-analysis	live birth	systematic review	treatment	live birth rate	controlled trial	quality evidence	clinical pregnancy	intervention	bias
17	PUBLIC HEALTH AND INFECTIOUS DISEASES	reproductive health	health	policy	health service	health care	intervention	family planning	community	cost	government
18	PREGNANCY COMPLICATIONS AND RISKS	pregnancy	case report	ectopic pregnancy	gestation	risk	hyperstimulation syndrome	diagnosis	syndrome	embryo transfer	treatment
19	CANCER	cancer	fertility preservation	fertility	treatment	cancer patient	chemotherapy	cancer treatment	diagnosis	breast cancer	risk
20	PSYCHOSOCIAL ASPECTS	child	donor	parent	family	oocyte donation	mother	recipient	oocyte donor	semen donor	couple
21	CANCER	endometriosis	surgery	disease	treatment	recurrence	hysterectomy	laparoscopy	cervical cancer	diagnosis	ovarian cancer
22	QUALITY, ETHICS AND LAW	resident	training	medical student	performance	trainee	obstetrics	score	skill	obstetrics and gynaecology	time

(continued on next page)

Table 1 – (continued)

Topic	Topic group	Term 1	Term 2	Term 3	Term 4	Term 5	Term 6	Term 7	Term 8	Term 9	Term 10
23	QUALITY, ETHICS AND LAW	obstetrics and gynaecology	resident	obstetrics	training	survey	trainee	residency	respondent	specialty	physician
24	BASIC RESEARCH AND GENETICS	chromosome	frequency	aneuploidy	chromosomal abnormality	FISH	carrier	translocation carrier	translocation	spermatozoon	karyotype
25	MALE FACTOR	men	azoospermia	semen extraction	obstructive azoospermia	FSH	spermatozoon	semen analysis	testis	non-obstructive azoospermia	male infertility
26	MALE FACTOR	oxidative stress	spermatozoon	oxygen species	semen motility	human spermatozoon	concentration	lipid peroxidation	apoptosis	semen function	motility
27	LABORATORY TECHNIQUES	ICSI	semen injection	oocyte	fertilization	fertilization rate	spermatozoon	couple	embryo quality	IMSI	fertilization failure
28	FEMALE FACTOR	follicular fluid	serum	concentration	endometriosis	granulosa cell	antibody	infertile woman	ASA	antisperm antibody	ELISA
29	FEMALE FACTOR	HSG	sensitivity	specificity	hysteroscopy	procedure	laparoscopy	tubal patency	HSG	fallopian tube	uterine cavity
30	FEMALE FACTOR	expression	endometrium	implantation failure	implantation	cell	infertile woman	unexplained infertility	endometriosis	endometrial receptivity	endometrial biopsy
31	LABORATORY TECHNIQUES	IUI	couple	unexplained infertility	treatment	pregnancy	clomiphene citrate	pregnancy rate	cycle	infertility	ovulation
32	LABORATORY TECHNIQUES	preimplantation genetic diagnosis	embryo	preimplantation genetic screening	aneuploidy	couple	day	biopsy	transfer	chromosome	advanced maternal age
33	MALE FACTOR	expression	testis	germ cell	spermatogenesis	Sertoli cell	cryptorchidism	cell	mRNA	spermatogonia	Leydig cell
34	PUBLIC HEALTH AND INFECTIOUS DISEASES	HIV	*C. trachomatis*	infection	*C. trachomatis* infection	infertility	risk	PCR	detection	ureaplasma urealyticum	transmission
35	FEMALE FACTOR	infertility	adenomyosis	uterus	fertility	myomectomy	fibroid	treatment	diagnosis	hysteroscopy	endometrial polyp
36	BASIC RESEARCH AND GENETICS	DNA methylation	gene	expression	methylation	RNA	telomere length	methylation pattern	IGF	gene expression	DNA
37	FEMALE FACTOR	AMH	ovarian reserve	antral follicle count	age	FSH	follicle count	AMH	DOR	AMH level	inhibin b
38	PREGNANCY COMPLICATIONS AND RISKS	tocolysis	delivery	pregnancy	preterm labour	preterm birth	gestational age	pregnant woman	Caesarean section	gestation	birth
39	PUBLIC HEALTH AND INFECTIOUS DISEASES	method	mtDNA	spermatozoon	DNA	mitochondrial DNA	detection	semen cell	cell	propofol	victim
40	BASIC RESEARCH AND GENETICS	stem cell	cell	human	oocyte	cumulus cell	cell type	mitochondrion	SCNT	somatic cell	granulosa cell
41	MALE FACTOR	device	method	exposure	laboratory	CASA	worker	variation	VCL	measurement	lead
42	MALE FACTOR	treatment	acupuncture	treatment group	andrology	SCI	significant difference	cord injury	medication	pregnancy rate	observation group

(continued on next page)

Table 1 – (continued)

Topic	Topic group	Term 1	Term 2	Term 3	Term 4	Term 5	Term 6	Term 7	Term 8	Term 9	Term 10
43	PUBLIC HEALTH AND INFECTIOUS DISEASES	HPV	vaccine	vaccination	human papilloma virus	HPV vaccine	HPV infection	HPV vaccination	hybrid	cervical cancer	growth
44	PUBLIC HEALTH AND INFECTIOUS DISEASES	celiac disease	hepatitis B virus	hepatitis B virus infection	medicine	gynaecology and obstetrics	hepatitis B virus DNA	activity	site	th century	film

AMH = anti-Müllerian hormone; ART = assisted reproductive technology; ASA = acetylsalicylic acid; BMI = body mass index; CASA = computer-assisted semen analysis; DOR = diminished ovarian reserve; ELISA = enzyme-linked immunosorbent assay; DFI = DNA Fragmentation Index; FISH = fluorescence in-situ hybridization; FSH = Follicle Stimulating Hormone; GnRH = gonadotrophin-releasing hormone; HCG = human chorionic gonadotrophin; HIV = human immunodeficiency virus; HPV = human papilloma virus; HSG = hysterosalpingography; ICSI = intracytoplasmic sperm injection; IGF = insulin-like growth factor; IMSI = intracytoplasmic morphologically selected sperm injection; IUI = intrauterine insemination; IVM = in-vitro maturation; mtDNA = mitochondrial DNA; PCR = polymerase chain reaction; SCI = Spinal Cord Injury; SCNT = somatic cell nuclear transfer; SNP = single-nucleotide polymorphism; TUNEL = terminal deoxynucleotidyl transferase-mediated dUDP nick-end labelling; VCL = curvilinear velocity.

World human assisted reproduction research production per category

Grouping topics into macro-categories provides an effective coarse-grained view of worldwide scientific production. In particular, it is possible to assess size and growth or decline trends of the categories. FIGURE 4 shows the percentage of publications linked to each of the macro-categories. It can be seen that most of the publications focus on the different techniques applied in assisted reproduction (i.e. IVF, intracytoplasmic sperm injection [ICSI], genetic tests), grouped in the macro-category 'Laboratory techniques'. Notably, 'Male factor' is the second largest macro-category in percentage of publications. Interestingly, the third category is 'Quality of ART, ethics and law'. This result indicates that ART is progressively including research on policies and on ethical implications. Conversely, the category 'Psychological aspects' is the third smallest, suggesting there could be room for growth for research dealing with the psychological aspects of assisted reproduction.

Human assisted reproduction research production per geographical region

Research in Western world regions (Western Europe, Northern America, and Australia and New Zealand) is fairly homogeneously distributed across macro-categories and relatively similar among different regions, with some minor differences (FIGURE 5). For instance, a relative skew in research on 'Laboratory techniques' can be detected in Western Europe and on 'Quality of ART, ethics and law' in Northern America, respectively. On the other side of the spectrum, Sub-Saharan Africa devotes most of their research efforts to 'Public health and infectious diseases' issues. It is also apparent how most of the countries outside the Western world are more skewed in their human assisted reproduction research towards 'Male factor' rather than 'Female factor'.

Similar findings are apparent when grouping countries by income (FIGURE 6). High-income countries have a fairly balanced research production, with a relative abundance of research on 'Laboratory techniques' and 'Quality of ART, ethics and law' (the latter, especially, driven by research in the USA). The relative importance of research on 'Public health and infectious diseases' seems to grow when the level of income decreases. Also, research on 'Male factor' has a higher weight in middle-income countries with respect to high-income countries. Similarly, and not surprisingly, the amount of research on 'Laboratory techniques' seems to increase with income.

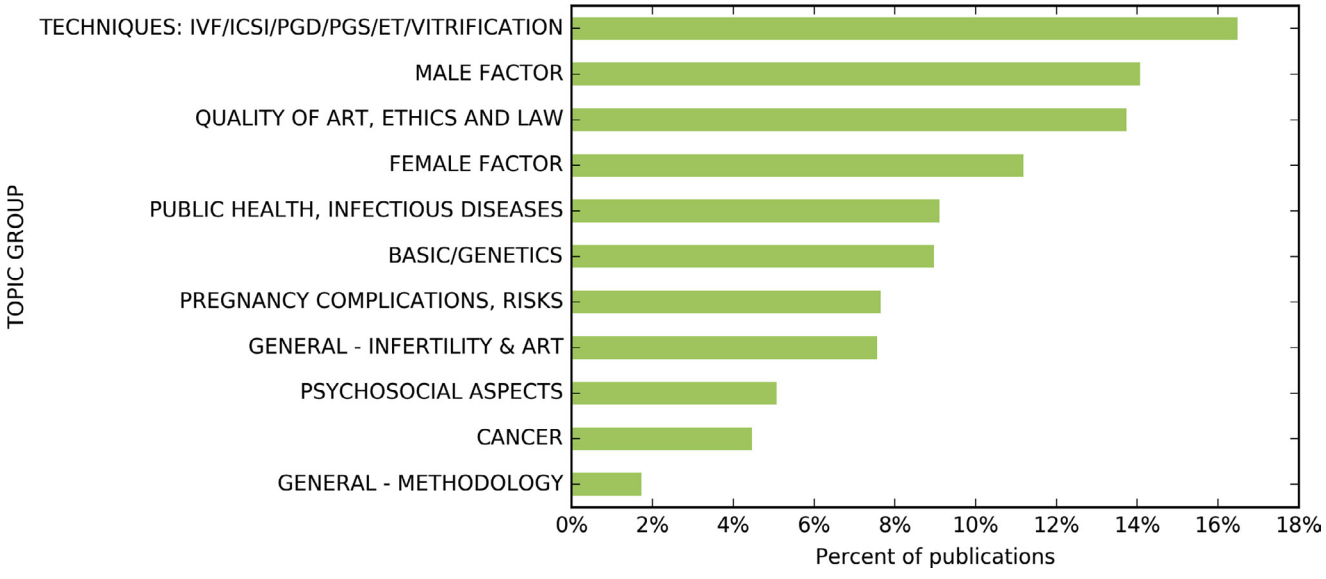


FIGURE 4 Percentage of publications for each of the macro-categories defined in the present work.

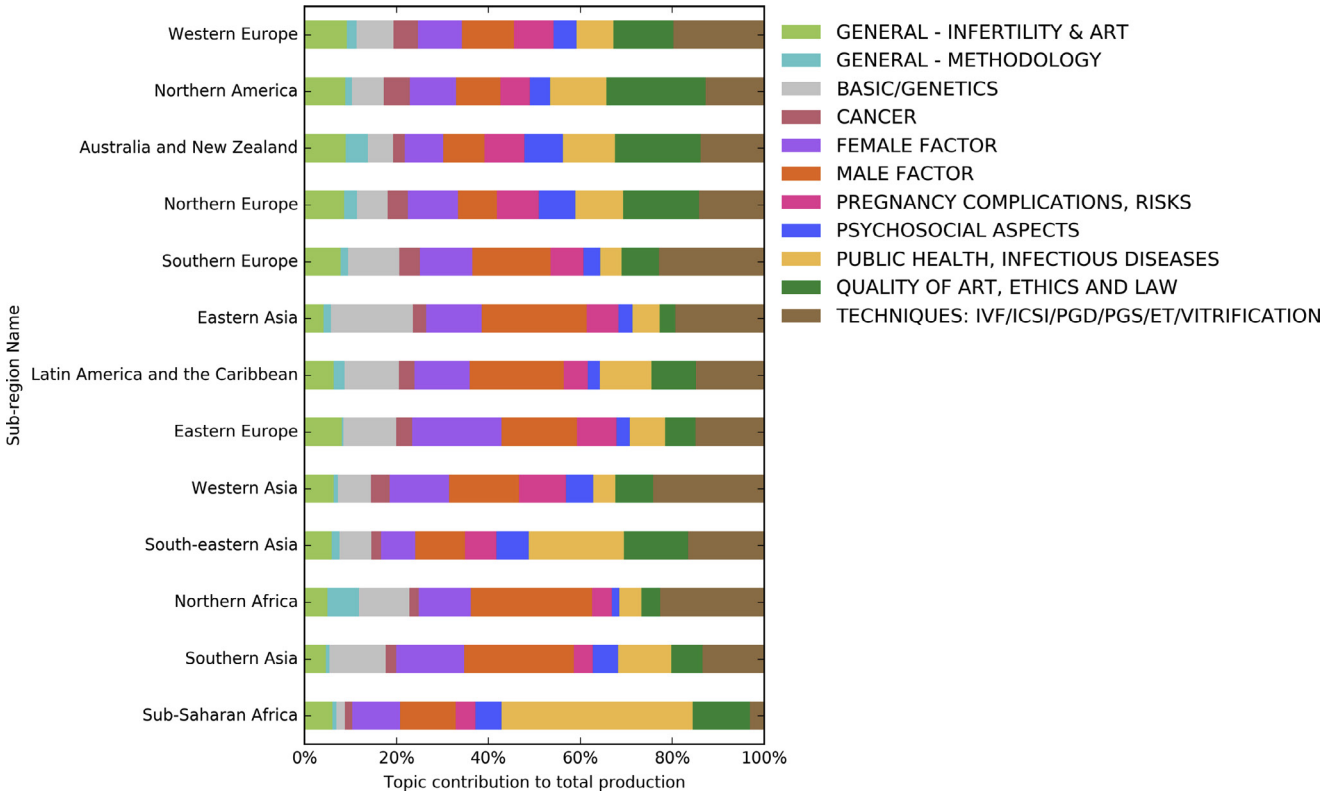


FIGURE 5 Percentage of publications of each macro-category at geographical grouping level. Each group is defined by the United Nations (*United Nations Statistics Division, 2017*); publications are associated with all affiliations.

Temporal trends in human assisted reproduction research

The analysis of temporal evolution of research output for each specific category can help to understand whether there are hot or cold themes in human assisted reproduction research. FIGURE 7 shows the temporal trends of each macro-category of research; FIGURE 7A represents the weighted number of publications and FIGURE 7B the evolution of the percentage weight of each of the

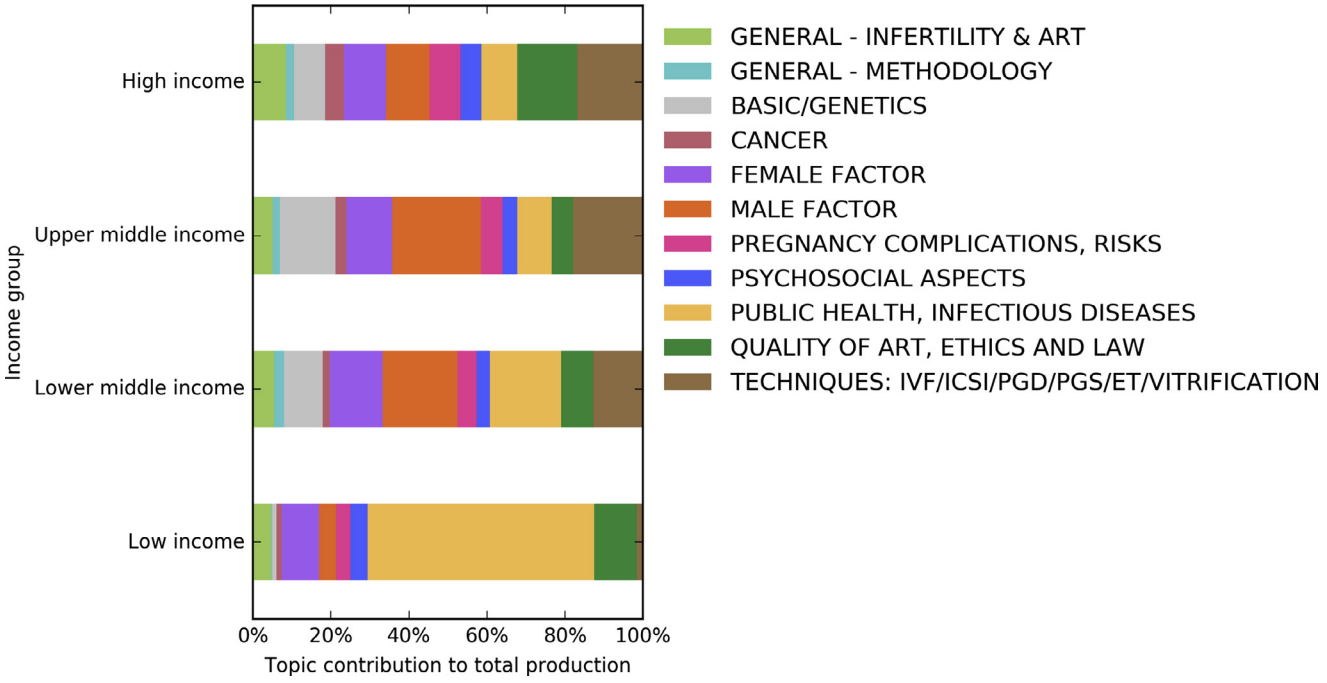


FIGURE 6 Percentage of publications of each macro-category for countries grouped for income level. Each group is defined by the World Bank (*WorldBankData 2017*), publications are associated to all affiliations.

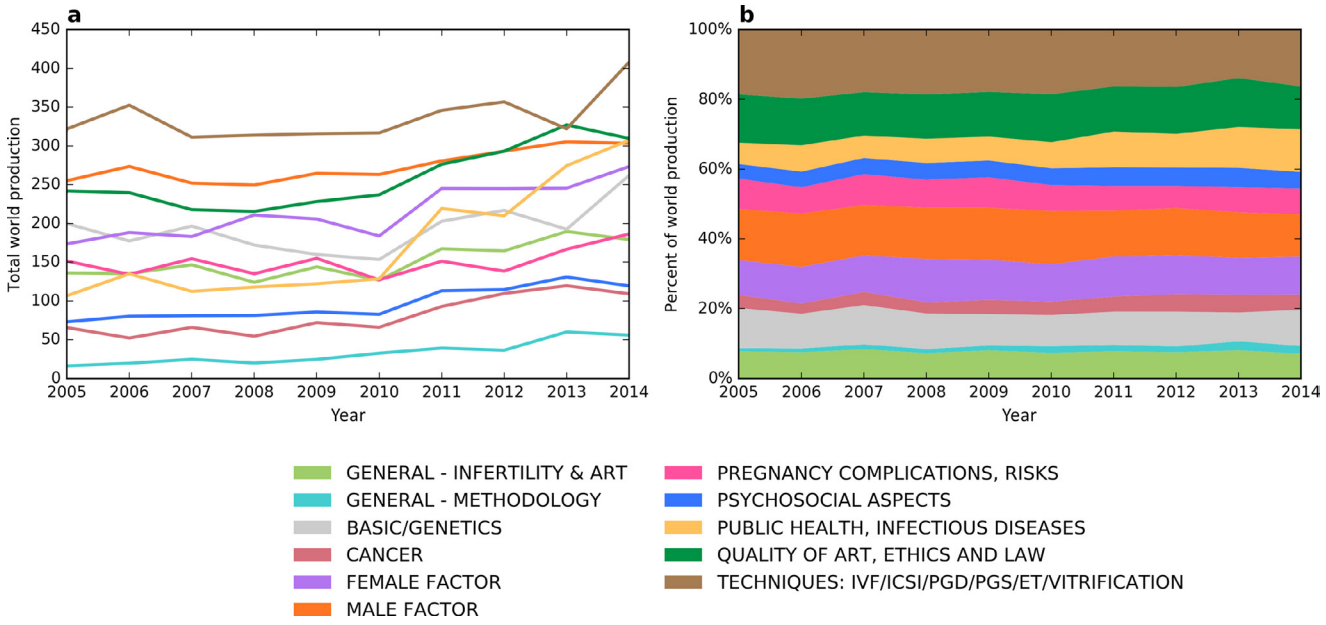


FIGURE 7 Temporal trends of each research macro-category in terms of weighted number of publications (a) and relative yearly weight (b). Some categories attain a constant evolution (e.g. techniques) and some categories grow. Because the total number of publications increases, categories with constant trends actually experience a decline in their percentage contribution to the total production, while those categories which are growing, such as the total production, keep a constant percentage weight.

categories. Overall, the patterns are quite stable. 'Laboratory techniques' is relatively constant and the most abundant on a yearly basis (FIGURE 7A). However, because production in most of the other categories is increasing, the relative contribution of this research category is actually decreasing slightly (FIGURE 7B). Research on 'Quality of ART, ethics and law' shows a slight increase, in line with the general trends, so that its weight remains roughly constant. Research on 'Public health and infectious diseases' shows a rapid increase and gains a conspicuous percentage of the total production already in 2011. Research on the 'Female factor' shows a similar trend, slowly becoming a major research theme.

To gain further insights, the analysis can again be broken down at the level of country income groups, as shown in FIGURE 8. One can see that, for high-income countries, research on 'Quality of ART, ethics and law' has caught up research on 'Laboratory techniques'; research on the 'Male factor' is slowly decreasing, while research on 'Public health and infectious diseases' is increasing. Middle-income countries have been slowly increasing their research on 'Laboratory techniques', catching up with the pattern shown by high-income countries. Lower middle-income countries experienced a peak of research on the 'Male factor' in 2011 and then a decrease and share, with low-income countries, a very marked growth of research on 'Public health and infectious diseases'.

DISCUSSION

This article identifies the current hot topics in research on human assisted reproduction worldwide and the temporal trends for the period 2005–2016. It provides an innovative picture of the current research in basic, clinical and psychosocial aspects of assisted reproduction. This could help explore the areas where further research is needed, provided questions in these areas are important, not already answered, and of interest to patients. The societal demand side of research is not analysed in the present article, because it requires the involvement of societal stakeholders, supported by extensive fieldwork, carefully designed surveys and the compiling of statistical data (*Ciarli, 2019*; *Petit-Zeman et al., 2010*; *Ràfols and Yegros, 2018*).

Human assisted reproduction includes infertility treatments for male, female or unknown factors, and also treatments for people who resort to ART due to a lack of the male and/or female gamete. It also includes gamete vitrification for medical and social reasons. Therefore, this analysis refers to research around human assisted reproduction and it is not limited to the scope of infertility. However, it has to be taken into account that, although infertility is a disease (*Zegers-Hochschild et al., 2017*) that affects couples and individuals globally, access to ART in low-income countries is limited, even non-existent (*Inhorn and Patrizio, 2015*). This may explain the low amount of research on ART performed in these countries; according to the results of this study, research in these countries is more limited and more focused on 'Public health and infectious diseases'. In agreement with *Aleixandre et al. (2015)*, it was observed that most research has been performed in North America, Western Europe and Australia, with some differences among world regions in the themes of research. In contrast to the mentioned article, however, it was found that China is among the three countries with the largest research output, maybe due to the different search strategy used in this study.

Globally, the four main areas of research are 'Laboratory techniques', 'Male factor', 'Quality of ART, ethics and law' and 'Female factor'. Research on 'Female factor' is chiefly developed by middle- to high-income countries, while research on 'Laboratory techniques' is more prevalent in high-income countries, a feature that could be related to the higher complexity

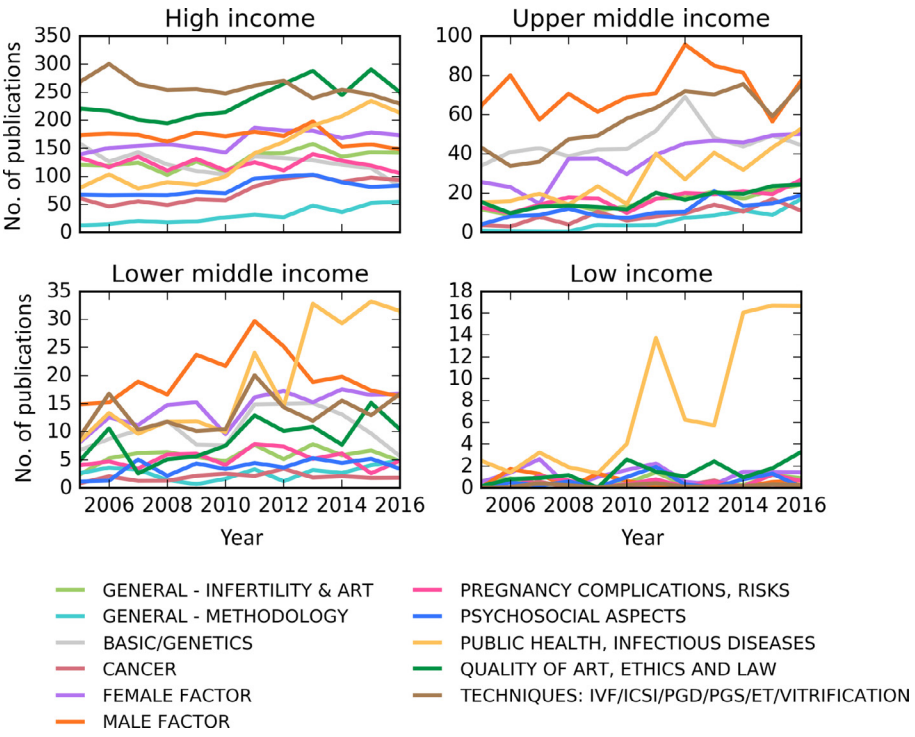


FIGURE 8 Temporal trends of each research macro-category analysed in the present paper, broken down at country groups for income level. Each group is defined by the World Bank (*WorldBankData 2017*); publications are associated with all affiliations.

of research on these topics and the larger associated costs. Relative research efforts on 'Male factor' compared with 'Female factor' are not related to income group, but clear differences are observed among regions. 'Male factor' predominates in all regions of Asia, Latin America, Northern Africa and Southern Europe, while 'Female factor' heads research only in Eastern Europe. In Western Europe, Northern America, Australia and Sub-Saharan Africa the ratio between male and female factor is nearly 1. This suggests that the decision to carry out ART research on a certain topic might, in part, be shaped by social and cultural aspects. One aspect to be taken into account because it could have biased these findings is that the definition of infertility as of male or female origin is not always consistent across the countries, and the prevalence of male infertility in the general population (whether in a relationship or not) has not yet been established (*Barratt et al., 2017*).

Regarding temporal trends, total scientific production worldwide is increasing yearly, especially in the areas of 'Public Health and infectious diseases' and 'Female factor'. As a consequence, the dominant area of 'Laboratory techniques' is losing relative prevalence at international level. These results are in line with previous work showing that ICSI is one of the most frequent keywords in reproductive medicine for the same period, with a downward trend (*Aleixandre-Benavent et al., 2015*).

There are three main factors that could limit the robustness of this piece of work: first, the textual corpus analysed is based on abstract and titles; second, the reproducibility of the stochastic algorithms applied, which may produce slightly differing results at each run, and third, the interpretation of the topics obtained. As a result, the analysis was focused on a number of publications. This excludes the grey literature, which could certainly add comprehensive information to the approach used here, but the non-systematized nature of this source of information (and its questionable quality) prevented its inclusion in the analysis. It is important to note that increased quality in studies also leads to improved quality of care (*Duffy et al., 2017*), although assessing how to improve quality of care is no easy task. For example, several treatments/techniques seem to increase pregnancy outcomes only slightly, but even these small effects need large (and expensive) trials to achieve statistical significance (*Evers, 2013*). Finally, it was not possible to distinguish between research performed with either public or private funding, therefore research undergone to answer societal needs may be biased by research in favour of the interests of private companies.

In conclusion, this work may help open up the debate on research focus and actual needs in the realm of human assisted reproduction, from the understanding of current ART research focus, and discussing future research directions.

ACKNOWLEDGEMENTS

We acknowledge the efforts of researchers in human assisted reproduction for their contribution to this work.

FAM acknowledges financial support from grant PTQ-14-06718 of the Spanish MINECO Torres Quevedo programme.

REFERENCES

Aleixandre-Benavent, R., Simon, C., Fauser, B.C. **Trends in clinical reproductive medicine research: 10 years of growth.** Fertil. Steril. 2015; 104

Arun, R., Suresh, V., Veni Madhavan, C.E., Narasimha Murthy, M.N. **On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations.** In Zaki M.J., Yu J.X., Ravindran B., Pudi V. Advances in Knowledge Discovery and Data Mining Springer Berlin Heidelberg Berlin, Heidelberg 2010: 391–402

Barratt, C.L.R., Bjorndahl, L., De Jonge, C.J., Lamb, D.J., Osorio Martini, F., McLachlan, R., Oates, R.D., van der Poel, S., St John, B., Sigman, M. **The diagnosis of male infertility: an analysis of the evidence to support the development of global WHO guidance-challenges and future research opportunities.** Hum. Reprod. Update 2017; 23: 660–680

Bird, S., Klein, E., Loper, E. 2009 **Natural language processing with Python: analysing text with the natural language toolkit.** 'O'Reilly Media, Inc.'

Bishop, C.M. 2006 **Pattern recognition and machine learning.** springer

Blei, D.M., Ng, A.Y., Jordan, M.I. **Latent dirichlet allocation.** Journal of machine Learning research 2003; 3: 993–1022

Borg, I., Groenen, P. **Modern Multidimensional Scaling: Theory and Applications.** Journal of Educational Measurement 2006; 40: 277–280

Börner, K., Chen, C., Boyack, K.W. **Visualizing knowledge domains.** Annual review of information science and technology 2003; 37: 179–255

Cassi, L., Lahatte, A., Rafols, I., Sautier, P., De Turckheim, E. **Improving fitness: Mapping research priorities against societal needs on obesity.** Journal of Informetrics 2017; 11: 1095–1113

Ciarli TaR, I. **The Relation between Research Priorities and Societal Demands: The Case of Rice.** Research Policy 2019; 48: 949–967

Duffy, J., Bhattacharya, S., Herman, M., Mol, B., Vail, A., Wilkinson, J., Farquhar, C., Cochrane, G., Fertility, G. **Reducing research waste in benign gynaecology and fertility research.** BJOG 2017; 124: 366–369

Evers, J.L. **The wobbly evidence base of reproductive medicine.** Reprod. Biomed. Online 2013; 27: 742–746

Griffiths, T.L., Steyvers, M. **Finding scientific topics.** Proceedings of the National Academy of Sciences 2004; 101: 5228–5235

Hofmann, T. **Probabilistic latent semantic analysis.** Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence 1999: 289–296

Inhorn, M.C., Patrizio, P. **Infertility around the globe: new thinking on gender, reproductive technologies and global movements in the 21st century.** Hum. Reprod. Update 2015; 21: 411–426

Lin, C.-J. **Projected Gradient Methods for Nonnegative Matrix Factorization.** Neural Computation 2007; 19: 2756–2779

Mccallum A.K. MALLET: A Machine Learning for Language Toolkit. 2002.

Petit-Zeman, S., Firkins, L., Scadding, J.W. **The James Lind Alliance: tackling research mismatches.** Lancet 2010; 376: 667–669

Ràfols I., and Yegros, A. Is research responding to health needs? Social Observatory of La Caixa Foundation (refereed online publication), March 2018. 2018.

Sarewitz, D., Pielke, R.A.Jr **The neglected heart of science policy: reconciling supply of and demand for science.** environmental science and policy 2007; 10: 5–16

Sievert, C., Shirley, K. **LDAvis: A method for visualizing and interpreting topics.** Proceedings of the workshop on interactive language learning, visualization, and interfaces 2014: 63–70

Steyvers, M., Griffiths, T. **Probabilistic topic models.** Handbook of latent semantic analysis 2007; 427: 424–440

United Nations Statistics Division. Standard country or area codes for statistical use (M49). https://unstats.un.org/unsd/methodology/m49/. 2017.

van Eck, N.J., Waltman, L. **VOS: A New Method for Visualizing Similarities Between Objects.** In Decker R., Lenz H.-J. Advances in Data Analysis: Proceedings of the 30th Annual Conference of the Gesellschaft für Klassifikation e.V., Freie Universität Berlin, March 8–10, 2006 Springer Berlin Heidelberg 2007Berlin, Heidelberg: 299–306

World Bank Data 2017. World Bank Country and Lending Groups 2018.

Zegers-Hochschild, F., Adamson, G.D., Dyer, S., Racowsky, C., de Mouzon, J., Sokol, R., Rienzi, L., Sunde, A., Schmidt, L., Cooke, I.D., Simpson, J.L., van der Poel, S. **The International Glossary on Infertility and Fertility Care, 2017.** Hum. Reprod. 2017; 32: 1786–1801

Received 19 July 2019; received in revised form 17 October 2019; accepted 18 October 2019.