Big data and artificial intelligence (AI) are poised to transform infertility healthcare

Ines de Santiago[1] & Lukasz Polansky[2]

[1]Cancer Research UK, Cambridge Institute. University of Cambridge, UK
[2]Assisted Conception Unit, Guy's Hospital, London, UK

Abstract

Advances in machine learning (ML) and artificial intelligence (AI) are transforming the way we treat patients in ways not even imagined a few years ago. Cancer research is at the forefront of this movement. Infertility, though not a life-threatening condition, affects around 15% of couples trying for a pregnancy. Increasing availability of large datasets from various sources creates an opportunity to introduce ML and AI into infertility prevention and treatment.

At present in the field of assisted reproduction, very little is done in order to prevent infertility from arising, with the main focus put on treatment when often advanced maternal age and low ovarian reserve make it very difficult to conceive. A shift from this disease-centric model to a health centric model in infertility is already taking place with more emphasis on the patient as an active participator in the process. Poor quality and incomplete data as well as biological variability remain the main limitations in the widespread and reliable implementation of AI in the field of reproductive medicine. That said, one of the areas where this technology managed to find a foothold is identification of developmentally competent embryos. More work is required however to learn about ways to improve natural conception, the detection and diagnosis of infertility, and improve assisted reproduction treatments (ART) and ultimately, develop clinically useful algorithms able to adjust treatment regimens in order to assure a successful outcome of either fertility preservation or infertility treatment.

Progress in genomics, digital technologies and advances in integrative biology has had a tremendous impact on research and clinical medicine. With the rise of 'big data', artificial intelligence, and the advances in molecular profiling, there is an enormous potential to transform not only scientific research progress, but also clinical decision making towards predictive, preventive, and personalized medicine. In the field of reproductive health there is now an exciting opportunity to leverage these technologies and develop more sophisticated approaches to diagnose and treat infertility disorders. In this review, we present a comprehensive analysis and interpretation of different innovation forces that are driving the emergence of a system approach to the infertility sector. Here we discuss recent influential work and explore the limitations of the use of Machine Learning models in this rapidly developing area.

Introduction

Infertility defined as failure to conceive after 1 year or more of regular unprotected sexual intercourse is estimated to affect 15% of couples of reproductive age[1,2], with regional variations affected mainly by socio-economic status[3]. Though robust epidemiological data is lacking, there is a possible trend in worsening of this pandemic, especially in sub-Saharan Africa and South Asia[4]. In most cases of infertility, the cause is known with tubal and male factors being most common. Tubal factor is more common in areas of high incidence of sexually transmitted infections and is strongly correlated with intraabdominal infections leading to adhesion formation. Over the last few decades, semen parameters have been deteriorating possibly contributing to the rise of male factor infertility[5]. In approximately a quarter of infertility cases, the cause cannot be established using conventional diagnostic methods[6]. The main cause for the increase in infertility rates over the last few decades however is postponement of childbearing due to various socio-economic and personal reasons[7]. Based on The Organization for Economic Co-operation and Developments (OECD) statistics, the mean maternal age of first child birth has increased by 2 to 5 years from 1970 to 2017, with majority of women now having a first child born at the age of 30 or more[8]. When the presumed cause of subfertility is established, or after a certain period of unexplained infertility, the only treatment of choice is ART in the form of *in vitro* fertilization (IVF) or intracytoplasmic sperm injection (ICSI). Failure of less invasive treatment options such as intrauterine insemination (IUI) or ovulation induction (OI), would also lead to the above therapies. Though ART has come a long way since the first IVF baby was born in 1978, the overall success rates per embryo transfer are in the region of 35% [9]. This is marginally better than nature, as the estimated monthly fecundity rate is 20% per cycle[10].

Throughout the last decades we have witnessed massive technological changes and innovations that have influenced the field of infertility, in particular the ways we perform IVF, handling and evaluating the oocytes, the sperm and the embryos. Examples of such advancements include the introduction of ICSI technology for male factor infertility in 1991[11] and the introduction of pre-implantation genetic testing or screening (PGT/S) of embryos in 1990's[12]. Machine learning (ML), artificial intelligence (AI), and other modern statistical data mining methodologies are now providing new opportunities in precision medicine to improve infertility care. The term of '*P4 medicine*' subsumes the preventative, personalised, participatory and predictive aspects of medicine, and was initially promulgated 17 years ago by the systems biologist Lee Hood[13]. This concept has mainly been applied to cancer and chronic diseases such as cardiovascular diseases which represent the major causes of death globally. An increasing availability of advanced technologies allowing for screening of the genome, transcriptome, microbiome and metabolome should find its way into infertility investigations and treatment. The utilization of big data and ML to screen the results of molecular phenotyping and other clinical analyses, as well as population-based data and registries will allow us to continue developing insights into additional causes of infertility and establish improved and personalised treatments. Understanding why infertility occurs may reveal some unknown factors that could be used to make diagnoses and treatments better, faster, less expensive, and as such, grant infertile couples, the family they so desire.

In this paper we refer to this notion as a '*systems medicine approach*' to infertility (Figure 1). This approach can be defined as the use of advanced computer algorithms such as ML/AI and big data mining techniques to integrate the diverse factors (genetic, lifestyle, environmental) implicated in this disease, with the aim to develop better tools for infertility prevention, prediction, diagnostics and treatment.
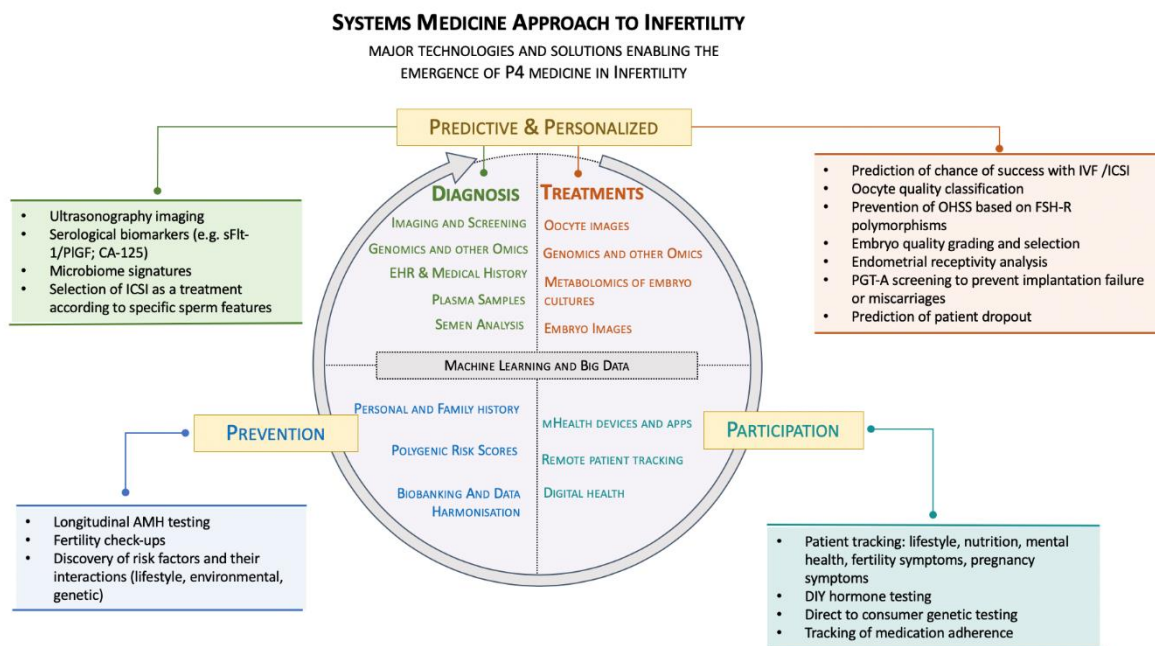
Figure 1: The systems approach to infertility. The figure shows the major technologies and solutions enabling the emergence of a P4 system in infertility. 'Prevention' refers to the ability to prevent infertility disorders or pregnancy complications. 'Predictive and Personalized' refer to the prediction and management of disease with greater granularity and the individualization of treatments. 'Participation' refers to patient participation in their own health. The text inside the circle indicates key data sources generated by each segment that can be integrated using ML and Big data mining techniques. Text boxes outside the circles include practical examples of the use of these data sources. At the centre of the circle we highlight 'Machine Learning and Big Data', which are essential to every of the four sections to power the innovation cycle of a 'systems medicine approach' to Infertility.

We aim to describe key developments, promises and limitations to this approach, and identify possible avenues to its implementation with detailed discussion of four innovation forces that are driving the emergence of a system approach to the infertility sector:

1. **A conceptual shift from a disease-centric to a health-centric model**: Infertility healthcare is moving beyond disease-based reactive model of healthcare to a pro-active model where consumers are actively engaged in the preservation and enhancement of their health and well-being. A shift in nomenclature should accompany this approach with '*infertility healthcare*' being replaced with '*fertility healthcare*'.

2. **Better prevention, diagnosis and treatments**: Sophisticated big data analysis of cohorts have allowed the development of predictive and prognostic biomarkers and strategies for early interventions based on molecular profiling rather than categories of symptoms. Infertility is being managed and treated in a more cost-effective manner.

3. **ML/AI in ART Treatments:** Perhaps the biggest breakthrough in the use of ML/AI in infertility is its application to IVF treatments across all stages: from IVF protocol design, through oocyte quality classification and sperm analysis, to embryo grading and the selection of the optimal window of implantation.

4. **The participatory citizen:** Participatory health[14] is a term that indicates the individual's empowerment in the realization of their own health. The individual is an instigator and driver of its own reproductive health and well-being.

1. A conceptual shift from a disease-centric to a health-centric model

The chance of natural conception for a couple with normal fertility, as a result of intercourse without contraception during the fertile phase, is about 20-25% for any given month[15]. Perhaps due to an emphasis on preventing unwanted pregnancies and sexually transmitted diseases, currently most people are well informed about contraceptive precautions but are generally unaware of or unconcerned with their fertility health until they start trying to conceive. At present, the approach to infertility is to treat, rather than prevent with very little effort put into the latter. Tubal factor being one of the leading causes of infertility is caused most commonly by sexually transmitted infections. Prevention of these (use of condoms) and or early treatment reduces the insult leading to tubal damage, which may cause tubal factor infertility in 15% of cases following a single infection[16,17]. Another example of a simple intervention to improve chances of pregnancy is a modest decrease of even 5% of total body weight in cases of obesity associated with anovulatory polycystic ovarian syndrome (PCOS), which has shown to restore ovulation in a large proportion of women, potentially sparing them the need for ART[18,19]. Unfortunately, there are very few other examples of infertility prevention and in order to minimize the spread of the pandemic, this attitude needs to change. Some signs of this can be seen in the introduction of fertility check-ups by IVF clinics, where a combination of blood tests and ultrasound scans is used to determine the infertility/ fertility potential of the woman or couple. In this context, markers of ovarian reserve (anti-Mullerian hormone - AMH and antral follicle count - AFC) are used to determine the expected duration of the woman's or couples' fertile years in order to guide them as to when to try and conceive in order to achieve the desired family size. When the ovarian reserve is found to be poor, fertility preservation may be advised. This can be done either in the form of oocyte or embryo freezing, with outcomes of both methods being very promising. In 2016 in the UK alone 1,310 oocyte freezing cycles were carried out representing a 17% increase from 2015 and doubling of numbers since 2013. In 74% of cases, these were privately funded, indicating that the likely reason for oocyte freezing were social circumstances[20]. However, it is often older women who seek this way of preventing future problems while it is in younger women that the success rate would be greatest. The 2016 data from the HFEA indicated that most oocyte freezing took place at the age of 38, which is past the optimal age for this procedure to confer the greatest chance of pregnancy (below the age of 35 would be optimal)[20]. Access to accurate information about reproductive health can empower patients and practitioners with the necessary knowledge, thereby motivating behavioral changes and a shift from a disease-based model to a health-model in fertility. In the UK, a need to overcome legal limitations of the 10-year storage limit on gametes must be addressed for women to be able to store their eggs without the pressure of using them with the stated timeframe, a consideration mostly for women in their early or mid 20's. This matter is currently under public debate and will be addressed in the UK parliament soon. A limitation of a preventative aspect of infertility is also related to the fact that it is assisted and as such, artificial. This may form barriers in uptake due to religious or personal beliefs and a perceived stigma of infertility. Irrespective of how oocytes or embryos are cryopreserved for future use, there is no guarantee that the person will be able to have a biological child. It has been estimated that when cryopreserving oocytes, a woman at the age 34 would need to store 20 mature oocytes in order to have a 90% likelihood of having at least one child. For women of 42, this likelihood would be only 37%[21]. While it is possible for a 34-year-old woman to achieve 20 mature oocytes during one stimulation, it may be risky and often will require 2 or more fresh COH cycles. In women of 42 years-old, multiple cycles will be required to achieve this number. This unfortunately forms a significant financial barrier for most women as the average cost of COH, freezing and storage of oocytes in the UK is estimated to be £4,600 per cycle[22].

As the value of preventive, predictive and personalized medicine is increasingly recognized, there exists a great opportunity to encourage this paradigm change in infertility, i.e. away from a disease-centred vision to a preventative health-centred vision. In this perspective, the role of the patient is

also shifting, from being a passive minimally informed recipient of healthcare to an active and engaged participant of their own fertility health and wellbeing. The legal and ethical changes are also required in parallel to the social aspects in order to assure a smooth uptake. This essential participatory aspect of the patient is discussed below.

2. The era of big data is enabling better prevention, diagnosis and treatments

Infertility can be attributed to men or women in equal proportions (20%-30%), or to both partners (20%) with the remainder being unexplained. Female infertility is mainly associated with tubal damage, endometriosis, ovulatory dysfunction, or rarely uterine or cervical abnormalities. Broadly termed, male infertility based on origin could be divided into hormonal, testicular and post-testicular affecting the quantity as well as quality of sperm[23]. A fundamental problem with developing personalised therapies or diagnostic tests for infertility is the limited understanding of the aetiology and pathophysiology of the disease. In fact, in about 10%-25% of the cases, the cause cannot be identified leading to the so called 'unexplained infertility' diagnosis. Refined diagnosis with clear understanding of the mechanisms and underlying symptoms can lead to specific and effective clinical treatments, higher safety for patients, and less costs. A good example is the severe male factor infertility due to presence of varicocele. Surgical or radiological varicocelectomy can lead to improvement in semen parameters and an increase in chances of natural conception, or reduction in the need for surgical sperm retrieval[24]. Treatment of varicocele has also been associated with a reduction in the risk of pregnancy loss in couples with recurrent miscarriages (13.3% versus 69.2%, P=0.001)[25]. If left untreated, varicocele may lead to clinical hypogonadism with associated negative effects of testosterone deficiency, poor sperm parameters or azoospermia[26]. Though varicocele is considered by some as a reversible cause of male infertility, often these men are guided towards ART ignoring the treatment that could restore normal testicular function and allow for natural onception[27].

The use of advanced next generation sequencing (NGS) techniques - such as genomics, transcriptomics, proteomics and metabolomics - allows the generation of large volumes of biological and clinically useful data. Figure 2 presents examples of the different types of data that are starting to be incorporated into an overall precision medicine picture in infertility. These different data-streams are used by ML and systems biology to model complex systems, understand the underlying disease mechanisms and identify novel predictive and prognostic biomarkers (Figure 2). Several examples already exist in different areas of the reproductive sector, such as endometrial receptivity analysis[28], male sperm selection[29], preeclampsia[30], and preterm birth[31].
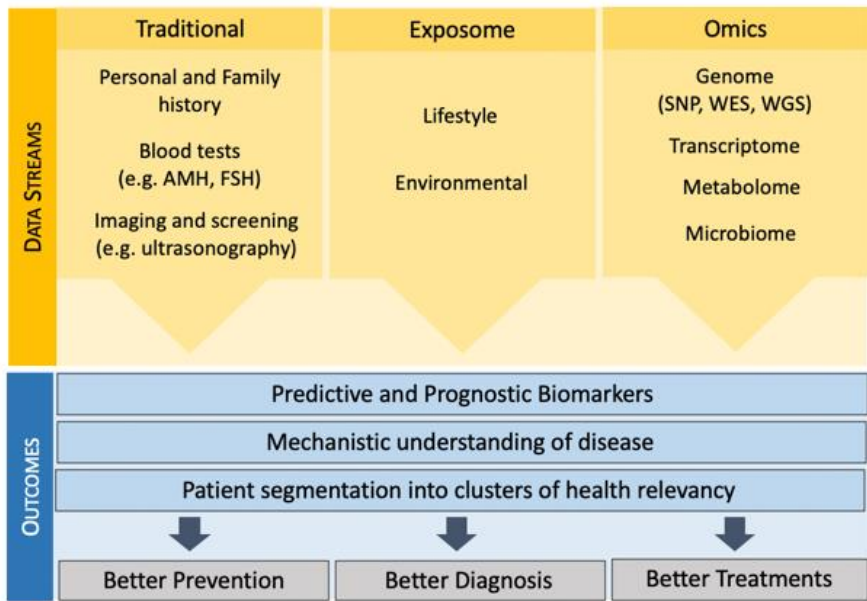
Figure 2: Data streams used to enable precision medicine in the Infertility sector. The newer health data streams such as genomics, the microbiome and the exposome are a complement to traditional approaches used in infertility investigations and management, to ultimately deliver better prevention, diagnosis and treatments.

Biological markers (biomarkers), originally defined by Hulka and Wilcosky in 1988[32], are essential tools and technologies in precision medicine. By definition, biomarkers are biological entities that can be measured objectively and can be used in the prediction, diagnosis and progression of a pathological process, and used as indicators of responses to therapeutic interventions. One of the best examples of the use of biomarkers in reproductive health is the discovery of the role of angiogenesis-related factors, soluble fms-like tyrosine kinase-1 (sFlt-1) and placental growth factor (PlGF) in preeclampsia. sFlt-1/PlGF ratio through ML has been demonstrated as a valid biomarker for identification of women at high risk of preeclampsia and intrauterine growth restriction[33]. The increased production of autoantibodies such as anti-phospholipid (APL) and/or anti-nuclear antibodies (ANA), have shown to be involved in infertility disorders including premature ovarian insufficiency (POI), unexplained infertility, as well as unsuccessful IVF treatments, preeclampsia, and spontaneous abortions[34,35]. For couples with recurrent miscarriages, detection of serum APL and subsequent treatment with combination of heparin and aspirin can lead to a reduction in pregnancy loss by 54%[36]. Biomarker discovery can be based on other types of information, such as non-coding RNAs, where some authors propose the use of non-coding RNAs and microRNAs as biomarkers predictive of male infertility[37,38], placental function during pregnancy[39] and of circulating small non-coding RNAs in the first trimester[40] as a non-invasive diagnostic tool for preeclampsia.

Access to a diverse set of health-related data, will not only be important for biomarker identification in isolation, but when applied together in an integrated analysis can aid in the improvement of predictive algorithms. For instance, the serum Cancer antigen 125 (CA-125) is the most extensively used peripheral biomarker in deep infiltrating endometriosis (DIE) for detection of the disease and evaluation of therapy[41]. While studies that evaluate the performance of CA-125 have presented different limitations for the diagnosis of early stages of the disease, mainly in relation to their sensitivity[42-44], recent data mining approaches modelling CA-125 levels in conjunction with other clinical features, such as transvaginal ultrasound, cyst or fallopian tube pathology, BMI and dyspareunia have been shown to improve the sensitivity of this biomarker[45,46].

Possibly one of the most promising interventions where ML was used to develop a screening test to improve IVF success was the Endometrial Receptivity Analysis (ERA). The test assesses the expression of 238 genes that have been demonstrated as potential transcriptomic predictors of endometrial

receptivity and allows for personalisation of embryo transfer to allow optimal synchrony between the embryo and endometrium[47,48]. The test still however requires validation in a randomized controlled trial versus conventional embryo transfer strategy in order to prove its clinical value. It is therefore important that the effect of any potential biomarker is replicated in a large independent cohort, with well-selected control populations before being adopted in the clinic.

Disease classification has always been challenging, especially in the context of biological heterogeneity where the same symptom might be generated by different disease mechanisms. It should be possible to expand beyond broad distinctions of infertility symptoms, such as pelvic pain, dysmenorrhoea, or metrorrhagia, and include molecular profiling and biomarker testing to classify diseases and disease subgroups. Genetic information, in the form of single nucleotide polymorphisms (SNP), whole exome sequencing (WES) or whole-genome sequencing (WGS) in combination with other omics data are crucial in this context (Figure 2). One of the best examples of the use of genomic profiling in personalised medicine is Herceptin targeted therapy used to treat HER2 positive breast cancer[49]. In infertility sector, endometriosis is a good example of the increased recognition of varied clinically informative molecular phenotypes which can be used to segment patients into clusters of health relevancy (as reviewed in [50]). Initiatives such as the World Endometriosis Research Foundation (WERF) Endometriosis Phenome and Biobanking Harmonisation Project (EPHect)[51] are paving the way to advance biomarker and targeted treatment discovery in endometriosis.

Several recent reviews have described the complexity of female and male fertility[52-54] revealing an emerging picture of high levels of genetic heterogeneity in certain diseases such as PCOS which can include variants in or near the Luteinizing hormone (LH), the FSH receptor (FSHR) genes or variants in the FSH-β gene[55]. Better understanding of disease mechanisms and related genes can guide the selection of drugs or treatment protocols, minimize harmful side effects, or ensure more successful outcomes[56]. For example, variants in the STK11 gene were associated with a decreased chance of ovulation in PCOS women treated with metformin[57]. Other examples include the identification of M2 carrier pregnancies by screening both partners for the M2 haplotype, which can be used to stratify couples for treatment with low-molecular-weight heparin (LMWH)[58].

Genomic data is not just useful for treating and diagnosing diseases, it's also important for disease prevention. Other data-streams such as the exposome could be integrated alongside genomic information to help elucidating the baseline status and response of individuals in specific contexts. The exposome is a measure of the impact of exposure (e.g. diet, lifestyle) or an individual's experience over their lives on their health[59]. In infertility, several lifestyle factors have been shown to be predictive or prognostic of disease such as the consumption of coffee (WHO recommends < 3 cups of coffee/day to minimise risk of miscarriage)[60], alcohol intake (moderate alcohol intake does not seem to be associated with infertility) [61], and even air pollution[62]. Actions in this direction are visible in cancer research, for example The Women Informed to Screen Depending on Measures of Risk (WISDOM) study evaluates the efficacy of risk prediction based on screening, clinical risk factors (e.g. BMI, age), breast density, and polygenic risk scores, to inform breast cancer screening[63,64]. It is clear that the complexity of interactions between genetic and non-genetic factors in infertility should be modelled integratively. This could be detrimental for developing preventive strategies such as improving diet structure or considering the use of fertility preservation techniques such as oocyte cryopreservation earlier for certain patients at higher risk of infertility.

The microbiome is another important emerging health data stream in infertility. The female reproductive tract microbiota accounts for 9% of all bacterial load in humans with *Lactobacilli* being the dominant genus in a healthy woman[65]. Bacteria have been isolated from every part of the female reproductive tract including even the peritoneal fluid of the pouch of Douglas, where the concentration of bacteria was 10 000 times lower compared to vaginal fluid[66]. Different bacteria have

also been isolated from seminal fluid of healthy and infertile men, highlighting that *Lactobacillus* also plays a protective role for sperm health[67,68]. NGS has allowed to characterise in detail the microbiota of the female reproductive tract and has shown that the percentage composition of *Lactobacillus* in the endometrium differs between healthy volunteers (85.7%), non-IVF patients (73.9%), and IVF patients (38%) indicating that majority of infertile patients demonstrate an abnormal endometrial bacterial profile[69,70]. Domination of *Lactobacillus* in the endometrial microbiota (96.5% ± 33.6%) in the same infertile population was associated with pregnancy, suggesting that *Lactobacillus* dominated bacterial flora might favour implantation[69]. Isolation of pathogenic bacteria or diagnosis of chronic endometritis, a condition associated with recurrent implantation failure, may improve the chance of healthy term pregnancy with a course of antibiotics before embryo transfer[71,72]. The use of probiotics containing the *Lactobacillus* genera could be an alternative in order to avoid the necessity for antibiotic treatment and its side-effects, however evidence to support this intervention is lacking. With mounting evidence for the importance of reproductive tract microbiota for reproductive health, more pressure should be put into assessment of this before treatment, especially in the population of couples with recurrent implantation failures or recurrent miscarriages. Data obtained during such tests, could add to the aspect of personalized medicine and with help of ML/AI models, an exact microbiome-print could be identified which assists conception and continuation of pregnancy to term.

3. Machine Learning is aiding ART Treatments

Since Louise Brown, the first IVF baby, was born in 1978, it is now estimated that over eight million babies worldwide have been conceived through assisted reproduction. The overall success rates of achieving a pregnancy in couples undergoing ART are marginally better than chances of natural conception. Whilst we are proficient in obtaining oocytes and spermatozoa and creating embryos, replacing them in the uterus, we have almost no influence and limited knowledge of events occurring thereafter. When research funding in Obstetrics and Gynaecology is concerned, fertility research is perceived as an ugly duckling compared with funding for gynaecological oncology. In 2018/19 Cancer Research UK, spend £546 M on research[73], whereas NIHR in the preceding year has spent only £21M in research grants across reproductive health and childbirth combined[74]. This disparity is likely originating in the fact that IVF centres are perceived as commercial entities with research not being a priority. This low-level funding for research in the field of ART is concerning when combined with the fact that in the last decade, a decline in worldwide IVF birth rates can be observed. This state is likely to be attributed to mild stimulation protocols, elective single embryo transfer (eSET), pre-implantation genetic testing for aneuploidy (PGT-A), freeze-all cycles, embryo banking, increased industrialization, and unchecked introduction of IVF add-ons, with very little good quality research being conducted to assess the actual benefit of any of these interventions on pregnancy outcomes[75].

Recent developments looking at the endometrial receptivity[48,76] in order to personalize embryo transfer and align it with the individual window of implantation have shed some light on the workings of the endometrium. Non-invasive assessment of the embryo culture fluid allows with some degree of accuracy to determine the ploidy status of the embryo without resorting to invasive biopsy[77]. Intracytoplasmic morphologically selected sperm injection (IMSI) utilizes high magnification microscopy to identify the most suitable sperm to be injected into the oocyte in the hope of maximising the chance of pregnancy. Time-lapse imaging allows for monitoring of the development of the embryos and as such, selection of the most promising one, without relying only on the subjective assessment of morphology by the embryologist. All these advances are carried out in order to improve the pregnancy rates; however, they seem to be carried out without guidance and selection. If we treat each of these interventions as a data point, ML/AI would help to select which of these treatments, if any, would benefit the couple seeking fertility without the need to include all of the expensive, novel and unproven add-ons (Figure 3).

Assessment of the couple should be the starting point with detailed history and examination as the foundation of treatment. Whilst the standard approach to ovulation induction or controlled ovarian hyperstimulation (COH) will work for majority of the women, a small proportion will require more or less of stimulating medications to achieve the same outcome. This is related to personal sensitivity to the given medication and is related to the metabolism or receptor variants. Clomiphene citrate (clomid) has been used as a first-line treatment for PCOS. Clomiphene resistance may be linked with cytochrome CYP2D6 variants, as the drug is mainly metabolized into its active components by the enzyme[78]. Small studies have however produced conflicting results[79,80]. Larger datasets may be a step towards personalized medicine in the field of ovulation induction for women with PCOS and in order to decrease the risk of cycle cancellation. In some cases, very high doses of gonadotrophins need to be used in order to achieve follicle growth despite objective good ovarian reserve. A small proportion of women undergoing COH with gonadotrophins may under respond or not respond at all to these medications. Mutations in the FSH receptor (FSH-R) gene have been found to contribute in this circumstance. Inactivating receptor mutations are known to occur in the FSH- R gene[81,82], but there are other more subtle alterations to the gene that modulate its activity. Single nucleotide polymorphism (SNP) (G/A) in the 5′ untranslated region (UTR) of the FSH receptor gene at the −29 position is associated with altered transcriptional activity of the receptor gene[83], with AA genotype being associated with poor response. In a small study of 50 women undergoing ART, one of the 7 women with AA genotype conceived (14%) versus 31% of the GG type and 59% of the GA type. AA genotype women seem to require higher doses of gonadotrophins with lower number of antral follicles and lower number of retrieved oocytes compared to the GG and GA genotypes[84]. When another SNP Asn680Ser was analysed, homozygous women for Ser680 achieved lower levels of oestradiol during COH[85], an effect that could be overcome by higher doses of gonadothrophins[86], with other studies refuting this result with the effect of the population studied as a possible explanation[87-89]. Homozygosity for A in the Thr307Ala SNP in another small study has been associated with Ovarian hyperstimulation syndrome (OHSS) (6 of 7 patients (86%), versus 3 of 12 subjects with homozygous T allele (25%) and 6 of 31 (20%) subjects with heterozygous TA allele)[89]. It would therefore be advantageous and immensely helpful to know these facts before starting treatment in order to improve the outcome and reduce frustration due to a trial and error approach. Blood tests with analysis of specific parts of the genome, would be able to identify the small proportion of women with these variants and lead to improved treatment results minimising time to pregnancy and associated cost. In order to make this mainstream, larger studies on a more diverse population are required, alternatively AI could be employed to assess the genetic diversity that may have eluded normal analysis. This approach is however not undertaken in reproductive medicine at present. Baseline characteristics entered into an AI algorithm would allow for appropriate counselling of the couple regarding their chance of success per treatment cycle started, if they require any additional treatments in order to improve their fertility potential, guide how many cycles they may require to achieve the family size they desire, and allow for financial planning around his treatment. In cases where the chance of success is very low (women with very poor ovarian reserve or significant comorbidities), with the help of these algorithms, the clinician could accurately counsel the couple regarding the best course of action, which in these cases may be oocyte donation, surrogacy or adoption. Time to pregnancy in such cases would be shortened and the significant financial impact on the couple mitigated. The same process could be applied to the selection of the most appropriate stimulation regime for the couple and if IVF or ICSI would provide the best chance of fertilization, especially in the unexplained infertility situation. There are many sources of information, however they all consider a very specific timepoint for analysis and risk/chance calculation. An overarching approach is not possible other than the clinicians' experience. ML/AI could incorporate all the minute details that are often overlooked and provide superior guidance for the fertility specialist.

Analysis of ultrasound images of the endometrium in combination with any molecular receptivity data in a combined algorithm, could provide more accurate assessment of endometrial receptivity than either of these techniques used alone. Such an approach is not yet being utilized in the fertility clinic.

Following oocyte collection, computer vision and AI that uses advanced machine learning algorithms to understand static images of day 5 blastocyst is already being utilized to predict the embryo viability and chance of pregnancy with a 24.7% improvement in accuracy over a trained embryologist [90]. This model is used commercially by fertility clinics worldwide. Static image analysis and AI algorithms have similarly been employed in assessing the risk of embryo aneuploidy and have reached concordance level with the biopsy result of 81.5% [91].

Once the pregnancy has started, there is very little that can be offered to the couple in terms of estimation of a successful outcome. Depending on the healthcare system, a pregnant woman contacts her gynaecologist or general practitioner and either has a blood test or a scan. Rough estimation of doubling of hCG levels every 48 hours has been used for a long time as a predictor of the risk of a viable intrauterine pregnancy versus an ectopic pregnancy[92]. This approach is mainly used in case of early pregnancy complications, such as pain and or bleeding and rarely employed in low risk pregnancies, other than maybe IVF pregnancies. Recently, ML has shown some promise however in providing some useful guidance on follow up frequency for patient thought to be at high risk of miscarriage after IVF[93]. Ultrasound imaging is the mainstay of pregnancy assessment; however, its accuracy is largely dependent on the operator. The advent of 3D ultrasound and the ability to store a volume of the area scanned allows for storage of a dynamic image of an entire organ and offline or even off-site analysis by different operators. It would therefore be possible to employ AI to combine biochemical markers of pregnancy with 2D or 3D images of the early pregnancy in order to provide the couple with risk of miscarriage, preterm delivery and ultimately, live term pregnancy. Utilization of algorithms in early pregnancy has been used for trisomy screening, where nuchal fold thickness in combination with free human chorionic gonadotrophin beta (beta-hCG) and pregnancy associated plasma protein A (PAPP-A) are used to calculate a risk of trisomy 21[94]. This approach is likely to be superseded by NIPT- non-invasive prenatal testing of free fetal DNA from maternal blood, which can be done as early as 7 weeks gestation It is far more reliable than the combined test for diagnosis of trisomy 21, 18 and 13, but at a higher overall cost[95]. Soft ultrasound markers of chromosomal anomalies are utilized by the clinicians as means to carry out more definitive tests, such as chorionic villous sampling or amniocentesis, both of which are associated with a 1-2% risk of miscarriage[96]. This may be unacceptable to some couples and they would welcome a non-invasive assessment combining what is already known about the pregnancy and providing them with a possible outcome. AI could incorporate all this particulate knowledge into one chance of healthy term pregnancy algorithm. This would have to be a dynamic process as the pregnancy develops and more items can be input. Having also access to this data by the patient with the ability to input new data, would make it a valuable tool for the couple to guide them as to the need for medical attention. Mobile applications, such as the QUIPP app[97] utilize the clinical data regarding patients' previous pregnancies, their duration, any cervical surgery, cervical length measurement and levels of fetal fibronectin in the vaginal secretions to predict the risk of preterm delivery with an area under the curve of 0.763 for delivery < 34 weeks and 0.746 for delivery < 37 weeks when the cervical length and fetal fibronectin were entered between 22+0 and 25+6 weeks gestation[98].
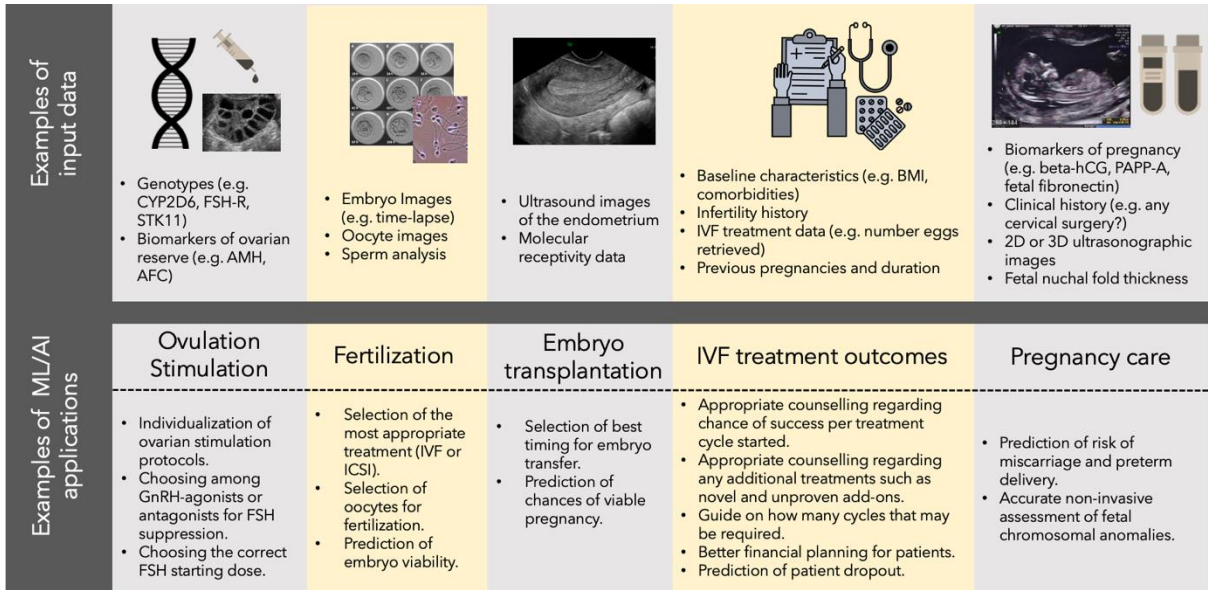
Figure 3: Examples of different types of input data and ML models that can be built and applied to improving ART treatments.

4. The participatory citizen: From a disease-centric model to active wellness

The term '*Participatory medicine*' was famously defined by the patient activist Giles Frydman in 2010 as '*a movement in which networked patients shift from being mere passengers to responsible drivers of their health, and in which providers encourage and value them as full partners*' [99] . The term is one of many used interchangeably, including 'patient-centric medicine', Health 2.0, Medicine 2.0, and eHealth. Among the various causes of infertility, some are related to factors that can be modulated by individuals themselves such as an unhealthy lifestyle and diet. In addition, lack of awareness and '*fertility education*' may hinder the search for treatments. The basic idea is that an active and engaged individual is able to become not only an active player in influencing their own health and well-being, but also a good advocate of preventative and early detection strategies. Participatory health efforts include social media, mobile health apps (often referred to mHealth), self-tracking records, social health networks and direct to consumer (D2C) tests like genomics and blood tests (Figure 1).

In recent years, we have seen an explosion of apps and wearables dedicated to optimizing the chance of pregnancy and reproductive health. The digital market and support resources for women seeking fertility services is skyrocketing. Period-tracking apps, which aim to help cycle monitoring and pinpointing ovulation, like Glow and Clue, have millions of users. Self-tracking apps are winners not only for female reproduction but for anything trackable, like steps, calories, diet, sleep patterns, mood, and physiological parameters such blood pressure, heart rate, and pulse rate. A large proportion of female digital healthcare companies focus on making health-related tests available directly to consumers. These can range from at-home blood tests for AMH and other hormones (e.g. Parla and Modern Family), to turning a smartphone into a powerful microscope to allow users to test their semen at home (e.g. YO test and ExSeed health). In 2017, Celmatix Inc, a USA based company, announced the first commercially available product that provides D2C genetic testing, the Fertilome. The Celmatix test currently examines 49 SNPs in 32 genes, which have been implicated in a variety of reproductive conditions, such as POI, recurrent pregnancy loss (RPL) and PCOS. This genomic test was one of the first ones to offer significant amounts of genomic data related to reproductive health directly to individuals without the mediation of medical professionals. Insights from this test could help identify infertility (such as diminished ovarian reserve) and pregnancy-related complications earlier. However, this level of access to genomic information has been criticized[100,101] as it may cause unnecessary anxiety for patients and clinicians, especially when it comes to explaining the implications of problematic results that might not be relevant to the patient. For example, in some cases there

might be inadequate evidence (e.g., lack of appropriate controls, underpowered studies, ethnic specificity) to accurately interpret the pathogenicity of a certain genetic polymorphism. In addition, we still have a very limited understanding of genetic risk. Probabilistic risk information for any health condition remains difficult to be actionable in the clinic, such as advising patients to make behavioural changes or guide clinicians into choosing risk-reducing interventions.

The COVID-19 pandemic has dramatically changed the story of telehealth. Patients and clinicians are now more receptive to virtual care. We are living in unprecedent times when it comes to embracing D2C and "at-home" testing, and thus we can trust that these technologies will continue gaining momentum in the future. eHealth and D2C tests may prove extremely helpful for enabling physicians to reach more patients and giving individuals more control of their reproductive health. The potential for widespread availability of D2C testing is big, however, the full capabilities of such tests and inevitable shortcomings must be properly validated in order to provide reliable counselling. Additionally, due to the incomplete understanding of the processes related to human fertility, results of such tests may be irrelevant until the woman actually tries to conceive. The implications of such tests may vary, for example finding out you are a carrier of a FSH-R mutation reducing its sensitivity to COH, which may or may not be needed in the future, is of lesser significance than finding out you are a carrier of the BRACA-1 gene mutation giving the person affected an up to 87% lifetime risk of developing breast and up to 63% risk of developing ovarian cancer[102]. Applications of such D2C tests require appropriate counselling of the patients and clinicians to interpret them correctly and stratify their significance in the context of the individual. In the future, ML/AI algorithms could aid in the process of assigning significance to these results and making them a valuable tool in e-medicine.

5. Limitations of the use Machine learning and Big data in the Infertility sector

Despite its numerous high-value benefits, ML and AI are not without challenges (Figure 4). Key challenges include (1) difficulties in obtaining the accurate and high-quality data, (2) the challenges of generalisation and unintended algorithmic and dataset biases, (3) validation of ML/AI algorithms and (4) translation into clinical practice. Developers of AI models must be cognizant of these limitations to capture the maximum amount of value from the predictive models.
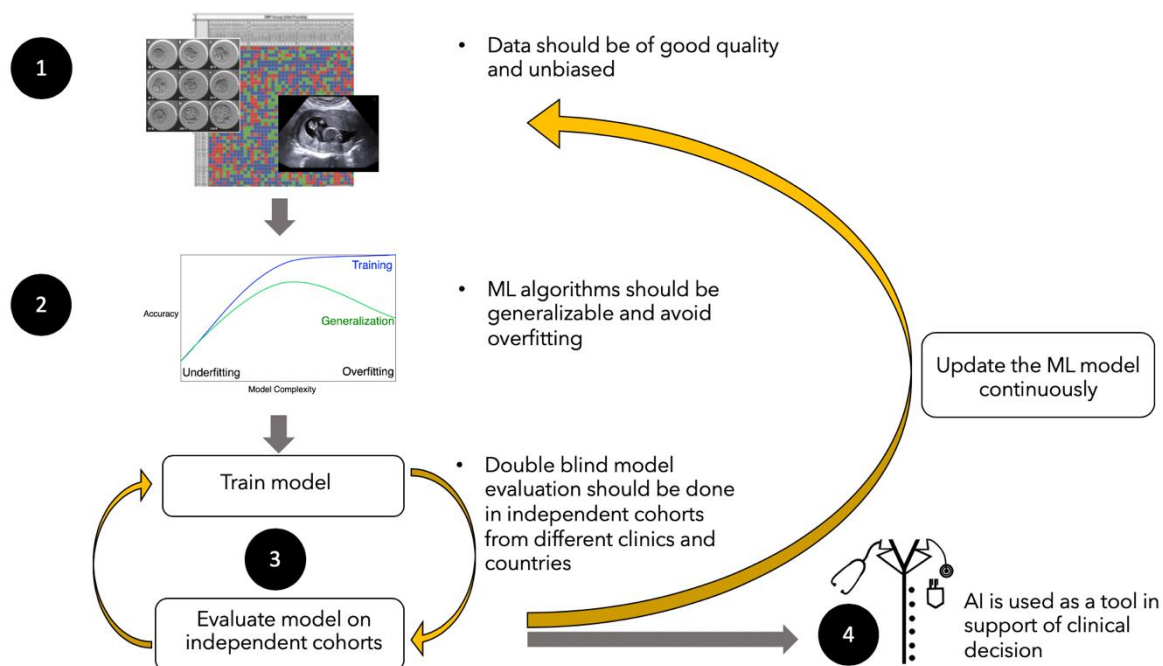
Figure 4: Challenges associated to the use of ML/AI algorithms. (1) Data should be labelled accurately and unbiased; (2) ML/AI algorithms should be generalisable and avoid overfitting; (3) Algorithms should be validated in independent cohorts and continuously updated; (4) Algorithms need to be translated and adopted into clinical practice.

5.1 High quality and quantities of data

The first step to implementing an ML/AI algorithm is access to Big data, as many clinics have not yet implemented rich electronic health record (EHR) systems. The UK Human Fertilisation and Embryology Authority (HFEA) prospectively collects baseline information and birth outcomes on all licensed fertility treatment cycles performed in the UK since 1991. It is one of the biggest data collections publicly available in the world, comprising more than 1.3 million treatment cycles up to 2016. However, many of the important predictors such as BMI, detailed clinical history, and lifestyle parameters, and many details regarding the sperm, oocyte and embryo quality assessments, are not robustly collected in this dataset, making its use limited for modern ML applications. In order for a ML model to be consistently successful at making predictions, data must be of high quality and of sufficient sample size. If data is not labelled or categorized appropriately, or if it is not in sufficient amounts to represent an unbiased sample of the general population, then the model produced will be unduly influenced by noise in the data that is used. The number of records required to reach sufficiently high model performance depends on the model itself, and the number of predictor variables required. The rule of thumb for a regression model is to have 20 observations for each parameter[103], however this number will vary if the task at hand is more complex, such as an image classification task using deep learning. Although deep learning requires larger amounts of data than traditional ML, large data does not always impact the performance. Using well understood, strong predictor variables can decisively affect model performance. A recent study in type 1 and type 2 diabetes has demonstrated that traditional logistic regression models can perform as well as complex machine learning algorithms like Neural networks and gradient boosting machines, provided that the data contains a set of strong predictor variables[104]. Similarly to this example, models of oocyte quality assessment could benefit from combining a typical '*black-box*' approach (in which we let the computer identify the general features important for a classification), with a more targeted approach where additional predictor variables are chosen based on domain-knowledge. These added features should be related to well-known factors that affect oocyte quality such as patient age, BMI and smoking status. In summary, the data used for training a ML algorithm should not only be of high-quality, but also contain the necessary predictor variables and be representative of the problem domain. Finally

and equally important, data should be compliant with the 'FAIR' data principles; meaning that it is Findable, Accessible, Interoperable, and Reusable[105]. The public availability of patient data is an issue that has been endlessly discussed. While it is important to restrict access to sensitive patient information, we need to improve and facilitate data access requests in the reproductive health field (in many cases access request is made by emailing the corresponding author of a publication). Making research datasets easily available to the wider community will no doubly help accelerating research progress.

5.2 Generalizability of learning

A generalizable model is a model that is not only representative of the population used for training but also transposable to other populations. For instance, It is well recognised that infertility investigations and treatment practices tend to vary across clinics (due to differences in population demographics or socio-economic statuses, differences in treatment regimens, measuring devices, laboratory protocols, or compliance with established guidelines), countries (e.g. due to differences in population or IVF regulations) and temporal periods (e.g. due to advances in technology). Thus, in order to build generalizable models, it is important to collect data across different clinical settings, populations, and subgroups of interest. If the data is small and based on local clinics or centres it may not represent the real word setting because there might be clinic- or demographics-specific nuances in certain clinics. In fact, the treatment centre is a well-known very strong predictor of the success of IVF[106] as many indications for IVF treatment are defined qualitatively, and their use might have varied among physicians of different clinics. On the other hand, if the sample size is large, the data should have a wide variety of records, including unique or odd cases, to produce a generalizable model. If not, models derived using multi-centre national databases could have poorer quality when compared to single-centre collected and physician-curated data. Hence, it is important to check that the population and data collection are in accordance with, and representative of, the envisioned application scenario, and therefore meaningful to patients in the context of their clinics.

5.2.1. Data biases introduced by population heterogeneity

There are several examples of heterogeneous model performance across different settings and populations. This is often defined as a '*spectrum effect*' or '*spectrum bias*', a term used to describe the situation where test performance varies across patient populations[107]. Failure to recognize and address heterogeneity will lead to models that are not generalizable[107,108].

One key source of heterogeneity is the average prevalence of treatment types performed in different countries, or even in different clinics of the same country. For instance, the usage rates of ICSI vs standard IVF vary from region to region, with a 55% usage in Asia, 65% in Europe, 73% in North America, and 86% in South America. The highest proportion of ICSI is utilized in the Middle East with almost all cycles using ICIS to fertilize oocytes [9,109]. Many other clinical and laboratory differences exist such as types of COH regimes, protocols for oocyte retrieval, including sedation equipment setup; differences in treatment strategies, clinical guidelines, methodological standards and experience; differences in disease definitions, etc. All these problems may lead to heterogeneity in the magnitude of the predictor effects[110], in the prevalence of the measured outcome, and in the distribution of predictor values themselves[111].

5.2.2. Non-stationarity in treatment data and historical biases

Historical biases can be engrained at any category of infertility-related data, but it is particularly relevant when developing models around IVF treatment data. As the regulatory environment around IVF practices continues to change and treatment practices continue to evolve, great care must be

taken when choosing which data to use and which to exclude. For instance, elective single-embryo transfer (eSET) is now a common practice across clinics in UK since the introduction of a policy to encourage fertility centres to eSET in 2009[112,113], the number of same-sex couples seeking IVF has increased substantially after 2008 when the legal requirement for a father figure was removed (between 2014 and 2013, IVF increased by 20.1% for same-sex female couples, and by 2.7% when considering all couples)[114], egg freezing has increased 10% from 2016 to 2017, egg thawing was used in 581 cycles in 2017 (compared to 159 in 2012)[114], and the number of frozen embryo replacement (FER) have increased by 11% between 2017 and 2016[114]. These changes and evolving trends require a close partnership with ML experts and clinics to continuously improve the predictions models through the so called '*model updating*'[115], where continuous adjustments of the prediction models are performed and re-calibrated in new populations from time to time, to provide accurate estimates of success rate changes over time.

5.3. Algorithm validation using double-blinded datasets

Evaluating model performance across populations, clinics and countries is extremely important. A model performance can be quantifiable using metrics such as positive-predictive value, sensitivity, and specificity. Probability-prediction models can be evaluated using discrimination (AUC or C statistic) and calibration metrics. The performance of a machine learning model can be assessed using data from the same source as the training sample (e.g. 80:20 data split for train and test data). That is, 20% of the cases are removed before the data was modelled; these removed cases were called the testing set. Once the model has been built using the 80% of the cases left (often called the training set), the cases which were removed (testing set) can be used to test the performance of the model on the "unseen" data (i.e. the testing set). However, a true evaluation of generalisability (also called transportability) typically requires external validation of a prediction model[116]. This external evaluation should be done in a completely blinded way, in some cases named '*double blinded*' evaluation[90], which consists of using data that not only has never been used in the training of the algorithm but also data originating from a completely different and independent clinical environments, and if possible different countries.

5.4. The challenge of translation to clinical practice

Challenges of translation of AI systems in infertility care include those inherent to the science of machine learning itself, such as difficulties in implementation and scalability in the deployment, but also the human barriers to AI adoption[117]. The human barriers to clinical adoption of an AI algorithm depend on two main factors: Clinical usefulness and trustworthiness[118]. Before implementation in the clinic we need to clearly understand how AI models will affect the quality of care; how they will improve on the efficiency and productivity of clinical practice, and, most importantly, how will they improve on patient outcomes. It is vital to demonstrate the added value of an AI algorithm when compared to legacy or conventional methods. Randomised controlled trials are viewed as the gold standard for robust clinical evaluation but conducting these in practice may not always be appropriate or feasible. If the algorithm is used to aid clinical decisions related to interventions that are expensive, invasive or have unwanted side-effects, we should be concerned with false-positive predictions that could result in unnecessary harm to the patient[118,121]. An example is the use of computer-aided diagnosis for mammography in the late 1990s, that due to false-positive predictions was found to significantly increase recall rate and surgical interventions without improving patient outcomes[119]. On the other hand, if the interventions are predominantly preventative or "assistive" (i.e aim to provide additional assistance to patients), we would want to minimise in the number of false negatives[118,121]. To this end, thoughtful post-market surveillance should be implemented to make sure that the algorithm does not create or exacerbate inequalities[118], for instance if certain groups of patients are deprived of access to beneficial innovations (e.g., in minority ethnic groups, or certain age groups[120]).

These so called 'unintended discriminatory bias' could happen when the algorithms are not generalizable to new populations. In order to overcome these potential pitfalls in the clinical adoption of any ML algorithm, it is important to systematically test the algorithms for bias and fairness[121,122].

Conclusions

Advanced algorithms associated with ML or AI have found their way into modern medicine and are transforming many aspects of patient care. Infertility and reproductive health are catching up, however much more can be done to improve the chance of conception for infertile couples. The major limitations include access to good quality and sufficient quantity of data to produce algorithms that would be reliable and reproducible on a global scale. Social understanding and knowledge of reproductive ageing is increasing and as such, a shift towards a more engaged individual in their own reproductive future can be seen. This translates to increased uptake of fertility MOTs and social egg freezing. As male reproductive ageing is not as pronounced, social sperm freezing is not a common practice. We live in an exciting time of telemedicine, ML/AI influence in every aspect of our lives. It is not unexpected, that attempts at integration of clinical experience and information technology into an amalgam aiming to improve health and prolong lifespan as well as affect our reproductive health are emerging everywhere. It is however important to critically appraise these and not blindly jump at any opportunity labelled 'AI will increase your chance of pregnancy'. One has to bear in mind that IVF is also a business and as such is governed by laws of the open market, where any new developments are firstly there to lure a customer in, with often no or limited regard for the service user.

In order to improve the development and uptake of AI algorithms in reproductive medicine, major obstacles would need to be overcome. Our understanding of the processes governing reproduction are not fully understood and as such the data that will form the basis of such algorithms, will be incomplete or on occasions incorrect. AI in the field or reproductive medicine may also be useful to sift through the plethora of available scientific data and help identify aspects of further research that would be worth pursuing. With expanding understanding of physiology, we would be able to identify pathology and devise individualized treatment protocols that would best suit the woman in order to guarantee reproductive success. Easy access to self-entered data via various apps may be of help in the former, but clinical input and experience in combination with data analysts and biostatisticians is necessary to increase the influence of ML/AI in reproductive health and move reproductive medicine into version 2.0.

References

1. Zegers-Hochschild F, Adamson GD, de Mouzon J, et al. The International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) Revised Glossary on ART Terminology, 2009. *Hum Reprod.* 2009;24(11):2683-2687.
2. Healy DL, Trounson AO, Andersen AN. Female infertility: causes and treatment. *Lancet.* 1994;343(8912):1539-1544.
3. Petraglia F, Serour GI, Chapron C. The changing prevalence of infertility. *Int J Gynaecol Obstet.* 2013;123 Suppl 2:S4-8.
4. Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S, Stevens GA. National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. *PLoS Med.* 2012;9(12):e1001356.
5. Levine H, Jorgensen N, Martino-Andrade A, et al. Temporal trends in sperm count: a systematic review and meta-regression analysis. *Hum Reprod Update.* 2017;23(6):646-659.
6. Quaas A, Dokras A. Diagnosis and treatment of unexplained infertility. *Rev Obstet Gynecol.* 2008;1(2):69-76.

7.   Baird DT, Collins J, Egozcue J, et al. Fertility and ageing. *Hum Reprod Update.* 2005;11(3):261-276.

8.   OECD. Age of mothers at childbirth and age-specific fertility. OECD - Social Policy Division - Directorate of Employment, Labour and Social Affairs. http://www.oecd.org/els/soc/SF_2_3_Age_mothers_childbirth.pdf. Published 2019. Accessed2020.

9.   De Geyter C, Calhaz-Jorge C, Kupka MS, et al. ART in Europe, 2014: results generated from European registries by ESHRE: The European IVF-monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE). *Hum Reprod.* 2018;33(9):1586-1601.

10.   Evers JL. Female subfertility. *Lancet.* 2002;360(9327):151-159.

11.   Palermo G, Joris H, Devroey P, Van Steirteghem AC. Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. *Lancet.* 1992;340(8810):17-18.

12.   Handyside AH, Kontogianni EH, Hardy K, Winston RM. Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature.* 1990;344(6268):768-770.

13.   Hood L. Systems biology: integrating technology, biology, and computation. *Mech Ageing Dev.* 2003;124(1):9-16.

14.   Kaye J, Curren L, Anderson N, et al. From patients to partners: participant-centric initiatives in biomedical research. *Nat Rev Genet.* 2012;13(5):371-376.

15.   Spandorfer SD, Avrech OM, Colombero LT, Palermo GD, Rosenwaks Z. Effect of parental age on fertilization and pregnancy characteristics in couples treated by intracytoplasmic sperm injection. *Hum Reprod.* 1998;13(2):334-338.

16.   Westrom L. Effect of pelvic inflammatory disease on fertility. *Venereology.* 1995;8(4):219-222.

17.   Tsevat DG, Wiesenfeld HC, Parks C, Peipert JF. Sexually transmitted diseases and infertility. *Am J Obstet Gynecol.* 2017;216(1):1-9.

18.   Norman RJ, Noakes M, Wu R, Davies MJ, Moran L, Wang JX. Improving reproductive performance in overweight/obese women with effective weight management. *Hum Reprod Update.* 2004;10(3):267-280.

19.   Balen AH, Dresner M, Scott EM, Drife JO. Should obese women with polycystic ovary syndrome receive treatment for infertility? *BMJ (Clinical research ed).* 2006;332(7539):434-435.

20.   Egg freezing in fertility treatment Trends and figures: 2010-2016. HFEA. https://www.hfea.gov.uk/media/2656/egg-freezing-in-fertility-treatment-trends-and-figures-2010-2016-final.pdf. Accessed.

21.   Goldman RH, Racowsky C, Farland LV, Munne S, Ribustello L, Fox JH. Predicting the likelihood of live birth for elective oocyte cryopreservation: a counseling tool for physicians and patients. *Hum Reprod.* 2017;32(4):853-859.

22.   HFEA. Egg freezing. https://www.hfea.gov.uk/treatments/fertility-preservation/egg-freezing/. Accessed2020.

23.   Tüttelmann F, Nieschlag E. Classification of Andrological Disorders. In: Nieschlag E, Behre HM, Nieschlag S, eds. *Andrology: Male Reproductive Health and Dysfunction.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2010:87-92.

24.   Yan S, Shabbir M, Yap T, et al. Should the current guidelines for the treatment of varicoceles in infertile men be re-evaluated? *Hum Fertil (Camb).* 2019:1-15.

25.   Mansour Ghanaie M, Asgari SA, Dadrass N, Allahkhah A, Iran-Pour E, Safarinejad MR. Effects of varicocele repair on spontaneous first trimester miscarriage: a randomized clinical trial. *Urol J.* 2012;9(2):505-513.

26.   Pastuszak AW, Wang R. Varicocele and testicular function. *Asian J Androl.* 2015;17(4):659-667.

27. Agarwal A, Deepinder F, Cocuzza M, et al. Efficacy of varicocelectomy in improving semen parameters: new meta-analytical approach. *Urology.* 2007;70(3):532-538.

28. Diaz-Gimeno P, Ruiz-Alonso M, Sebastian-Leon P, Pellicer A, Valbuena D, Simon C. Window of implantation transcriptomic stratification reveals different endometrial subsignatures associated with live birth and biochemical pregnancy. *Fertil Steril.* 2017;108(4):703-710 e703.

29. Heidari M, Darbandi S, Darbani M, et al. Evaluating the Potential of Three Sperm Surface Antigens as Egg-adhesion Biomarkers for Human Sperm Selection. *J Reprod Infertil.* 2018;19(4):203-210.

30. Nair TM. Statistical and artificial neural network-based analysis to understand complexity and heterogeneity in preeclampsia. *Comput Biol Chem.* 2018;75:222-230.

31. Gracie S, Pennell C, Ekman-Ordeberg G, et al. An integrated systems biology approach to the study of preterm birth using "-omic" technology--a guideline for research. *BMC Pregnancy Childbirth.* 2011;11:71.

32. Hulka BS, Wilcosky T. Biological markers in epidemiologic research. *Arch Environ Health.* 1988;43(2):83-89.

33. Sufriyana H, Wu YW, Su EC. Prediction of Preeclampsia and Intrauterine Growth Restriction: Development of Machine Learning Models on a Prospective Cohort. *JMIR Med Inform.* 2020.

34. Haller-Kikkatalo K, Salumets A, Uibo R. Review on autoimmune reactions in female infertility: antibodies to follicle stimulating hormone. *Clin Dev Immunol.* 2012;2012:762541.

35. Deroux A, Dumestre-Perard C, Dunand-Faure C, Bouillet L, Hoffmann P. Female Infertility and Serum Auto-antibodies: a Systematic Review. *Clin Rev Allergy Immunol.* 2017;53(1):78-86.

36. Empson MB, Lassere M, Craig JC, Scott JR. Prevention of recurrent miscarriage for women with antiphospholipid antibody or lupus anticoagulant. *Cochrane Database of Systematic Reviews.* 2005(2).

37. Jodar M, Selvaraju S, Sendler E, Diamond MP, Krawetz SA, Network RM. The presence, role and clinical use of spermatozoal RNAs. *Hum Reprod Update.* 2013;19(6):604-624.

38. Yang Q, Hua J, Wang L, et al. MicroRNA and piRNA profiles in normal human testis detected by next generation sequencing. *PLoS One.* 2013;8(6):e66809.

39. Mouillet JF, Ouyang Y, Coyne CB, Sadovsky Y. MicroRNAs in placental health and disease. *Am J Obstet Gynecol.* 2015;213(4 Suppl):S163-172.

40. Yoffe L, Gilam A, Yaron O, et al. Early Detection of Preeclampsia Using Circulating Small non-coding RNA. *Sci Rep.* 2018;8(1):3401.

41. Gupta S, Agarwal A, Sekhon L, Krajcir N, Cocuzza M, Falcone T. Serum and peritoneal abnormalities in endometriosis: potential use as diagnostic markers. *Minerva Ginecol.* 2006;58(6):527-551.

42. Kitawaki J, Ishihara H, Koshiba H, et al. Usefulness and limits of CA-125 in diagnosis of endometriosis without associated ovarian endometriomas. *Hum Reprod.* 2005;20(7):1999-2003.

43. Hirsch M, Duffy JMN, Deguara CS, Davis CJ, Khan KS. Diagnostic accuracy of Cancer Antigen 125 (CA125) for endometriosis in symptomatic women: A multi-center study. *Eur J Obstet Gynecol Reprod Biol.* 2017;210:102-107.

44. Karimi-Zarchi M, Dehshiri-Zadeh N, Sekhavat L, Nosouhi F. Correlation of CA-125 serum level and clinico-pathological characteristic of patients with endometriosis. *Int J Reprod Biomed (Yazd).* 2016;14(11):713-718.

45. Knific T, Vouk K, Vogler A, et al. Models including serum CA-125, BMI, cyst pathology, dysmenorrhea or dyspareunia for diagnosis of endometriosis. *Biomarkers in medicine.* 2018;12(7):737-747.

46. Farkas Z, Török P. Role of CA 125 level in management of endometriosis. *European Journal of Obstetrics and Gynecology and Reproductive Biology.* 2019;234:e109-e110.

47.   Diaz-Gimeno P, Horcajadas JA, Martinez-Conejero JA, et al. A genomic diagnostic tool for human endometrial receptivity based on the transcriptomic signature. *Fertil Steril.* 2011;95(1):50-60, 60 e51-15.

48.   Diaz-Gimeno P, Ruiz-Alonso M, Blesa D, et al. The accuracy and reproducibility of the endometrial receptivity array is superior to histology as a diagnostic method for endometrial receptivity. *Fertil Steril.* 2013;99(2):508-517.

49.   Reynolds K, Sarangi S, Bardia A, Dizon DS. Precision medicine and personalized breast cancer: combination pertuzumab therapy. *Pharmgenomics Pers Med.* 2014;7:95-105.

50.   Zondervan KT, Becker CM, Missmer SA. Endometriosis. *N Engl J Med.* 2020;382(13):1244-1256.

51.   Becker CM, Laufer MR, Stratton P, et al. World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: I. Surgical phenotype data collection in endometriosis research. *Fertil Steril.* 2014;102(5):1213-1222.

52.   Yatsenko SA, Rajkovic A. Genetics of human female infertility†. *Biol Reprod.* 2019;101(3):549-566.

53.   Miyamoto T, Minase G, Shin T, Ueda H, Okada H, Sengoku K. Human male infertility and its genetic causes. *Reprod Med Biol.* 2017;16(2):81-88.

54.   Gajbhiye R, Fung JN, Montgomery GW. Complex genetics of female fertility. *NPJ Genom Med.* 2018;3:29.

55.   Laven JSE. Follicle Stimulating Hormone Receptor (FSHR) Polymorphisms and Polycystic Ovary Syndrome (PCOS). *Front Endocrinol (Lausanne).* 2019;10:23.

56.   Zhang P-Y, Yu Y. Precise Personalized Medicine in Gynecology Cancer and Infertility. *Frontiers in Cell and Developmental Biology.* 2019;7.

57.   Legro RS, Barnhart HX, Schlaff WD, et al. Ovulatory response to treatment of polycystic ovary syndrome is associated with a polymorphism in the STK11 gene. *J Clin Endocrinol Metab.* 2008;93(3):792-800.

58.   Fishel S, Baker D, Elson J, et al. Precision Medicine in Assisted Conception: A Multicenter Observational Treatment Cohort Study of the Annexin A5 M2 Haplotype as a Biomarker for Antithrombotic Treatment to Improve Pregnancy Outcome. *EBioMedicine.* 2016;10:298-304.

59.   Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev.* 2005;14(8):1847-1850.

60.   Chen LW, Wu Y, Neelakantan N, Chong MF, Pan A, van Dam RM. Maternal caffeine intake during pregnancy and risk of pregnancy loss: a categorical and dose-response meta-analysis of prospective studies. *Public Health Nutr.* 2016;19(7):1233-1244.

61.   Gaskins AJ, Rich-Edwards JW, Williams PL, Toth TL, Missmer SA, Chavarro JE. Prepregnancy Low to Moderate Alcohol Intake Is Not Associated with Risk of Spontaneous Abortion or Stillbirth. *J Nutr.* 2015;146(4):799-805.

62.   Carré J, Gatimel N, Moreau J, Parinaud J, Léandri R. Does air pollution play a role in infertility?: a systematic review. *Environmental Health.* 2017;16(1):82.

63.   Shieh Y, Eklund M, Madlensky L, et al. Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. *J Natl Cancer Inst.* 2017;109(5).

64.   Esserman LJ, Investigators WSaA. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer.* 2017;3:34.

65.   Moreno I, Simon C. Deciphering the effect of reproductive tract microbiota on human reproduction. *Reprod Med Biol.* 2019;18(1):40-50.

66.   Chen C, Song X, Wei W, et al. The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases. *Nat Commun.* 2017;8(1):875.

67.   Hou D, Zhou X, Zhong X, et al. Microbiota of the seminal fluid from healthy and infertile men. *Fertil Steril.* 2013;100(5):1261-1269.

68.   Weng SL, Chiu CM, Lin FM, et al. Bacterial communities in semen from men of infertile couples: metagenomic sequencing reveals relationships of seminal microbiota to semen quality. *PLoS One.* 2014;9(10):e110152.

69.   Kyono K, Hashimoto T, Nagai Y, Sakuraba Y. Analysis of endometrial microbiota by 16S ribosomal RNA gene sequencing among infertile patients: a single-center pilot study. *Reprod Med Biol.* 2018;17(3):297-306.

70.   Kyono K, Hashimoto T, Kikuchi S, Nagai Y, Sakuraba Y. A pilot study and case reports on endometrial microbiota and pregnancy outcome: An analysis using 16S rRNA gene sequencing among IVF patients, and trial therapeutic intervention for dysbiotic endometrium. *Reprod Med Biol.* 2019;18(1):72-82.

71.   Vitagliano A, Saccardi C, Noventa M, et al. Effects of chronic endometritis therapy on in vitro fertilization outcome in women with repeated implantation failure: a systematic review and meta-analysis. (1556-5653 (Electronic)).

72.   Cicinelli E, Matteo M, Trojano G, et al. Chronic endometritis in patients with unexplained infertility: Prevalence and effects of antibiotic treatment on spontaneous conception. *Am J Reprod Immunol.* 2018;79(1).

73.   UK CR. We're saving lives through research: Annual Report & Accounts 2018/19. https://www.cancerresearchuk.org/sites/default/files/ec1060588_cruk_ar_2019_interactive .pdf. Published 2019. Accessed.

74.   NIHR. Reproductive Health. https://www.nihr.ac.uk/explore-nihr/specialties/reproductive-health.htm. Accessed 02.07.2020, 2020.

75.   Gleicher N, Kushnir VA, Barad DH. Worldwide decline of IVF birth rates and its probable causes. *Hum Reprod Open.* 2019;2019(3):hoz017.

76.   Ruiz-Alonso M, Blesa D, Diaz-Gimeno P, et al. The endometrial receptivity array for diagnosis and personalized embryo transfer as a treatment for patients with repeated implantation failure. *Fertil Steril.* 2013;100(3):818-824.

77.   Xu J, Fang R, Chen L, et al. Noninvasive chromosome screening of human embryos by genome sequencing of embryo culture medium for in vitro fertilization. *Proc Natl Acad Sci U S A.* 2016;113(42):11907-11912.

78.   Ghobadi C, Gregory A, Crewe HK, Rostami-Hodjegan A, Lennard MS. CYP2D6 is primarily responsible for the metabolism of clomiphene. *Drug Metab Pharmacokinet.* 2008;23(2):101-105.

79.   Ji M, Kim K-R, Lee W, Choe W, Chun S, Min W-K. Genetic Polymorphism of CYP2D6 and Clomiphene Concentrations in Infertile Patients with Ovulatory Dysfunction Treated with Clomiphene Citrate. *J Korean Med Sci.* 2016;31(2):310-314.

80.   Murdter TE, Kerb R, Turpeinen M, et al. Genetic polymorphism of cytochrome P450 2D6 determines oestrogen receptor activity of the major infertility drug clomiphene via its active metabolites. *Hum Mol Genet.* 2012;21(5):1145-1154.

81.   Simoni M, Gromoll J, Nieschlag E. The follicle-stimulating hormone receptor: biochemistry, molecular biology, physiology, and pathophysiology. *Endocr Rev.* 1997;18(6):739-773.

82.   Themmen APN, Huhtaniemi IT. Mutations of gonadotropins and gonadotropin receptors: elucidating the physiology and pathophysiology of pituitary-gonadal function. *Endocr Rev.* 2000;21(5):551-583.

83.   Nakayama T, Kuroi N, Sano M, et al. Mutation of the follicle-stimulating hormone receptor gene 5'-untranslated region associated with female hypertension. *Hypertension.* 2006;48(3):512-518.

84.   Achrekar SK, Modi DN, Desai SK, Mangoli VS, Mangoli RV, Mahale SD. Poor ovarian response to gonadotrophin stimulation is associated with FSH receptor polymorphism. *Reprod Biomed Online.* 2009;18(4):509-515.

85. Perez Mayorga M, Gromoll J, Behre HM, Gassner C, Nieschlag E, Simoni M. Ovarian response to follicle-stimulating hormone (FSH) stimulation depends on the FSH receptor genotype. *J Clin Endocrinol Metab.* 2000;85(9):3365-3369.

86. Behre HM, Greb RR, Mempel A, et al. Significance of a common single nucleotide polymorphism in exon 10 of the follicle-stimulating hormone (FSH) receptor gene for the ovarian response to FSH: a pharmacogenetic approach to controlled ovarian hyperstimulation. *Pharmacogenet Genomics.* 2005;15(7):451-456.

87. Sudo S, Kudo M, Wada S, Sato O, Hsueh AJ, Fujimoto S. Genetic and functional analyses of polymorphisms in the human FSH receptor gene. *Mol Hum Reprod.* 2002;8(10):893-899.

88. Laven JS, Mulders AG, Suryandari DA, et al. Follicle-stimulating hormone receptor polymorphisms in women with normogonadotropic anovulatory infertility. *Fertil Steril.* 2003;80(4):986-992.

89. Achrekar SK, Modi DN, Desai SK, Mangoli VS, Mangoli RV, Mahale SD. Follicle-stimulating hormone receptor polymorphism (Thr307Ala) is associated with variable ovarian response and ovarian hyperstimulation syndrome in Indian women. *Fertil Steril.* 2009;91(2):432-439.

90. VerMilyea M, Hall JMM, Diakiw SM, et al. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. *Hum Reprod.* 2020;35(4):770-784.

91. VerMilyea MD, Hall JM, Perugini D, et al. Artificial intelligence: non-invasive detection of morphological features associated with abnormalities in chromosomes 21 and 16. *Fertility and Sterility.* 2019;112(3):e237-e238.

92. Diagnosis and Management of Ectopic Pregnancy. *BJOG: An International Journal of Obstetrics & Gynaecology.* 2016;123(13):e15-e55.

93. Liu L, Jiao Y, Li X, Ouyang Y, Shi D. Machine learning algorithms to predict early pregnancy loss after in vitro fertilization-embryo transfer with fetal heart rate as a strong predictor. *Comput Methods Programs Biomed.* 2020;196:105624.

94. Wald NJ, Hackshaw AK. Combining ultrasound and biochemistry in first-trimester screening for Down's syndrome. *Prenat Diagn.* 1997;17(9):821-829.

95. Health Quality O. Noninvasive Prenatal Testing for Trisomies 21, 18, and 13, Sex Chromosome Aneuploidies, and Microdeletions: A Health Technology Assessment. *Ont Health Technol Assess Ser.* 2019;19(4):1-166.

96. Gynaecologists RCoOa. Amniocentesis and Chorionic Villus Sampling. Green-top guideline No.8. *RCOG Press: London.* 2010.

97. Kuhrt K, Smout E, Hezelgrave N, Seed PT, Carter J, Shennan AH. Development and validation of a tool incorporating cervical length and quantitative fetal fibronectin to predict spontaneous preterm birth in asymptomatic high-risk women. *Ultrasound Obstet Gynecol.* 2016;47(1):104-109.

98. Watson HA, Seed PT, Carter J, et al. Development and validation of predictive models for QUiPP App v.2: tool for predicting preterm birth in asymptomatic high-risk women. *Ultrasound Obstet Gynecol.* 2020;55(3):348-356.

99. Frydman G. A patient-centric definition of participatory medicine. *Retrieved May.* 2010;12:2016.

100. Stuppia L, Gatta V. Genetic testing in infertile couples. *The EuroBiotech Journal.* 2017;1(s2):151.

101. Yurttas Beim P, Parfitt DE, Tan L, et al. At the dawn of personalized reproductive medicine: opportunities and challenges with incorporating multigene panel testing into fertility care. *J Assist Reprod Genet.* 2017;34(12):1573-1576.

102. Petrucelli N, Daly MB, Pal T. BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews((R)).* Seattle (WA)1993.

103.   Harrell Jr FE. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis.* Springer; 2015.

104.   Lynam AL, Dennis JM, Owen KR, et al. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. *Diagn Progn Res.* 2020;4:6.

105.   Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018.

106.   te Velde ER, Nieboer D, Lintsen AM, et al. Comparison of two models predicting IVF success; the effect of time trends on model performance. *Hum Reprod.* 2014;29(1):57-64.

107.   Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med.* 2002;137(7):598-602.

108.   Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med.* 1978;299(17):926-930.

109.   Dyer S, Chambers GM, de Mouzon J, et al. International Committee for Monitoring Assisted Reproductive Technologies world report: Assisted Reproductive Technology 2008, 2009 and 2010. *Hum Reprod.* 2016;31(7):1588-1609.

110.   Sauerbrei W. Prognostic factors. Confusion caused by bad quality design, analysis and reporting of many studies. *Adv Otorhinolaryngol.* 2005;62:184-200.

111.   Knottnerus JA. Between iatrotropic stimulus and interiatric referral: the domain of primary care research. *J Clin Epidemiol.* 2002;55(12):1201-1206.

112.   HFEA. Multiple births minimisation strategy. In. https://ifqlive.blob.core.windows.net/umbraco-website/1314/2009-12-09_authority_papers_-_527_multiple_births.pdf2009.

113.   Braude P. One child at a time: reducing multiple births after IVF. In: IVF EGoMBa, ed. https://ifqlive.blob.core.windows.net/umbraco-website/1311/one-child-at-a-time-report.pdf2006.

114.   HFEA. Fertility   treatment   2017: trends and figures. In. https://www.hfea.gov.uk/media/2894/fertility-treatment-2017-trends-and-figures-may-2019.pdf2017.

115.   Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol.* 2008;61(1):76-86.

116.   Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ.* 2016;353:i3140.

117.   Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17(1):195.

118.   Vollmer S, Mateen BA, Bohner G, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ.* 2020;368:l6927.

119.   Lehman CD, Wellman RD, Buist DS, et al. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. *JAMA Intern Med.* 2015;175(11):1828-1837.

120.   Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Intern Med.* 2018;178(11):1544-1547.

121.   Saleiro P, Kuester B, Hinkson L, et al. Aequitas: A bias and fairness audit toolkit. *arXiv preprint arXiv:181105577.* 2018.

122.   Chen IY, Szolovits P, Ghassemi M. Can AI Help Reduce Disparities in General Medical and Mental Health Care? *AMA J Ethics.* 2019;21(2):E167-179.