Personalized ovarian stimulation for assisted reproductive technology: study design considerations to move from hype to added value for patients

Ben W. Mol, M.D., Ph.D., a Patrick M. Bossuyt, Ph.D., b Sesh K. Sunkara, M.D., c Juan A. Garcia Velasco, M.D., Ph.D., d Christos Venetis, M.D., Ph.D., e Denny Sakkas, Ph.D., f Kersti Lundin, Ph.D., g Carlos Simón, M.D., Ph.D., h Hugh S. Taylor, M.D., i Robert Wan, Ph.D., j Salvatore Longobardi, M.D., Ph.D., k Evelyn Cottell, Ph.D., k and Thomas D’Hooghe, M.D., Ph.D. k,l,m

a Monash University, Clayton, Victoria, Australia; b University of Amsterdam, Amsterdam, the Netherlands; c Queen’s Hospital, Barking, Havering and Redbridge University Hospitals, Essex, United Kingdom; d IVI RMA Madrid and Rey Juan Carlos University, Madrid, Spain; e University of New South Wales, Sydney, New South Wales, Australia; f Boston IVF, Waltham, Massachusetts; g Sahlgrenska University Hospital, Goteborg, Sweden; h Igenomix, Valencia University, Valencia, Spain; i Yale University Medical School, New Haven, Connecticut; j EMD Serono, Billerica, Massachusetts; k Global Medical Affairs Fertility, Research and Development, Merck KGaA, Darmstadt, Germany; l Research Group Reproductive Medicine, Department of Development and Regeneration, Organ Systems, Group Biomedical Sciences, KU Leuven (University of Leuven), Leuven, Belgium; and m Department of Obstetrics and Gynecology, Yale University, New Haven, Connecticut

Although most medical treatments are designed for the average patient with a one-size-fits-all–approach, they may not benefit all. Better understanding of the function of genes, proteins, and metabolite, and of personal and environmental factors has led to a call for personalized medicine. Personalized reproductive medicine is still in its infancy, without clear guidance on treatment aspects that could be personalized and on trial design to evaluate personalized treatment effect and benefit–harm balance. While the rationale for a personalized approach often relies on retrospective analyses of large observational studies or real-world data, solid evidence of superiority of a personalized approach will come from randomized trials comparing outcomes and safety between a personalized and one-size-fits-all–strategy. A more efficient, targeted randomized trial design may recruit only patients or couples for which the personalized approach would differ from the previous, standard approach. Multiple monocenter studies using the same study protocol (allowing future meta-analysis) might reduce the major center effect associated with multicenter studies. In certain cases, single-arm observational studies can generate the necessary evidence for a personalized approach. This review describes each of the main segments of patient care in assisted reproductive technologies treatment, addressing which aspects could be personalized, emphasizing current evidence and relevant study design. (Fertil Steril 2018;109:968–79. Copyright © 2018 The Authors. Published by Elsevier Inc. on behalf of the American Society for Reproductive Medicine. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).)

Key Words: Personalized medicine, precision medicine, trial design, ART

Received March 13, 2018; revised April 23, 2018; accepted April 25, 2018.

B.W.M. is supported by a NHMRC Practitioner Fellowship (GNT1082548) and has provided consultancy for ObsEva, Merck KGaA and Gubernet. P.M.B. has nothing to disclose. S.K.S. has received grants and non-financial support from Merck and Ferring. J.A.G.V. has received grants from Merck KGaA, Merck, Sharp & Dohme and Ferring. C.V. is supported by a NHMRC Early Career Fellowship (GNT1147154) and has received personal fees and non-financial support from Merck Serono and Merck, Sharp & Dohme, and non-financial support from Ferring. D.S. has received grants from Finox Pharmaceutical, has served on a scientific advisory board for Origio, and on a speakers panel for EMD Serono. K.L. is engaged in the Scientific Educational Board for the Nordic branch of Merck-Serono and is a principle investigator in a study sponsored by Vitrolife. C.S. is scientific officer of Igenomix and has received grants from Merck, MSD and Ferring. H.S.T. has nothing to disclose. R.W. is an employee of EMD Serono, an affiliate of Merck KGaA, Darmstadt, Germany. S.L. is an employee of Merck KGaA, Darmstadt, Germany. E.C. is an employee of Merck KGaA, Darmstadt, Germany. T.D.H. is an employee of Merck KGaA, Darmstadt, Germany. Medical writing for this article funded by Merck KGaA, Darmstadt, Germany. Reprint requests: Thomas D’Hooghe, M.D., Ph.D., Merck KGaA, Frankfurturter Str. 250, F135/002, 64293 Darmstadt, Germany (E-mail: thomas.dhooghe@merckgroup.com).

Fertility and Sterility® Vol. 109, No. 6, June 2018 0015-0282
Copyright ©2018 The Authors. Published by Elsevier Inc. on behalf of the American Society for Reproductive Medicine. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). https://doi.org/10.1016/j.fertnstert.2018.04.037

Most medical treatments are designed for the average patient, with a one-size-fits-all–approach. Though successful for many, this approach may not benefit all patients. An improved understanding of the function of genes, proteins, metabolites, and personal and environmental factors has led to a call for personalized medicine: a tailored approach to disease prevention and treatment that considers inter-individual differences in patients. Despite some successes, especially in oncology, personalized medicine is
still in its infancy in reproductive medicine, possibly because treatment success (live birth) is determined by many baseline factors (in both males and females) and treatment factors (surgical, endocrinological, gamete, embryo, and uterine) (1). This review describes each of the main segments of the assisted reproductive technology (ART) patient journey, starting from ovarian stimulation and ending in live birth, addressing which aspects of patient care could be personalized, both with respect to clinical care and to future research.

WHAT ARE THE AIMS OF FERTILITY TREATMENT?

The main aim of fertility treatment is to enable infertile couples to have a baby, and live birth is therefore the preferred primary outcome of clinical trials evaluating the effectiveness of fertility treatments (2). In reality, intermediate pregnancy outcomes (including chemical pregnancy rate, clinical pregnancy rate [CPR], and ongoing pregnancy rate [OPR]) are most commonly evaluated, with only a minority of randomized controlled trials (RCTs) in reproductive medicine reporting live birth outcomes (3). However, as these intermediate endpoints are, overall, strongly positively correlated with live birth rates (LBRs) (4, 5), a comparison between treatment groups may not be compromised. The positive association between number of oocytes, number of good quality embryos (GQE) and (cumulative) LBRs (6–8) suggests that the number of oocytes retrieved and number of embryos available may be relevant intermediate outcomes for the comparison of various ART treatments. However, it is uncertain if these outcomes are solely a consequence of the treatment, or whether they are significantly influenced by the intrinsic characteristics of the woman.

The safe delivery of a healthy baby may take several treatment cycles; therefore, the cumulative LBR per patient starting treatment is a key outcome variable. These cumulative LBRs are influenced by treatment efficacy (LBR per started treatment cycle) and intra- and inter-cycle discontinuation of treatment. While there is a lot of focus on efficacy, patient care is equally as important, as shown by the high discontinuation rates observed during ART treatment in both countries with low reimbursement rates (45% in the U.S. [9]) and countries with public reimbursement (48%–65% in the Netherlands [10], Canada [11], and Sweden [12]). Trials have shown that in systems without funding limitations, psychological stress is the most common reason for drop-out from ART treatment (13–16). A personalized management strategy, taking into account quality of care (17), targeting patient expectations, and aimed at minimizing the treatment burden and side effects as much as possible, with informed and shared decision-making between patients and healthcare professionals, could alleviate the psychological stress. Such a strategy could lead to reduced discontinuation rates and impact favorably on ART treatment outcomes by shortening the time interval between the start of treatment and conception leading to live birth. Pregnancy loss should always be incorporated in clinical decision making, owing to its psychological impact (4).

Another important consideration is safety, and particularly the potential for complications arising either directly from fertility treatment or through pregnancy or delivery. For example, while two healthy babies born from a twin pregnancy at term may be considered a good outcome for some patients (18), multiple pregnancy increases the risk of complications for both mother and babies. Achieving a singleton pregnancy should therefore be considered an important aim of treatment. Additional considerations that are relevant for patients and professionals are the time taken to achieve a live birth (19), cost and quality of care, and quality of life (20).

Time to pregnancy is of particular importance because infertility and its treatment are distressing, with couples wanting pregnancy to occur as soon as possible (19). When endpoints such as time to pregnancy are evaluated, a standard time frame should be used (e.g., time after randomization) to ensure comparability. This has not yet been adequately defined, with the latest International Committee Monitoring Assisted Reproductive Technologies (ICMART) glossary defining time to pregnancy as, “The time taken to establish a pregnancy, measured in months or in numbers of menstrual cycles” (21). We propose that time from treatment initiation (defined as the start of hormonal treatment for ovarian stimulation) to a clinical pregnancy (diagnosed by ultrasonographical visualization of one or more gestational sacs or definitive clinical signs of pregnancy [21]) that results in a live birth should be reported rather than the time to live birth, as this provides a standard starting point and will control for the gestational period.

Financial costs are also an important outcome: in a private healthcare system costs can determine whether a treatment is affordable and in a public healthcare system they can determine the overall accessibility of fertility treatment, including the number of ART cycles that are reimbursed per patient.

The individual expectations of patients should be well understood before treatment is started, to better enable personalization of the approach, and success based upon these expectations should form the basis of any treatment evaluation (22, 23).

EVALUATING PRECISION MEDICINE DURING THE ART TREATMENT JOURNEY

Controlled Ovarian Stimulation

The aim of controlled ovarian stimulation (COS) is to optimize the number of oocytes retrieved, so that sufficient oocytes can be safely obtained for ART treatment. Ovarian response is associated with the success of ART treatment (8, 24, 25).

Pituitary suppression protocol. According to three meta-analyses (26–28), gonadotropin-releasing hormone (GnRH) antagonist treatment is associated with a similar LBR compared with GnRH agonist treatment but with a lower incidence of any grade of ovarian hyperstimulation syndrome (OHSS). The reduced risk for OHSS is possibly also related to the use of a GnRH agonist instead of human chorionic gonadotropin (hCG) to trigger final oocyte maturation (29). In one meta-analysis (28)—the methodology of which has been
criticized (30)—a lower OPR was observed after antagonist treatment in the overall population but not in women with polycystic ovary syndrome (PCOS) or in women with poor ovarian response (POR). Overall, the GnRH antagonist protocol can be proposed for patients with expected normal or high ovarian response. However, there are not sufficient data to warrant abstaining from using antagonist protocols in poor responders.

**Gonadotropin dose: starting dose and dose adjustment.** Ovarian reserve markers, including basal follicle-stimulating hormone (FSH), basal estradiol, inhibin B, antral follicle count (AFC), and anti-Müllerian hormone (AMH), are often used as predictors of ovarian response and for selection of the FSH dose. AFC and AMH have the highest accuracy for predicting poor and excessive response following ovarian stimulation (OS) (31). Individual patient data meta-analyses have indicated that poor response is predicted by AFC <7 and AMH <1.1 ng/mL (32–34) and hyper response by AFC >14 (35, 36) and AMH >3.5 ng/mL (37, 38).

An individualized follitropin alfa starting dose based on a nomogram (age, basal FSH, and AMH) resulted in a higher proportion of women obtaining the stated target of 8–14 oocytes compared with a fixed 150 IU dose (63% vs. 42%) (39). However, this superiority could not be confirmed for a similar reproductive outcome (proportion of women obtaining 5–12 oocytes) in another trial comparing treatment with either a fixed or an individualized gonadotropin starting dose (29). This was possibly because a mixture of follitropin alfa originator and biosimilar drugs were used in both treatment arms (and biosimilars differ from originators with respect to chemical structure, preclinical activity, and clinical effect) and/or a different nomogram was used (age, body mass index [BMI], AFC, and AMH). In non-inferiority trials evaluating individualized versus fixed starting-dose regimens for the primary outcomes of OPR (40) or number of oocytes retrieved (41), any conclusions suggesting non-inferiority (40) or lack of non-inferiority (41) can be challenged; both treatment arms were different with respect to starting dose, type of gonadotropin used (40) and dose adjustment policy (only allowed in control group) (40, 41).

However, no study, due to the small sample size, has shown that individualizing FSH starting dose based on ovarian response markers improves LBR (42). For predicted normal responders, more oocytes were retrieved with 200–225 IU FSH daily compared with 100–150 IU daily, with no significant difference observed with 225 IU compared with 300 IU daily (42). In predicted low responders, daily gonadotropin doses ≥300 IU were associated with a greater number of oocytes retrieved and significantly lower cycle cancellation due to poor response compared with doses of 150 IU, but LBRs were not improved by the higher dose (42). Similarly, the OPTIMIST study in predicted poor responders (AFC <10) showed that individualized FSH dosing (225 or 450 IU) was associated with a greater number of oocytes retrieved, a reduced incidence of cycle cancellation (4–8% vs. 30%), and a similar LBR compared with a standardized dose of 150 IU FSH (43).

In a meta-analysis (44) of three RCTs in poor responders, a higher number of oocytes retrieved but similar CPRs were observed in women after COS with a “higher” gonadotropin starting dose (450 IU [two RCTs] or 600 IU [one RCT]) compared with COS using a “lower” gonadotropin starting dose (300 IU [two RCTs] or 150 IU [one RCT]). The data presented suggest recommended starting doses of 150–225 IU and 225–300 IU for predicted normal responders and predicted poor responders, respectively.

In expected hyper responders, lower FSH doses should be used, as they reduce the risk for OHSS without compromising success rates. The OPTIMIST study (45) in predicted hyper responders (AFC >15) showed that a 150 IU daily dose of FSH, compared with a 100 IU daily dose, significantly increased the number of oocytes retrieved and was associated with a lower first cycle cancellation rate (12% vs. 24%). However, the higher daily dose was also associated with a higher risk of mild-to-moderate OHSS (11% vs. 4%). More studies are needed to identify the optimal gonadotropin starting dose in expected high responders, taking into account not only reproductive outcomes but also the risk for OHSS and cycle cancellation.

Although dose adjustment has been allowed in most current trials, no robust trials have evaluated the effect on LBR of FSH dose adjustment during treatment (45). Individual dose adjustment after a first failed cycle and dose adjustment during OS are important areas of future research.

**Minimal/mild ovarian stimulation.** Minimal/mild OS regimens are advocated as a cost-effective alternative to conventional OS, but there is no standardized regimen for minimal stimulation. Suggested strategies include the use of anti-estrogens (e.g. clomiphene citrate) and/or aromatase inhibitors (e.g. letrozole), either alone or in combination with low-dose gonadotropins. While a recent meta-analysis observed that mild OS protocols reduced the amount of gonadotropin required and the incidence of OHSS, these protocols were also associated with a significant increase in the incidence of cycle cancellations, as well as reductions in the mean number of oocytes retrieved. Furthermore, no conclusive evidence for live-birth or pregnancy rates in both the general population and women with POR was seen (46). More evidence is needed before minimal/mild regimens are adopted into clinical practice (46); this should also take into account time to pregnancy and treatment discontinuation rate.

**Oocyte triggering.** In spontaneous ovulation cycles, FSH and luteinizing hormone (LH) show a mid-cycle surge, whereas triggering of final oocyte maturation in COS protocols traditionally relied solely on the activity of hCG to mimic the LH surge. More recently, triggering with a GnRH agonist has been employed in GnRH antagonist cycles to stimulate the LH surge. A meta-analysis of 17 trials observed that triggering with a GnRH agonist instead of hCG in fresh autologous in vitro fertilization (IVF)/intracytoplasmic sperm injection (ICSI) cycles (13 trials) reduced the risk for OHSS but resulted in a much lower LBR (29). However, a more recent meta-analysis of five trials of a GnRH agonist trigger followed by luteal phase support (LPS) with LH activity versus hCG trigger
followed by standard LPS, observed that the LBR did not differ significantly between the groups. Optimization of LPS is needed to further limit OHSS in normal responder patients. In addition, no difference in the risk of OHSS was reported (OHSS occurred in 4/413 patients receiving GnRH agonist and 7/413 patients receiving hCG) (47). In donor–recipient cycles (four trials) there was no difference in LBR or OPR (29).

A meta-analysis of four trials observed a significantly higher pregnancy rate with dual trigger (hCG and GnRH agonist) compared with hCG alone (48). However, no significant differences were observed between the groups in the number of oocytes retrieved, number of mature oocytes retrieved, number of fertilized oocytes, number of good-quality embryos or implantation rate (IR) (48). Collectively, these data suggest that a GnRH agonist trigger could be useful for women undergoing freeze-all cycles, women donating oocytes, and women freezing their oocytes for fertility preservation. However, despite the reduced risk for OHSS, a few cases of severe early-onset OHSS have been reported, indicating that further fine-tuning of individualized LPS might improve results even further.

Fresh embryo transfer (ET) should not be disregarded and a GnRH agonist trigger can be used to obtain a high LBR with a low risk for OHSS. An analysis of the cost-effectiveness and impact on the patient of GnRH agonist trigger and LPS compared with hCG (the gold standard trigger) would better enable the clinical significance of both options to be judged.

**ART Laboratory Procedures**

**Sperm diagnostic tests.** Compared with oocyte yield optimization and embryo selection, considerably less research effort has gone towards improving sperm yield and selection. Tests have been proposed to assess aspects including motility patterns (using computer-derived measures), chromatin errors, and apoptotic markers. However, none of these tests has been evaluated thoroughly and few are used routinely in the clinic.

A high degree of sperm DNA fragmentation is associated with poorer pregnancy rates and LBRs in couples trying to conceive naturally, by intrauterine insemination (IUI) and undergoing routine IVF, but not when ICSI was used (49–51). Meta-analysis also showed that the assay used to analyze DNA fragmentation can also have an impact on the correlations: the terminal deoxynucleotidyl transferase (TdT) dUTP nick-end labeling (TUNEL) assay and the single-cell gel electrophoresis (COMET) assay have better predictive capacity than the sperm chromatin structure assay (SCSA) and sperm chromatin dispersion (SCD) test (52). A major disadvantage of DNA fragmentation testing is that the assay renders the tested sample unsuitable for clinical use. Studies are needed to confirm prospectively the hypothesis (49–51) that, in couples with a high degree of sperm DNA fragmentation, ART treatment with ICSI leads to a higher OPR and LBR than ART treatment with IVF.

When DNA damage is suspected, intra–cytoplasmic morphologically selected sperm injection (IMSI) could be used; for example, in cases of oligoasthenoteratozoospermia, with or without recurrent implantation failure after ICSI. A meta-analysis reported improved implantation and pregnancy rates after motile sperm organelle morphology examination (MSOME) for couples with previous ICSI only or male factor infertility (53). These findings were reiterated in a systematic review of 22 trials (54), but studies reporting on LBR after MSOME followed by ICSI (IMSI) compared with ICSI alone are of poor quality (55). In summary, while the concept of advanced sperm selection shows potential for personalizing the treatment approach, there are no studies that demonstrate its clinical value.

**Type of insemination/fertilization.** The type of fertilization method is another possibility for individualization, whereby the selection criteria for IUI, IVF, or ICSI is often based on a couple’s reproductive and clinical history, sperm diagnostic tests, and “post-preparation” sperm recovery and motility rates (56). Variable concentrations of viable sperm have been used as indicators for changing from IUI to IVF. One study reported that IVF should be selected if fewer than 10 million viable sperm are present in the total ejaculate (57), whereas other studies recommend this concentration in a processed sample (58). For ICSI, there is no cut-off value except sufficient viable sperm, and ICSI should be considered in cases where the sperm sample is poor (59), there is low “post-preparation” yield (often specified as a total number of sperm less than 1 million), previous fertilization failure (56) or in cases of surgically retrieved sperm (60).

Regarding implantation, a large data analysis showed that, compared with conventional IVF, ICSI was not associated with improved LBRs per transfer, irrespective of the presence of a diagnosis of male factor infertility (61). This finding was also supported by an RCT in couples with non-male factor infertility (62). However, a meta-analysis of studies where sibling oocytes were split between IVF and ICSI suggested that ICSI might increase fertilization rates and decrease risk of total failure to fertilize in couples with well-defined unexplained infertility (63), although this has not been evaluated in a RCT.

**Embryo culture, selection, and day of transfer.** Since only a limited number of fertilized oocytes/cleavage stage embryos will become good quality blastocysts on day 5 (56, 64), the number of retrieved oocytes, 2PN oocytes and/or the number of good GQE on day 2/3 can be used to individualize the selection of the day of transfer (65). However, it is well known that a proportion of non-GQE at day 2/3 will become good quality blastocysts (64).

Continuous embryo monitoring (CEM) systems enabled by time-lapse technology (TLT) hold promise to improve ART (66, 67). Such technology confers several practical benefits to the IVF laboratory (68) and enables the identification of abnormal embryo developments, including direct or reverse cleavage, which have been shown to be associated with reduced IR in retrospective studies (69, 70). However, its potential to improve embryo selection and therefore clinical outcomes is still under debate.

Recent studies have shown that fresh and frozen embryo transfer (FET) result in similar LBR, opening the way to single embryo transfer (SET) during both fresh and frozen cycles (71, 72). Assessment of embryo quality to identify the embryo
with the highest implantation potential is especially relevant for SET, as appropriate embryo selection may reduce the time to pregnancy and live birth by improving implantation rate and reducing early pregnancy loss.

**Pre-implantation genetic testing for aneuploidy.** As over half of the embryos produced by IVF are aneuploid, pre-implantation genetic testing for aneuploidy (PGT-A) may identify embryos with optimal implantation potential, reduce the number of ET, and time to pregnancy in ART (73–75). Clinical outcomes are calculated as sustained implantation, referring to the number of fetal sacs over 12 or 20 weeks of gestation, and as pregnancy rates per transfer, per started cycle or per patient (76).

Mosaicism originates from the first embryo cleavage and can be identified by a trophectoderm biopsy analyzing several cells. However, it is challenging to calculate the real incidence of mosaicism in preimplantation embryos because of technical limitations (77, 78). It is important to note that the true incidence of mosaicism that could impact the accuracy of the diagnosis should not be higher than 6% (79) and the clinically recognizable error rate after PGT-A is low (80). More recently, several studies have proposed the possibility of transferring some types of mosaic embryo as they might develop into healthy euploid newborns, in particular when no euploid embryos resulted from the aneuploidy testing and low levels of mosaicism are detected in the biopsy (81–83).

In patients at advanced maternal age, overall, one or two COS cycles are needed to obtain five MII oocytes that result in at least one euploid blastocyst (76). Obviously, in low responders the number of cycles needed to obtain one euploid embryo will increase but the only other option is to transfer euploid embryos which are more likely to result in miscarriage or no pregnancy.

In both cleavage stage and trophectoderm biopsy, extended embryo culture is needed, and ET is performed at the blastocyst stage. This strategy could decrease the number of embryos available for transfer in patients at advanced maternal age, especially women <40 years old who are undergoing PGT-A with a small cohort of available blastocysts for biopsy. However, embryo IR with embryos not reaching the blastocyst stage are poorer. In addition, it has been recently suggested that morula can be biopsied on day 6, but these morulae are associated with higher aneuploidy rates and lower IRs compared with blastocysts (84).

A meta-analysis evaluated whether PGT-A with comprehensive chromosome screening (PGT-A-CCS) improves clinical IR and sustained IR (beyond 20 weeks) compared with routine care for embryo selection in IVF cycles. In three RCTs (n=659; all patients with normal ovarian reserve and good prognosis) at the blastocyst level, PGT-A-CCS was associated with a significantly higher clinical IR (73). However, screening with PGT-A as well as extended culture increases the risk of having no embryos for transfer, and comparative trials should be based on LBRs per started cycle. In addition, there are at present no conclusive strategies for reducing the impact of varying levels of mosaicism. Therefore, although PGT-A and selection of chromosomally euploid embryos may reduce the miscarriage rate and shorten time to pregnancy, especially in women at advanced age (74, 75), it has not been shown to improve cumulative LBR.

**Embryo culture media.** The addition of granulocyte-macrophage colony-stimulating factor (GM-CSF), which is reported to have a protective effect on embryo stress to the embryo culture medium has been observed to have a modest positive effect on ongoing IR (OIRs) at week 7 and also on LBR, compared with control (85). Post-hoc analyses showed that GM-CSF increased OIRs in women with more than one miscarriage. However, this study had an adaptive design and the positive effect of GM-CSF was not seen when human serum albumin concentration in culture media was increased. The addition of hyaluronic acid to the culture medium showed moderate quality evidence of improvements in CPTs and LBRs in a meta-analysis of 17 trials (86). Again, further RCTs are required to assess whether these treatments are of benefit, specifically in women with unexplained miscarriage.

**Personalizing Endometrial Receptivity**

Embryo implantation depends not only on the embryo but also on the endometrial window of implantation (W0I). Supraphysiological levels of hormones during COS are associated with modifications to the endometrium, including greater endometrial advancement and altered gene expression (87–89). These modifications have been suggested to have an impact on implantation and obstetric and perinatal outcomes (90). The use of freeze-all strategies has been suggested to optimize endometrial receptivity, as the transfer will happen into a more “natural” endometrium. The protocols that can be used for FET are natural cycle, modified natural cycle (ovulation is induced during a natural cycle), artificial (hormone replacement treatment [HRT]) cycle (endometrial receptivity is induced by exogenous progesterone exposure after proper priming with exogenous estradiol; GnRH agonist co-treatment may be employed to down-regulate the pituitary and prevent follicular growth) and stimulated cycles (gonadotropins are used to induce follicular development and ovulation is triggered with hCG).

A retrospective study of 1926 FET cycles performed using either an artificial or stimulated protocol, observed that artificial cycles were associated with a greater incidence of early pregnancy loss and a lower LBR compared with stimulated cycles (91). A meta-analysis that included 20 studies did not show differences in outcomes between the natural-cycle protocol, the modified natural cycle protocol or the artificial protocol (92). These data suggest that stimulated cycles might provide improved outcomes compared with natural and artificial cycles, but this should be studied by direct comparisons of the different protocols.

Another proposed method to improve endometrial receptivity is mechanical endometrial injury, also called ‘scratching’. This has been proposed to positively affect the chance of implantation after ET, but the currently available evidence (after 14 years of use) is not yet conclusive (93, 94). Ongoing large clinical trials (e.g. The SCRaTCH study) (94–96) could be
analyzed to identify any treatment selection markers that might be related to beneficial treatment outcomes.

**Identifying the window of implantation.** Failure of implantation is considered a major cause of unexplained infertility, with inadequate endometrial receptivity responsible for about two-thirds of implantation failure (97). Although most women (70%) will reach receptivity after 5 full days (120 hours; P+5) of progesterone administration in HRT cycles, or 7 days after the LH surge (168 hours; LH+7) in natural cycles, some others show a displaced (pre-receptive or post-receptive) WOI, possibly leading to recurrent implantation failure. Identification of a personalized WOI has been proposed, to synchronize ET with the optimal receptive period in a strategy known as personalized ET (pET). Endometrial transcriptomics has been suggested as a reliable and objective method for endometrial assessment according to its gene expression pattern (98, 99), leading to the development of a number of assays, including the endometrial receptivity analysis (ERA) test (100), the E-tegrity test® (101), the endometrial function test® (EFT) (102) and ReveptivaDx (103).

These tissue profiling technologies use genomic and molecular markers to evaluate the endometrium and aim to assist in optimizing treatment, including the implantation date, according to each individual’s characteristics. The ERA test (https://www.igenomix.com/provider-tests/endometrial-receptivity-test-era) is based on next generation screening (NGS) of 236 genes identified as being involved in endometrial receptivity and has shown clinical promise in patients who repeatedly failed IVF (104, 105). The E-tegrity test is based on immunohistochemical staining of single molecules such as alpha-1, alpha-4, and beta-3 integrins (http://www.etegrity.com) (102), the EFT is based on detection of cyclin E and cyclin-dependent kinase inhibitor p27 (http://klimanlab.yale.edu/infertility/eft) (102) and endometrial BCL6 testing is proposed for ReveptivaDx (103). Endometrial receptivity has also been analyzed by immunohistochemical detection of uterine natural killer (uNK) cells. However, the prognostic value of measuring total uNK cells or CD56(+) cells in endometrial specimens remains uncertain (106).

**Embryo Transfer**

The number of embryos transferred can be individualized according to the viability of the embryos, the age of the woman, and the number of previous (successful or unsuccessful) transfers. In some parts of the world, SET policies, mandating the transfer of a single embryo in the majority of cases, have successfully reduced the incidence of multiple pregnancies. However, in most countries, multiple pregnancies and births continue to be a burden, and concentrated efforts are needed to reduce this risk. Stricter SET policies for particular women, including younger women, and encouraging women with a high number of GGE towards SET and cryopreservation of the excess embryos, rather than toward multiple ET, would reduce multiple pregnancies. Good quality blastocysts obtained during the first or second cycles in women aged <36 years should be considered for SET, whereas women with a POR phenotype receiving double-embryo cleavage stage transfer during the first cycle have only a modestly increased risk of multiple births (107).

**Freeze all strategies.** Data from two RCTs suggest that freeze-all strategies do not benefit ovulatory women (71, 72). However, these trials did not analyze whether the effectiveness of fresh or frozen transfer success varied according to baseline characteristics that are biologically linked to implantation, for example, increased progesterone concentrations at the time of hCG triggering (108) or a thin endometrium (109, 110). Indeed, observational data support a negative impact on endometrial receptivity and reproductive outcome from elevated progesterone concentrations or endometrial thickness <7 mm determined by vaginal ultrasound at the end of COS (109–113). These biomarkers should be incorporated into the design of RCTs, to evaluate if they can guide better treatment decisions.

Freeze-all strategies followed by FET could also be beneficial for ovulatory women with increased risk for OHSS, but, the cost-effectiveness of these strategies still requires evaluation. Women with PCOS may benefit from a freeze–all strategy with subsequent FET, as a higher frequency of live births and a lower frequency of pregnancy losses has been observed after FET rather than fresh transfer (114).

**Early pregnancy Management**

According to most current guidelines, early pregnancy management for both natural and ART conceived pregnancies is similar, with the exception of LPS for ART pregnancies, as indicated.

**Luteal phase support.** LPS aims to compensate for the dysfunctional corpus luteum resulting from supra-physiological estradiol levels during OS suppressing LH levels.

Traditionally, patients who receive hCG to trigger oocyte maturation receive LPS (either vaginal, oral, intra-muscular, subcutaneous, or a combination) with progesterone (115). If a GnRH agonist is used as a trigger, LPS may be challenging, as there will be luteal phase insufficiency, although individual variability in steroid production has been described (116) and an ideal LPS protocol for use after GnRH trigger has not yet been defined. Patients can be treated with either intensive steroid support (high doses of oral estradiol plus intra-muscular progesterone) (117) or with low dose hCG, with the dose individualized according to the number of oocytes retrieved (118). Providing LPS via LH activity after agonist trigger is a possibility, as LH would stimulate corpora lutea to produce progesterone and other steroids. This could be done with LH directly, but this is inconvenient as it requires LH injections every 48 hours and is unlikely to be cost-effective. An alternative suggested by Humaidan et al. is to use low dose hCG (118), but this may increase the risk for OHSS after using a GnRH agonist trigger if hCG dosing is inadequate. Otherwise, all oocytes or embryos can be frozen rather than transferred, to avoid secondary OHSS if the patient gets pregnant (119). During HRT FET cycles, the luteal phase needs to be fully supported rather than supplemented, and progesterone is always needed after priming of the endometrium with estrogens.
Traditionally, patients start progesterone supplementation around the time of oocyte pick-up after hCG trigger, although starting on the day of the oocyte pick-up does not improve outcomes compared with starting 6 days later (115). Once LPS is initiated, most physicians do not monitor progesterone serum levels to titrate the dose of medication until withdrawal, rather serum levels are maintained empirically until weeks 10–12 of pregnancy. However, individual differences in absorption of the medication and patient compliance may have a clinical impact or affect the timing of treatment, meaning individualization of LPS is needed.

It is well known that vaginal progesterone and uterine levels do not correlate well with serum levels (120, 121), however, testing serum samples is the only way to monitor individual absorption into peripheral blood. Labarta et al. (122) recently demonstrated that serum levels of progesterone vary on the day of ET in women undergoing oocyte donation supplemented with micronized vaginal progesterone (400 mg twice per day), with lower pregnancy rates in patients with serum progesterone <9.2 ng/mL compared with patients with progesterone ≥9.2 ng/mL. The WOI may also vary among women with infertility, but more research is needed to evaluate if and how serum progesterone and estradiol levels can be used to guide personalized LPS (116).

Stimulated ART cycles in which the corpus luteum is present do not need additional progesterone support after the pregnancy test, as the hCG used for triggering supports the corpus luteum for 5–7 days. Exogenous progesterone will then support the endometrium until pregnancy, after which sufficient hCG is secreted to support the pregnancy. Once hCG is detected in the blood, there is no need to maintain progesterone supplementation (123–126). In HRT FET cycles, in which there is no corpus luteum, support is required until the luteo–placental shift takes place.

Obstetric Management

Multiple pregnancies resulting from the transfer of multiple embryos are a cause of significant adverse obstetric, fetal, perinatal and neonatal risks, but can be resolved with SET and FET. Other ART procedures, such as oocyte donation, have consistently been associated with increased risk of adverse obstetric and perinatal outcomes (127, 128). There is accumulating evidence of a higher risk of large-for-gestational-age infants and higher birth weight with FET compared with fresh ET, and fresh transfers are associated with higher risks of preterm birth, low birth weight and born small-for-gestational-age (129). When attributing risks to ART, it is important to note that the above-mentioned associations are based on observational studies, and that the pregnancy complications may be related to infertility itself and not to the treatment. Moreover, it is currently unknown whether higher surveillance is warranted with ART pregnancies. The most important group of infertile women that require special attention in pregnancy are women with PCOS who become pregnant, specifically those with the hyperandrogenic phenotype (130, 131).

**How can the superiority of personalized ART treatment compared with standardized ART treatment be established?**

A number of factors may influence the outcome of treatment, including the characteristics of the couple and the treatment strategy selected, and these are also likely to interact. It may be possible to identify specific patient characteristics (e.g. female age, BMI, and duration or cause of infertility, AFC, AMH or previous ovarian response) that enable optimal treatment selection (stratified treatment). It must be emphasized, however, that when evaluating personalized medicine the diagnosis should not be considered alone, and the benefit–harm balance of the treatment relative to the individual patient profile should be a major consideration for treatment selection. As such, personalized treatment requires sound evidence from strong scientific research to identify these subgroups, and to show that they genuinely differ in outcomes or benefit–harm balance.

The rationale for a personalized approach is likely to come from retrospective analyses of large observational studies or real-world data. Such analyses may suggest interaction between baseline features, markers or test results on one hand and treatment outcomes on the other. Analytical methods such as propensity scoring can be used to reduce biases to a lack of comparability in baseline and treatment characteristics between groups. Given the often small differences that are expected between personalized and standard approaches, such methods will rarely provide definitive evidence.

Thus, the value of personalized treatment selection markers should be evaluated in properly designed clinical trials. While it might seem obvious to directly compare a personalized strategy and the standard, one-size-fits-all strategy in eligible patients, one should realize that differences in outcome will only occur in those patients in whom the personalized treatment strategy differs from the one-size-fits-all strategy. As one may consider changing treatment for a subgroup only, a more efficient, targeted comparative trial would randomize only patients or couples for whom the personalized approach would differ from the previous, standard approach (132).

In certain cases, single-arm observational studies can generate the knowledge necessary for a personalized approach. This may apply if such studies can convincingly demonstrate the absence of a desired outcome after specific treatments in one subgroup, but not in others, demonstrating the absence of a treatment benefit in the former.

Despite the interest in personalized medicine, many trials do not report outcomes according to patient characteristics; rather, the focus is on reporting only the main outcomes without secondary analyses. We recommend that every RCT should include secondary analyses exploring potential treatment selection markers that identify differences in benefits across subgroups. These secondary analyses should ideally be performed with properly a priori specified hypotheses, if well-supported by prior evidence, or by original and relevant scientific rationale if prior evidence is not available. Where
possible these secondary analyses should be done with sufficient power and precision, enabling separate publication. However, it will not always be possible to provide sufficient power for secondary analyses, and in such situations exploratory secondary analyses (ideally pre-specified in study protocol, but also valuable as post hoc analyses to explain unexpected results) can still be useful to inspire further investigation and eventual confirmation in other trials, and the results could be included in a systematic review.

As ART is a stepwise process and involves a number of important treatment decisions at different time points, optimization at each time point may be expected to, but not necessarily, result in a globally optimized treatment protocol. As a result of this, it would be nearly impossible to compare the differences between all different treatment options throughout the ART treatment journey using a standard study design, owing to the unrealistic sample size and number of treatment arms required, in addition to concerns of heterogeneity.

Heterogeneity not only exists at the level of patients (age, duration and cause of infertility, ovarian reserve, number and type of previous infertility therapies, and related responses, with ovarian response to OS possibly influenced by genetic polymorphisms for LH, FSH and their receptors) (133), but can also arise from the knowledge, skill and opinions of different healthcare professionals (doctors, embryologists and nurses) in different centers. This heterogeneity is known to affect intermediate outcomes during ART treatment, such as the number of follicles observed by ultrasound scan during OS, number of oocytes retrieved after oocyte aspiration, pregnancy rate per ET, oocyte damage rate and fertilization rate after ICSI, and survival rate of cryopreserved embryos after warming/thawing. This heterogeneity in treatment effect, contributed to by both patients and healthcare professionals, results in the known center effect (134, 135), which explains why it has been impossible so far to apply models predicting pregnancy or live birth in the real world (134).

As the benefits from personalized ART treatment may be small yet clinically relevant, future multicenter trials could try to minimize the variability from center effects. Proof-of-principle evidence of the benefits of using a personalized approach can also be obtained in separate trials using the same study protocol in selected centers, with invitations based upon their established interest, commitment, trial expertise, and experience in addressing the primary study question. The selection of centers needs to be as fair as possible and efforts would be needed to ensure the selection is not influenced by economic, political and academic bias. The studies should aim to demonstrate proportional improvements in intermediate outcomes (i.e. number and quality of oocytes/embryos) and ultimately in CPRs and LBRs, as well as safety, effectiveness in terms of patient-centered outcomes, and cost-effectiveness. In addition, when personalized ART treatment is compared with standard ART treatment, these studies should be conducted in well characterized patient phenotypes. In any case, the outcomes observed in trials in selected centers should subsequently be confirmed using registries reflecting care in the real world.

Overall, this highlights that both innovative study designs and outcome assessment methods need to be further developed before a globally optimized, personalized treatment protocol can be easily produced.

CONCLUSION

One can wonder whether any form of standardized care can exist without taking the personal characteristics, concerns and aspirations of a couple into account, and whether personalized care can exist without any form of standardization. Any form of quality care will be, to some extent, both standardized and personalized.

In this review, we have identified the most relevant and emerging areas of personalized ART and have proposed several approaches, taking the specific characteristics and aspirations of the couple into account, to develop more solid evidence of the (differential) effectiveness, safety, cost and treatment burden for the couples involved.

Acknowledgments: Medical writing support provided by Alexander Jones and Steven Goodrick, inScience Communications.

REFERENCES


