Rethinking the approach to endometrial non-receptivity: treating the cause instead of the consequence

Samuel dos Santos Ribeiro

Orientadores: Prof. Doutor Christophe Blockeel
            Prof. Doutor Carlos Calhaz Jorge
            Prof. Doutor Herman Tournaye

Tese especialmente elaborada para obtenção dos graus de Doutor em Ciências Médicas (Vrije Universiteit Brussel) e Medicina – Especialidade de Ginecologia e Obstetricia (Universidade de Lisboa)

2017
Rethinking the approach to endometrial non-receptivity: treating the cause instead of the consequence

Samuel dos Santos Ribeiro

Orientadores: Prof. Doutor Christophe Blockeel
               Prof. Doutor Carlos Calhaz Jorge
               Prof. Doutor Herman Tournaye

Tese especialmente elaborada para obtenção dos graus de Doutor em Ciências Médicas (Vrije Universiteit Brussel) e Medicina – Especialidade de Ginecologia e Obstetrícia (Universidade de Lisboa)

Júri:
Presidente: Prof. Doutora Hilde Van de Velde (Vrije Universiteit Brussel, Bruxelas, Bélgica)

Vogais: Prof. Doutor Carlos Simón (Universitat de València, València, Espanha)
        Prof. Doutor Carlos Eugénio Plancha dos Santos (Universidade de Lisboa, Lisboa, Portugal)
        Prof. Doutor Alberto Manuel Barros da Silva (Universidade do Porto, Porto, Portugal)
        Prof. Doutor Willem Verpoest (Vrije Universiteit Brussel, Bruxelas, Bélgica)
        Prof. Doutora Anick De Vos (Vrije Universiteit Brussel, Bruxelas, Bélgica)
        Prof. Doutor Christophe Blockeel (Vrije Universiteit Brussel, Bruxelas, Bélgica)
        Prof. Doutor Carlos Calhaz Jorge (Universidade de Lisboa, Lisboa, Portugal)

2017
Samuel dos Santos Ribeiro was born on May 30th 1983, in Benoni (Johannesburg). In 1993, he moved to Portugal where he performed most of his education until ultimately obtaining his Medical Degree in 2007 from Nova Medical School (Lisbon). Between 2008 and 2015, he followed a Specialty Training Programme in Obstetrics and Gynaecology at Hospital Santa Maria (Lisbon). In 2013, he initiated a two-year Clinical Scholars Research Training organized by Harvard Medical School (Boston) and his Subspecialty Training in Reproductive Medicine at the Centre for Reproductive Medicine of the Universitair Ziekenhuis Brussel (Brussels). He is currently a consultant of Reproductive Medicine and Genetics at Universitair Ziekenhuis Brussel.

The primary focus of his ongoing research is understanding better the relationship between infertility and its treatment on endometrial receptivity. Specifically, one of his major fields of interest is unravelling the potential effect of circulating progesterone during ovarian stimulation on the chances of pregnancy following assisted reproduction. He is also a lead investigator of two clinical trials and one prospective study assessing the potential benefit of the following interventions on the clinical outcomes after in-vitro fertilization: endometrial scratching with same-cycle endometrial expression analysis, female urogenital microbiome assessment and the “freeze-all” embryos strategy. He has co-authored 19 publications in peer-reviewed scientific journals (7 of which as first author) and performed 48 scientific presentations in international conferences.
To my parents and brothers, to whom I will be forever thankful.
And to Marta, who made it all worthwhile.
CONTENTS

1 Extended abstract........................................................................................................ II

2 List of abbreviations .................................................................................................. 15

3 List of publications and manuscripts ..................................................................... 17

4 Introduction............................................................................................................. 21
  4.1 The prevalence and burden of infertility............................................................ 24

4.2 In-vitro fertilization for infertility: benchmarks of success............................... 25
    4.2.1 Early pregnancy complications associated with in-vitro fertilization......... 28
    4.2.1.1 Ectopic pregnancy.................................................................................. 28
    4.2.1.2 Monozygotic twinning.......................................................................... 29

4.3 How may in-vitro fertilization affect endometrial receptivity?......................... 30
    4.3.1 Circulating sexual hormones and endometrial receptivity........................ 32
    4.3.1.1 Estradiol.................................................................................................. 33
    4.3.1.2 Progesterone........................................................................................ 36
    4.3.2 Ovulation triggering.................................................................................... 41
    4.3.3 Fresh embryo transfer and its predictors of success.................................... 44
    4.3.4 Embryo cryopreservation and the “freeze-all strategy”.............................. 45

4.4 Novel methods to predict endometrial receptivity............................................. 47
    4.4.1 Molecular diagnostic tools.......................................................................... 47
    4.4.1.1 The controversy around endometrial scratching.................................. 47
    4.4.1.2 Steps towards same-cycle endometrial receptivity assessment............ 48
    4.4.2 Microbiome of the female genital tract...................................................... 49

5 Thesis aims............................................................................................................. 51

6 Materials and methods ......................................................................................... 55
  6.1 Ovarian stimulation and pituitary suppression................................................ 55
  6.2 Cycle monitoring and final oocyte maturation triggering................................ 55
  6.3 Insemination, embryo quality assessment and cryopreservation................... 56
  6.4 Endometrial preparation for frozen embryo transfer.................................... 57
  6.5 Progesterone immunoassay............................................................................... 58

6.6 Main outcome measures...................................................................................... 58

6.7 Study population, design and statistical analysis.............................................. 59
    6.7.1 Specific aim 1............................................................................................ 59
    6.7.1.1 Study 1-A............................................................................................ 59
    6.7.1.2 Study 1-B............................................................................................ 60
    6.7.1.3 Study 1-C............................................................................................ 61
    6.7.2 Specific aim 2............................................................................................ 63
    6.7.2.1 Study 2-A............................................................................................ 63
    6.7.2.2 Study 2-B............................................................................................ 65
    6.7.3 Specific aim 3............................................................................................ 65
EXTENDED ABSTRACT

Approximately 9% of all couples suffer from infertility. Of these, up to 30% will eventually require in-vitro fertilization. Interestingly, following a steady increase of live birth delivery rates until the late 1990s, these have remained disappointingly below 30% per cycle. The fact that up to two thirds of all morphologically top-quality embryos eventually fail to implant is a testament that embryo morphologic assessment alone has a limited capacity to predict pregnancy.

Many authors have postulated that the supraphysiologic milieu of hormones produced during ovarian stimulation may affect endometrial receptivity and hinder both embryo implantation and neonatal outcomes. Of all the candidate biomarkers evaluated during the monitoring of ovarian stimulation, the abnormal production of progesterone during the later stages of the follicular phase has been, thus far, proposed most frequently as the best surrogate for endometrial receptivity. Specifically, late-follicular elevated progesterone (usually defined as a circulating progesterone above 1.50 ng/mL) has been linked to lower pregnancy rates following a fresh embryo transfer. Although the complete mechanism behind this hindered pregnancy outcome remains elusive, some investigators consider that this issue may be, at least in part, amenable by elective embryo cryopreservation, given that previous translational studies have shown an abnormal gene expression in the luteal phase of endometria exposed prematurely to progesterone. This has led many centres to change their clinical practice by measuring serum progesterone levels on the day of ovulation triggering and adopting a freeze-all strategy whenever the threshold of 1.50 ng/mL is exceeded. Meanwhile, others have proposed additional measures to enhance endometrial receptivity during in-vitro fertilization overall, among which endometrial scratching has become an increasingly widespread, although broadly misunderstood, alternative.

The general aim of this doctoral thesis was to improve our understanding on how controlled ovarian stimulation affects endometrial receptivity and to what extent one can change clinical practice to enhance pregnancy outcome. To this effect, we formulated specific objectives aimed to tackle the potential mechanisms behind stimulation-related
endometrial non-receptivity from multiple angles. Overall, our intention was mainly to address two overarching hypotheses: 
a) which are the best tools a physician may use to predict and optimize endometrial receptivity during a fresh embryo transfer? (studied in specific aims 1 and 2) and b) what are the clinical implications of a pragmatic approach (i.e. the freeze-all strategy) towards hindered endometrial receptivity during a fresh embryo transfer? (evaluated in specific aims 3 and 4).

Specific aim 1

In specific aim 1 of this thesis, we sought to better understand the relationship between the pregnancy outcome following assisted reproduction and circulating progesterone during ovarian stimulation. Attempting to determine whether circulating progesterone is detrimental for embryo quality and/or endometrial receptivity, we hypothesized that progesterone may affect (cumulative) live birth rates and that the mechanism behind it may be better explained by evaluating the cumulative exposure to abnormal progesterone. We performed three retrospective studies which associated progesterone produced during ovarian stimulation to hindered outcomes on multiple levels, including both embryo quality and endometrial receptivity. This result effectively extended the potential damaging effect of elevated progesterone beyond the first embryo transfer onto subsequent frozen embryo transfers as well. Furthermore, we also perceived that the relationship between abnormal late-follicular progesterone and live birth following assisted reproduction seems to be non-linear and that it may be of varying importance according to the number of oocytes retrieved. Finally, we also concluded that exposure to progesterone during ovarian stimulation could affect live birth rates even in women who would be originally expected to have unaltered chances of conception if only late-follicular progesterone levels is to be measured (i.e. those with serum progesterone levels <1.50 ng/mL on the day of ovulation triggering).

Specific aim 2

The evaluation of endometrial receptivity in a fresh embryo transfer cycle has been mostly limited to the assessment of endometrial thickness. In specific aim 2, we proposed to re-evaluate how measuring endometrial thickness performed in the prediction of live birth rates when accounting for the supraphysiologic endocrine milieu present during ovarian
stimulation. Furthermore, we pitted this measurement against a more novel approach of in-cycle endometrial receptivity assessment, *i.e.* transcriptomic expression analysis. Our theory was that endometrial thickness would no longer predict pregnancy outcomes when the hormonal effects of ovarian stimulation were taken into consideration and that endometrial transcriptomic profiling would prove to be more reliable. Additionally, we formulated an ongoing randomized control trial to evaluate prospectively the potential benefit of in-cycle endometrial scratching for same-cycle clinical pregnancy rates (the REFRESH trial, with 197 of 360 patients recruited so far) which we describe in detail. Following a large retrospective analysis, endometrial thickness was reconfirmed as a valuable prognostic tool for in-vitro fertilization, namely in the prediction of live birth rates and neonatal birthweight. This predictive capacity was independent of the effect of the supraphysiologic endocrine milieu on endometrial receptivity, a conclusion that goes against the evidence thus far pointing to the limited value of endometrial evaluation using ultrasonography in predicting treatment outcomes beyond clinical pregnancy.

**Specific aim 3**

An increasing number of researchers have alluded to the potential benefit of distancing the transfer of embryos produced during assisted reproductive technologies away from the ovarian stimulation cycle. The scientific evidence frequently used to justify this recent trend of the so-called elective “freeze-all strategy” includes randomized controlled trials demonstrating an increase in live birth after elective frozen embryo transfers and large retrospective population studies associating frozen embryo transfer cycles with less pregnancy-related complications such as prematurity, low birth weight and ectopic pregnancy. That said, another objective of this thesis (specific aim 3) was to compare the early pregnancy outcomes between fresh and frozen embryo transfers. Our hypothesis was that endometria subject to ovarian stimulation would be associated with worse early pregnancy outcomes (*i.e.* lower clinical pregnancy rates and higher miscarriage, ectopic pregnancy and monozygotic twinning rates) when compared to unstimulated uteri. This was, thus, a pragmatic approach to evaluate the effect of ovarian stimulation on endometrial receptivity. In order to test this scientific question, we designed an ongoing randomized controlled trial (the ICE trial, with 210 of 212 patients recruited so far) evaluating the clinical pregnancy rates among high-responder women performing either a fresh embryo transfer or a freeze-all approach. However, given that both ectopic
pregnancy and monozygotic twinning are rare events, two additional retrospective studies were also designed to evaluate each of these outcomes more extensively. Within the studies included in this specific aim, we observed that ectopic pregnancy rates following assisted reproduction have decreased overtime and may no longer be related to recent ovarian stimulation. Conversely, the transfer of a frozen embryo seemed to be associated with lower monozygotic twinning rates. Whether this latter result is due to an effect caused by the embryo cryopreservation method or by the transfer of an embryo in an unstimulated endometrium requires further investigation.

**Specific aim 4**

Finally, taking advantage of the large dataset of treatment cycles performed yearly at Universitair Ziekenhuis Brussel, we were interested in evaluating whether the empirical decision that physicians frequently take to postpone frozen embryo transfers in order to avoid any potential residual effect of ovarian stimulation served any purpose (specific aim 4). Our hypothesis was that ovarian stimulation had no carry-over effect to subsequent menstrual cycles and that the intentional deferral of embryo transfers beyond the menstrual cycle that follows and in-vitro fertilization attempt would play no role in increasing pregnancy rates. Given that the luteal phase of cycles triggered by exogenous human chorionic gonadotropin and gonadotropin-releasing hormone agonists vary substantially, two studies were conceived to evaluate the differences between immediate and deferred frozen embryo transfers on pregnancy rates in these two settings. In both instances, our results confirmed that intentionally deferring frozen embryo transfers may be of no benefit for in-vitro fertilization pregnancy rates.
# List of Abbreviations

Manuscript-specific abbreviations are defined at first appearance

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAHRPP</td>
<td>Association for the Accreditation of Human Research Protection Program</td>
</tr>
<tr>
<td>AE</td>
<td>Adverse events</td>
</tr>
<tr>
<td>AH</td>
<td>Assisted hatching</td>
</tr>
<tr>
<td>AIC/BIC</td>
<td>Akaike’s/Bayesian’s information criteria</td>
</tr>
<tr>
<td>AMP</td>
<td>Adenosine monophosphate</td>
</tr>
<tr>
<td>ART</td>
<td>Assisted reproductive technologies</td>
</tr>
<tr>
<td>BESST</td>
<td>Birth Emphasizing a Successful Singleton at Term</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>CL</td>
<td>Corpus luteum</td>
</tr>
<tr>
<td>CLBR</td>
<td>Cumulative live birth rate</td>
</tr>
<tr>
<td>COC</td>
<td>Cumulus-oocyte complex</td>
</tr>
<tr>
<td>COS</td>
<td>Controlled ovarian stimulation</td>
</tr>
<tr>
<td>CP</td>
<td>Clinical pregnancy</td>
</tr>
<tr>
<td>CPR</td>
<td>Clinical pregnancy rates</td>
</tr>
<tr>
<td>CRG</td>
<td>Centrum voor Reproductieve Geneeskunde</td>
</tr>
<tr>
<td>E2</td>
<td>Estradiol</td>
</tr>
<tr>
<td>EMT</td>
<td>Endometrial thickness</td>
</tr>
<tr>
<td>EP</td>
<td>Ectopic pregnancy</td>
</tr>
<tr>
<td>EQ</td>
<td>Embryo quality</td>
</tr>
<tr>
<td>ER</td>
<td>Endometrial receptivity</td>
</tr>
<tr>
<td>ET</td>
<td>Embryo transfer</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>FET</td>
<td>Frozen embryo transfer</td>
</tr>
<tr>
<td>FP</td>
<td>Fractional polynomials</td>
</tr>
<tr>
<td>FSH</td>
<td>Follicle-stimulating hormone</td>
</tr>
<tr>
<td>GEE</td>
<td>Generalized estimating equations</td>
</tr>
<tr>
<td>GnRH</td>
<td>Gonadotropin-releasing hormone</td>
</tr>
<tr>
<td>hCG</td>
<td>Human chorionic gonadotropin</td>
</tr>
<tr>
<td>HFEA</td>
<td>Human Fertilisation and Embryology Authority</td>
</tr>
<tr>
<td>HMG</td>
<td>Human menopausal gonadotropins</td>
</tr>
<tr>
<td>ICE</td>
<td>Implantation enhancement by elective Cryopreservation of all viable Embryos (clinical trial)</td>
</tr>
<tr>
<td>ICM</td>
<td>Inner cell mass</td>
</tr>
<tr>
<td>ICMART</td>
<td>International Committee for Monitoring Assisted Reproductive Technology</td>
</tr>
<tr>
<td>ICSI</td>
<td>Intracytoplasmic sperm injection</td>
</tr>
<tr>
<td>ITT</td>
<td>Intention-to-treat</td>
</tr>
<tr>
<td>IVD</td>
<td>In-vitro decidualization</td>
</tr>
<tr>
<td>IVF</td>
<td>In-vitro fertilization</td>
</tr>
<tr>
<td>KPI</td>
<td>Key performance indicators</td>
</tr>
<tr>
<td>LB</td>
<td>Live birth</td>
</tr>
<tr>
<td>LBR</td>
<td>Live birth rate</td>
</tr>
<tr>
<td>LBW</td>
<td>Low birthweight</td>
</tr>
<tr>
<td>LFEP</td>
<td>Late-follicular elevated progesterone</td>
</tr>
<tr>
<td>LH</td>
<td>Luteinizing hormone</td>
</tr>
<tr>
<td>MZT</td>
<td>Monozygotic twinning</td>
</tr>
<tr>
<td>NC</td>
<td>Natural cycle</td>
</tr>
<tr>
<td>NK</td>
<td>Natural killer</td>
</tr>
<tr>
<td>OHSS</td>
<td>Ovarian hyperstimulation syndrome</td>
</tr>
<tr>
<td>PGD</td>
<td>Pre-implantation genetic diagnosis</td>
</tr>
<tr>
<td>PIGF</td>
<td>Placental growth factor</td>
</tr>
<tr>
<td>PP</td>
<td>Per protocol</td>
</tr>
<tr>
<td>PR</td>
<td>Pregnancy rates</td>
</tr>
<tr>
<td>PRL</td>
<td>Prolactin</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td>PV-OS</td>
<td>Progesterone variation during ovarian stimulation</td>
</tr>
<tr>
<td>REFRESH</td>
<td>Receptivity Enhancement by Follicular-phase Renewal after Endometrial ScratChing (clinical trial)</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species</td>
</tr>
<tr>
<td>SAE</td>
<td>Serious adverse events</td>
</tr>
<tr>
<td>TF-FET</td>
<td>Timing of the first frozen embryo transfer</td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>UZ Brussel</td>
<td>Universitair Ziekenhuis Brussel</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular endothelial growth factor</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>WOI</td>
<td>Window of implantation</td>
</tr>
</tbody>
</table>
# LIST OF PUBLICATIONS AND MANUSCRIPTS

## Published Manuscripts (ordered by publication date)


5. **Santos-Ribeiro S, Polyzos NP, Haentjens P, Smitz J, Camus M, Tournaye H, Blockeel C.** *Live birth rates after IVF are reduced by both low and high progesterone levels on the day of human chorionic gonadotrophin administration.* Hum Reprod. 2014;29(8):1698-705.

## Manuscript proofs submitted to indexed medical periodic journals


3. Costa Ribeiro V, Santos-Ribeiro S, De Munck N, Drakopoulos P, Polyzos NP, Schutyser V, Verheyen G, Tournaye H, Blockeel C. *Should we continue to measure endometrial thickness in modern-day medicine?* (accepted following revision in RBM online).

**Trial protocols submitted to indexed medical periodic journals**


2. Santos-Ribeiro S, Tournaye H, Blockeel C. *Elective blastocyst vitrification for endometrial receptivity enhancement in high-responders: the Implantation enhancement by elective Cryopreservation of all viable Embryos (ICE) trial protocol.*

**Other published manuscripts not included in thesis (but relevant to the field of study)**


In 2009, award-winning Julian Jones wrote, directed and produced the movie *The Great Sperm Race*, which, corroborated by the scientific support of doctors Allan Pacey, Chris de Jonge, Joanna Ellington, Geoffrey Miller and Alex Travis, detailed the sequence of events required for a live birth to occur (Jones, 2009). Although narrated using simple terms, the author brilliantly described one of the largest marvels of human existence: that every single human being is unique and derives from the fertilization of an oocyte by a sperm cell, which themselves, are alike none. While the documentary gives most emphasizes to the uniqueness of each human being, it unwittingly brings into focus how little we know about human fertility, a process which, at times, seems intrinsically rigged to fail, demanding perfect synchrony to develop the biological equivalent of a lottery win that is the human being (Evers, 2002).

Fecundity, which is essentially defined as the capacity of a species to deliver a liveborn, is the best endpoint to measure the ability of spontaneous reproduction. While in other mammals, fecundity is as high as 80% per menstrual cycle (Stevens, 1997), normal and otherwise fertile human couples are not necessarily as successful. Specifically, the monthly reproductive efficiency of humans is approximately 20% (Evers, 2002) and, even following timed coitus, does not exceed 33% (Wilcox, et al., 1995). These figures put into light the exclusive nature of fertility and why it is oddly a time-based diagnosis, normally defined as the failure to conceive despite regular unprotected sexual intercourse over a period of at least 12 months (Zegers-Hochschild, et al., 2017).

The female natural menstrual cycle entails an intricate cross-talk between the hypothalamus, pituitary, ovaries and uterus that occur simultaneously to ensure mono-follicular development and endometrial receptivity (ER). Although a detailed explanation of the human menstrual cycle would be beyond the scope of this thesis, it is helpful to summarize the key features of what pertains to assisted reproductive technologies (ART). The female ovarian reproductive cycle can be divided in the follicular phase, ovulation and the luteal phase (Figure 1, p. 22). When describing the modifications occurring in the
endometrium during the cycle, the phase occurring between menses and ovulation is commonly referred to as the proliferative phase, while the interval following ovulation is called the secretory phase. The key hormones responsible for the crosstalk between these organs during the cycle are the pituitary-secreted gonadotropins follicle-stimulating hormone (FSH) and luteinizing hormone (LH) and the sexual ovary-derived steroid hormones estradiol (E2) and progesterone (P). At the beginning of each menstrual cycle, the secretion of FSH promotes the development of a small cohort of recruited antral follicles (Peters, et al., 1975). These follicles, in response, produce multiple paracrine and endocrine hormones, including E2 and inhibin, which collectively are responsible for the

**Figure 1 – The female menstrual cycle**
Gonadotropin control of the ovarian and endometrial cycles. The ovarian-endometrial cycle has been structured as a 28-day cycle. The follicular phase (days 1 to 14) is characterized by rising estrogen levels, endometrial thickening, and selection of the dominant “ovulatory” follicle. During the luteal phase (days 14 to 21), the corpus luteum produces E2 and P, which prepare the endometrium for implantation. If implantation occurs, the developing blastocyst begins to produce hCG and rescues the corpus luteum, thus maintaining P production. Reprinted from Cunningham (2014) with permission.
Introduction

optimal synchronization between a growing dominant follicle and the proliferating endometrium (Fritz and Speroff, 2012, Visser, et al., 2012).

FSH and LH have a synergistic action during the follicular phase which is conducive to adequate follicular development and the production of sexual hormones by the two layers of cells surrounding the follicles – the inner granulosa and outer theca (Fritz and Speroff, 2012). While FSH stimulates the production of E2 by the granulosa cells, LH promotes the conversion of cholesterol to androgens, which will be metabolized further after being transported into the granulosa cells (Figure 2, p. 23). The model proposed to describe the interaction between these key-players is called the two-cell, two-gonadotropin theory (Fevold, 1941).

Over the course of approximately 14 days in a 28-day cycle (World Health Organization, 1983), these endocrine and paracrine signalling pathways ensure, in most cases, the dominance of a single follicle while the remaining recruited follicles ultimately arrest, causing the ovulation of a single oocyte (Visser, et al., 2012). Follicular arrest is thought to be caused by an E2-regulated progressive decrease of endogenous FSH below which only

Figure 2 – The cross-talk between gonadotropins and sexual hormones throughout the follicular phase (the two-cell, two-gonadotropin theory)

During the early phases of the follicular phase, LH and FSH work synergistically to convert cholesterol into androgens (i.e. androstenedione and testosterone) in the theca cells which will later be converted into estrogens (i.e. E2 and estrone) in the granulosa cells. Inhibin (and, possibility, activin as well), serve as FSH auto-regulators by negative feedback. In the late follicular phase, the granulosa cells commence a LH-dependent production of P alongside the FSH-dependent production of E2, marking the initial stages that will eventually lead to luteinization. Reprinted from Fritz and Speroff (2012) with permission.
dominant follicles are able to continue development [coined by Brown (1978) as the *FSH threshold*]. When E2 reaches its maximum concentration, a switch in production of the pituitary occurs, causing the LH peak (World Health Organization, 1980). This LH peak is responsible for the completion of oocyte maturation (meiosis) which ultimately culminates in ovulation. The oocyte, once entering a permeable fallopian tube, is then ready to be fertilized by sperm (Halbert, et al., 1976). In the meantime, the ruptured follicle transforms into a corpus luteum (CL), increasing the production of the chief hormone of the luteal phase, P, which is responsible for the endometrial secretory changes essential for embryo implantation. An implanting embryo will produce increasing amounts of human chorionic gonadotropin (hCG), which can interact via the same receptor LH binds to (the LH/hCG receptor) and regulate CL function during the initial stages of pregnancy. In the absence of circulating hCG (*i.e.* an evolving implanted embryo), the CL will eventually become atretic, leading to the cessation of the production of P and endometrial shedding.

### 4.1 The prevalence and burden of infertility

Parenthood will eventually become a lifetime goal for up to 95% of all stable reproductive age couples (Stobel-Richter, et al., 2005, Lampic, et al., 2006). Overall, approximately 9% of these couples are currently suffering from infertility. Additionally, for up to 26% of all prospective parents, their desire to conceive at some point will ultimately be met with frustration and failure at least once during their lifespan (Boivin, et al., 2007).

Beyond childlessness, infertility leads to a wider, yet commonly overlooked set of burden, which includes depression (Klemetti, et al., 2010), sexual dysfunction (Nelson, et al., 2008), treatment-related distress and iatrogenesis (Verhaak, et al., 2007), stigmatization (Mascarenhas, et al., 2012), decrease in quality of life (Luk and Loke, 2015), divorce (Dyer, et al., 2002) and even a potential overall increased risk of cancer and death (Ehrlich, 2015). The World Health Organization (2006) has exposed the ever-growing financial and societal repercussions of this condition since early on, namely by declaring it a global health issue and actively promoting its diagnosis and treatment worldwide. At a time where already as many as 186 million people suffer from infertility (Inhorn and Patrizio, 2015), the increase in delay in age of child-wish (Klein and Sauer, 2001) and progressive decrease in fecundity (Levine, et al., 2017) has led health care providers to become
increasingly concerned with the unpredictable increase of this burden on the overall society (Connolly, et al., 2010, Chambers, et al., 2013, Ehrlich, 2015). This factor is particularly worrisome when one accounts for the fact that, ultimately, only 51-56% will actually seek medical care (Boivin, et al., 2007).

4.2 In-vitro fertilization for infertility: benchmarks of success

In the context of ART, in-vitro fertilization (IVF) involves the in-vitro handling of both human oocytes and sperm to ensue fertilization and development of viable embryos which eventually be transferred into the uterus (Zegers-Hochschild, et al., 2017). In essence, it is constituted by three critical steps: the retrieval of both male and female gametes (the latter are obtained following ovarian stimulation, as explained below) for in-vitro manipulation, IVF and, lastly, an embryo transfer (and eventual cryopreservation of supernumerary embryos).

Whilst originally developed to obviate the need for a functional fallopian tube in women with tubal factor infertility (Steptoe and Edwards, 1978), the application of IVF has ever since been widely extended to several other cases of infertility, resulting in the birth of over 5 million babies worldwide since its genesis, half of which were born after 2007 (Adamson, et al., 2013), representing up to 4% of childbirths annually in Europe (Ferraretti, et al., 2012). In cases of male factor infertility, intracytoplasmic sperm injection (ICSI) – an extension of IVF – in which a single spermatozoon is injected into the oocyte cytoplasm, is also frequently performed (Palermo, et al., 1992).

In 2001, the European Society of Human Reproduction and Embryology (2001) estimated that approximately 1500 couples per million will eventually require one or more cycles of IVF per annum to conceive. This estimate, however, has since been considered too conservative, in light of the increasing awareness of both the diagnosis of infertility and the available treatments alternatives (Connolly, et al., 2010). Furthermore, this level of treatment availability is largely unmet, namely in the United States (US) and United Kingdom (UK), where, despite a 4% to 5% annual increase in the use of ART (Calhaz-Jorge, et al., 2016), only 40% of the demand of infertility treatments are actually met (Chambers, et al., 2009), with couples needing to spend up to 50% of their income to perform such treatments (Connolly, et al., 2010). That said, a more accurate inference might be that approximately 25% to 30% of all infertile couples will eventually require IVF (Collins and

The most obvious expected outcome of ART for a couple is the delivery of a child and, when visiting a fertility centre, patients seek counselling on their likelihood of achieving a live birth after a single treatment of IVF (Min, 2004). Although such an inquiry seems fair, multiple limitations in the reporting of ART outcomes and benchmarking of treatment success render this request “easier said than done”, with multiple authors proposing a large number of discrepant ideal benchmarks (Schieve and Reynolds, 2004, Abdalla, et al., 2010, Maheshwari, et al., 2015, Davies, et al., 2017, Wilkinson, et al., 2017). Furthermore, there is also a strong controversy around the question of the ideal way of reporting outcomes in IVF following suggestions of how results can be presented in a more favourable light by manipulating their presentation (Abdalla, et al., 2010).

Both effectiveness (i.e. pregnancy rates) and safety (i.e. incidence of ovarian hyperstimulation syndrome) key performance indicators (KPI) of ART are usually
presented as indexes with multiple possible numerators and denominators (Schieve and Reynolds, 2004, Wilkinson, et al., 2017). In the numerator, normally the last of all outcomes is preferred. For the effectiveness of an embryo transfer (ET) this would naturally be the live birth of at least 1 child (Figure 3, p. 26). However, since live birth delivery distances many months from the treatment performed and frequently occurs in a different setting, loss to follow-up is common as it mostly depends on the willingness of others to provide information to the fertility centres and competent authorities, making it difficult to adequately report overall cycle outcomes. For this reason, other less-preferable outcomes (i.e. positive hCG, clinical and ongoing pregnancy rates) are frequently preferred (or the only available option) – in spite of their artificial overestimation of IVF success.

For the denominator, the ideal scenario for both effectiveness and safety would be to evaluate all outcomes per started cycle (Schieve and Reynolds, 2004), with the obvious exception when the matter of study cannot be assessed from the start – e.g. in a study evaluating the likelihood of each individual oocyte to result in a live-birth, only women who completed ovarian stimulation and performed an oocyte retrieval can be included (De Geyter, et al., 2016).

Women with supernumerary embryos who did not achieve a live birth after their first ET, may try consecutive frozen embryo transfers (FETs) as well (explained below). Given the added benefit of this approach, many authors consider that, whenever feasible, follow-up should be continued further as cumulative live birth rates (CLBR, in which all ETs performed using embryos derived from a single cycle until a live birth delivery or exhaustion of all available embryos are to be considered). A classic example that demonstrates the value of this approach is the case study of the relationship between the number of oocytes retrieved and conception. In summary, while having more than 15 oocytes provides no additional benefit for the chances of conception after the first ET (Sunkara, et al., 2011), each extra oocyte is beneficial for the cumulative chance of conceiving (Drakopoulos, et al., 2016). For this reason, this approach is considerably more interesting for patients (Schieve and Reynolds, 2004, Wilkinson, et al., 2017), although it does bring other challenges to the benchmarking process: the understanding of how to best deal with women who abandon treatment prior to a live birth and without exhausting all available embryos.
4.2.1 Early pregnancy complications associated with in-vitro fertilization

Regarding the safety of IVF, relevant complications can be either iatrogenic events specifically related to ART (e.g. ovarian hyperstimulation syndrome, ovarian torsion, pelvic infection, pelvic injury, hemorrhage) (Roest, et al., 1996, El-Shawarby, et al., 2004, Maxwell, et al., 2008), or early pregnancy (early miscarriage, ectopic pregnancy and monozygotic twinning), maternal (e.g. gestational diabetes, pre-eclampsia) or neonatal complications (e.g. multiple pregnancy, congenital malformations, prematurity, low birth weight, late miscarriage or perinatal death). These complications, although not exclusive to ART, are more common following it (Kallen, et al., 2002, Olivennes, et al., 2002, Pelkonen, et al., 2010, Kalra, et al., 2011, Maheshwari, et al., 2012, Pandey, et al., 2012, Pinborg, et al., 2013, Wennerholm, et al., 2013, Hansen and Bower, 2014, Ishihara, et al., 2014, Mainigi, et al., 2014, Peeraer, et al., 2014, Pelkonen, et al., 2014, Weinerman and Mainigi, 2014). Amongst these, two specific early pregnancy complications, while being intrinsically related to ART and ER during IVF, are exceedingly difficult to study given their rare occurrence: ectopic pregnancy (EP) and monozygotic twinning (MZT).

4.2.1.1 Ectopic pregnancy

EP, which occurs in 2% of all spontaneous conceptions (Farquhar, 2005), is the most common cause of mortality during the first trimester of pregnancy in the UK (Cantwell, et al., 2011) and is responsible for up to 7% of all pregnancy-related deaths in the US (Grimes, 2006).

Anecdotally, the first successful human IVF treatment resulted in an EP (Steptoe and Edwards, 1976). Ever since then, ART have been consistently associated with increased EP rates, reaching as high as 8.6% (Clayton, et al., 2006). Although the reasons behind this increased incidence of EP following ART are not completely understood (Shaw, et al., 2010), several studies have demonstrated that this association may be related to the confounding effect of other known infertility risk factors (Strandell, et al., 1999). Such risk factors, particularly tubal disease, smoking and advanced maternal age, are both more prevalent in infertile women (Strandell, et al., 1999) and also described as risk factors for EP (Farquhar, 2005).

Confounding, however, does not account for the overall increased risk of EP that is associated with ART, suggesting that other contributing factors may still exist. In an attempt to provide a reasonable explanation for this higher incidence, investigators have
assessed the potential role of several other possible mechanisms. Specifically, the increased number of embryos transferred (Clayton, et al., 2006), a change in the expression of the adhesion protein e-cadherin in embryos derived in-vitro (Revel, et al., 2008) and a detrimental effect of the supraphysiologic sex steroid levels during ovarian stimulation on both endometrial receptivity (Killick, 2007) and tubal ciliary function (Paltieli, et al., 2000) have all been proposed as plausible causes for the increased EP rate following ART. Nonetheless, most of these studies have been based on small cohorts with a very small number of events (i.e. EPs), given the relatively low prevalence of EP among all pregnancies derived from ART.

Over the last 20 years, ART has changed significantly with a tendency towards fewer embryos being transferred (Thurin, et al., 2004), extended embryo culture (Papanikolaou, et al., 2006), milder ovarian stimulation (Heijnen, et al., 2007) and improved cryopreservation techniques (Cobo, et al., 2012). In addition, nationwide campaigns have successfully managed to reduce the incidence of known risk factors for EP such as smoking (Richardson, et al., 2014) or chlamydial infection and chlamydia-associated tubal disease (French, et al., 2011). A recent study attempted to evaluate whether these changes, implemented to both the ART procedure itself and the lifestyle of women in reproductive age, had any impact on the overall incidence of EP in the US (Perkins, et al., 2015). The authors concluded that EP rates decreased between 2001 and 2011 (from 2.0% to 1.6%) and that factors such as the transfer of multiple embryos seemed to have been associated with an increased risk of EP during this period. Nonetheless, since treatment policies and pregnancy outcomes are known to vary significantly between the US and Europe (Gleicher, et al., 2007, Baker, et al., 2010), these results may not be directly generalizable to the European situation.

### 4.2.1.2 Monozygotic twinning

Although the risk factors and mechanisms responsible for embryo splitting are still unclear, it has been argued that specific patient characteristics and ART procedures might play a role in the development of MZT. These reports warrant further exploration into this matter, since monozygotic pregnancies are associated with an increased risk of maternal and fetal complications (Rao, et al., 2004) such as fetal growth restriction, preterm delivery and perinatal mortality (Aston, et al., 2008). Specifically, previous researchers have pointed out that oocyte age (Knopman, et al., 2014), assisted hatching

4.3 How may in-vitro fertilization affect endometrial receptivity?

While multiple male gametes are directly obtainable in most cases by retrieving semen samples following masturbation, the retrieval of multiple oocytes from a given women requires a more complicated process that generally entails the administration of supraphysiologic doses of gonadotropins and surgical retrieval using transvaginal ultrasound guidance. Another essential difference between the male and female reproductive systems involves sheer numbers: while fertile males continuously produce millions of spermatozoa daily, fertile females are born with a limited number of oocytes which are subject to cyclic waves of recruitment, dominance and atresia until a critical threshold of depletion ensues the end of the reproductive age and – ultimately – menopause (Fritz and Speroff, 2012).

Controlled ovarian stimulation (COS) – i.e. the administration of supraphysiologic doses of exogenous gonadotropins – is historically regarded as a milestone for the treatment of infertility (Lunenfeld, et al., 2004), as it allowed physicians to bypass the natural phenomena of follicular dominance and atresia by artificially overcoming the FSH threshold and inducing multi-follicular (instead of mono-follicular) development (Macklon, et al., 2006). Following the initial introduction of human menopausal gonadotropins by Wortham, et al. (1983), the development in the field has progressed immensely, introducing more patient-friendly and less side-effect prone exogenous gonadotropins alternatives, namely highly purified human menopausal gonadotropin (HP-hMG) and recombinant FSH (rFSH), which are, collectively, the most-commonly used exogenous gonadotropins nowadays. Despite fervid debate, clinical trials and meta-analyses thus far have failed to demonstrate clear superiority of one gonadotropin formulation over the other (van Wely, et al., 2003, Andersen, et al., 2006, Devroey, et al., 2009). However, even without knowing which drug is the overall best, the generalized use
of exogenous gonadotropins during IVF has led to a substantial increase in ART pregnancy rates, from the typical 3-10% range (using no or minimal stimulation) to 20-50% (Barbieri and Hornstein, 1999).

While there is a clear advantage in using COS in routine IVF, one must be aware that its administration requires a comprehensive knowledge of how the ovaries react to hyperstimulation, as an unsuspecting user may cause the occurrence of both premature luteinization and ovarian hyperstimulation syndrome (OHSS). Specifically regarding the former, as the presence of unopposed supraphysiologic circulating E2 derived from multifollicular development may cause an unpredictable LH surge and premature luteinization (Stanger and Yovich, 1985), a strategy to avoid such is traditionally required.

Parallel to the administration of exogenous gonadotropins for ovarian hyperstimulation, other researchers were evaluating the use of exogenous gonadotropin-releasing hormone (GnRH) agonists in women with hypogonadotropic hypogonadal anovulation (Leyendecker, et al., 1980, Schoemaker, et al., 1981). However, these investigators noted that such treatment was only possible if the GnRH agonist was administered in a pulsatile manner, since a continuous administration would cause GnRH receptor depletion with subsequent pituitary desensitization and abolition of LH and FSH production. What was initially perceived as a failure in treatment for hypogonadotropic hypogonadal women has since then been cherished as an important enhancer of multi-follicular recruitment for IVF (Fleming, et al., 1982), being later developed into what later became the long-agonist GnRH protocol (Porter, et al., 1984, Fleming and Coutts, 1986), a mainstream of COS cotreatment for at least the first 20 years of IVF (Macklon, et al., 2006). By adding GnRH analogues to COS, physicians were able to reduce cycle cancellation rates by 67%
while increasing clinical pregnancy rates by 70% (Hughes, et al., 1992). Furthermore, since the year 2000, a new approach has been developed with the advent of GnRH antagonists (Albano, et al., 2000, Borm and Mannaerts, 2000). Conversely to GnRH agonists, GnRH antagonists cause immediate pituitary suppression, obviating the need for a more prolonged and cumbersome administration of GnRH analogues (Figure 4, p. 31). This more patient-friendly approach has shown to have comparable pregnancy outcomes, leading many to recommend that it become the new “standard” in IVF (Macklon, et al., 2006, Al-Inany, et al., 2011, Toftager, et al., 2017, Wang, et al., 2017).

4.3.1 Circulating sexual hormones and endometrial receptivity

ART have developed vastly since the first live birth following IVF in 1978. This multidirectional improvement led to the optimisation of ovarian stimulation and to a better assessment of embryo quality, ultimately causing a steady increase of live birth delivery rates until the late 1990s (Diedrich, et al., 2007). Despite that, LBRs have remained relatively low and, since the year 2000, rather stagnant (Andersen, et al., 2005, Nyboe Andersen, et al., 2009).

Monitoring of both the endometrial and ovarian responses to COS with transvaginal ultrasound has become an important predictor of ART success (McWilliams and Frattarelli, 2007). In addition, while not being unanimously applied (Murad, 1998, Vandekerckhove, et al., 2014), many agree that a concomitant hormonal assessment may also be of benefit to predict ART outcome (Hardiman, et al., 1990, Rizk and Smitz, 1992, Loumaye, et al., 1997).

While the duration of the follicular phase of the menstrual cycle may be subject to both inter-cycle and inter-women variability, endogenous sexual hormones remain within relatively low levels prior to ovulation (Figure 5, p. 33), specifically a median (5th and 95th percentiles) of 182.80 pg/mL (131.30 pg/mL – 388.28 pg/mL) for E2 and 0.80 ng/mL (0.39 ng/mL – 1.30 ng/mL) for P (Stricker, et al., 2006). Meanwhile, these same hormones reach substantially higher levels with COS, namely E2, which may raise 10- or even 20-fold (Pittaway and Wentz, 1983, Wang, et al., 2017). These supraphysiologic hormones concentrations present during COS have been strongly associated with a negative effect on endometrial receptivity (ER) (Devroey, et al., 2004, Horcøjadas, et al., 2007), with the exceeding majority of biopsies performed on the day of oocyte retrieval demonstrating both histologic and gene expression patterns suggestive of endometrial advancement
Studies in oocytes donors have also shown that COS may accelerate the appearance of endometrial pinopods and nucleolar channel systems (Kolb, et al., 1997, Nikas, et al., 1999, Zapantis, et al., 2013) – ultrastructural formations which are closely related to ER and the window of implantation (WOI) (Nikas, et al., 1995) – which promote abnormal intracavity interleukin-1 (IL-1) expression (Simon, et al., 1996), and cause abnormal placentation and embryonic growth (Mainigi, et al., 2014).

4.3.1.1 Estradiol

E2 plays an important role in the preparation of the uterus for implantation (Lutjen, et al., 1984, Navot, et al., 1986). Its key functions involve the regulation of endometrial proliferation and vascular perfusion during the follicular phase (Joo, et al., 2010), both of which may be assessed indirectly during COS via transvaginal ultrasound, namely by the measurement of endometrial thickness (Kasius, et al., 2014) and doppler imaging techniques (Yang, et al., 1999).

Although the crucial role of this hormone is concentrated in the follicular phase, E2 continues to be produced during the luteal phase as well, where it may still influence endometrial function according to its circulating concentration. While most of the clinical

Figure 5 – Normal hormone values throughout the menstrual cycle

Solid lines represent median values; dotted lines depict the 5th and 95th percentiles. Reprinted from Stricker, et al. (2006) with permission.
data in humans supports the notion that E2 in the luteal phase has a non-regulatory but merely facilitating role (de Ziegler, 1995, Ghosh and Sengupta, 1995, Friedler, et al., 2005, Moini, et al., 2011, Huang, et al., 2015), translational evidence from the early 2000s showed that this hormone may play an important role in determining the beginning and duration of the WOI in the animal model (Ma, et al., 2003). Specifically, these researchers administered increasing doses and durations of E2 to oophorectomized mice with artificially stimulated luteal phases followed by an embryo transfer to evaluate ER in each experimental model. What they concluded was that luteinized endometria required E2 levels to be within a certain interval for the adequate opening and closing of the WOI (Figure 6, p. 34). The authors further noticed that the refractory changes associated with excessive E2 supplementation could be due to an abnormal endometrial expression of Lif, Ptgs2 and Hegfl induced by a differential expression of progesterone and estrogen receptors. Studies in oocyte donors have also shown that increasing luteal phase E2 concentrations may have a direct toxic effect on the embryo itself, hindering adhesion (Valbuena, et al., 2001). Furthermore, in-vitro proteomic studies have also shown an E2-dependent differential profile in supraphysiologic versus physiologic E2-treated Ishikawa cells (Ullah, et al., 2017), although the clinical relevance of such results require further investigation.

These results raised an interesting clinical hypothesis, specifically if one reduces late-follicular E2 concentrations, will that limit the “carryover” activity of E2 into the luteal phase and, consequentially, enhance ER? Such a hypothesis was indirectly evaluated in a
prospective trial by Simon, et al. (1998) in the human. In women with a previously-failed IVF attempt with at least 3 good-quality embryos and hCG-trigger-day E2 levels above 3000 pg/mL, the researchers randomized the subjects to either a standard fixed-dose or step-down COS protocol in a subsequent IVF cycle. What the investigators noticed was a decrease in the circulating levels of E2 during the preimplantation period in the step-down group, which was also accompanied by an increase in pregnancy rates (64.2% versus 24.2%). While the use of step-down protocols is common in clinical practice, this was, to our knowledge, the only instance in which it was used with the deliberate intent to reduce the overall dose of exogenous gonadotropin and curtail the carry-over of E2 from the follicular phase to the luteal phase. However, a small pilot study with 10 patients in each arm revealed similar benefits of reducing circulating E2 during IVF using a different approach – specifically, the co-administration of aromatase inhibitors during COS. Following their trial, these investigators found that the co-administration of letrozole resulted in both an increase in endometrial thickness (EMT, 10.3 mm versus 8.1 mm) and a trend towards better clinical pregnancy rates (CPR, 5 versus 2 clinical pregnancies), albeit severely limited by the sample size (Verpoest, et al., 2006). However, subsequent larger retrospective (Ecemis, et al., 2016) and prospective (Lossl, et al., 2008, Yasa, et al., 2013) studies failed to show any benefit in using aromatase inhibitors and its co-administration has progressively fallen into disuse for normal/high responders ever since (Papanikolaou, et al., 2011, Garcia-Velasco, 2012).

These studies collectively transpire what was considered to be the chief concern of supraphysiologic E2 levels during COS: that these excessive E2 concentrations may subsist throughout the luteal phase and affect implantation. However, more recent work in the animal model have also postulated that supraphysiologic E2 may affect ART outcome already in the proliferative phase, specifically via the production of detrimental intracavity fluid (Akman, et al., 2005) mediated by the E2-dependent water channel genes aquaporin-5 and aquaporin-8 (Zhang, et al., 2015). Furthermore, as E2 levels usually decline drastically (approximately 95%) following ovulation triggering in COS (Friedler, et al., 2005), it seemed interesting to evaluate the direct relationship between excessive late-follicular E2 (i.e. on the day of ovulation triggering) and embryo implantation. In an attempt to translate this basic scientific evidence into clinical practice, the effect of late-follicular E2 on pregnancy rates has been fairly comprehensively assessed – albeit, mostly in retrospective studies – since the genesis of IVF. In 2004, Kosmas, et al. (2004) noted the disagreement and lack of strong evidence at the time, in which studies were either failing to relate IVF
pregnancy to circulating E2 (Mettler and Tavmergen, 1989, Dor, et al., 1992, Sharara and McClamrock, 1999, Levi, et al., 2001, Papageorgiou, et al., 2002, Chen, et al., 2003) or associating it with both a hindered (Simon, et al., 1995, Kolibianakis, et al., 2003) or even enhanced (Chenette, et al., 1990, Gelety and Buyalos, 1995) change of pregnancy. These discrepancies in results were originally considered to be due to methodological issues among the studies, namely the frequent failure to adequately account for ovarian response and late-follicular P levels. That said, most of the subsequent studies [with the notable exception of a small retrospective cohort including 455 cycles (Joo, et al., 2010)] overwhelmingly failed to find any strong association between late-follicular circulating E2 and IVF pregnancy rates (Wu, et al., 2007, Wu, et al., 2012, Erzincan, et al., 2014, Taskin, et al., 2014, Zavy, et al., 2014, Singh, et al., 2015, Siddhartha, et al., 2016, Wang, et al., 2017), including a prospective observational trial comprising 207 women (Kyrou, et al., 2009). Furthermore, a subsequent meta-analysis failed to show any benefit of monitoring E2 during IVF in terms of clinical pregnancy, oocyte retrieval, cycle cancellation or OHSS rates (Kwan, et al., 2008). While indeed some researchers still consider that the inconsistency in results may be related to an incorrect methodological evaluation of the dynamics of E2 during COS (Kim, et al., 2010), the accumulation of negative studies has effectively deviated the attention on the issue of hindered ER away from E2 and refocused it towards another hormone – P.

In terms of neonatal outcomes, E2 serum levels above 2500-3500 pg/mL have being previously shown to be moderately predictive (20.8% versus 8.5%; area under the curve: 0.735) for pregnancy complications related to abnormal placentation (Farhi, et al., 2010, Royster, et al., 2016), including pregnancy-induced hypertension/pre-eclampsia (Imudia, et al., 2012), small-for-gestational-age fetuses (Imudia, et al., 2012) and low birth weight (Pereira, et al., 2017). Others, however, have failed to show this effect (Wang, et al., 2017).

### 4.3.1.2 Progesterone

Following adequate priming by E2, P secreted by a recently luteinized follicle(s) will regulate an intricate sequence of molecular events in preparation for implantation (Sonigo, et al., 2014, Lawrenz and Fatemi, 2017). After ovulation, P limits any further proliferation of the endometrium over the course of the next three days by directly restricting the function of E2 and the expression of E2-binding receptors, while stimulating further the development of both vascular and glandular structures
(Tabibzadeh, 1990, Kirkland, et al., 1992). This selective growth of the vascular and glandular components within a confined structure causes coiling of the spiral vessels and increases glandular tortuosity (Fritz and Speroff, 2012). At the moment of implantation, the endometrium has become differentiated into two distinct layers (Lawrenz and Fatemi, 2017): the basalis (most adjacent to the myometrium, a layer which remains relatively unchanged throughout the menstrual cycle and is critical for endometrial regeneration following the cyclical sloughing events endured during menstruation) and the functionalis (constituted by the stratum spongiosum and an overlaying stratum compactum). These successive morphological changes which P sets into motion follow a predictable day-by-day sequence of development which were the basis of the histological endometrial dating classification system developed by Noyes, et al. (1975). This classic endometrial classification system remained, until recently (Diaz-Gimeno, et al., 2013), an unchallenged gold standard for the evaluation of luteal function, where endometrial samples with a developmental stage two-days apart from the expected were considered to be “out of phase” (Wentz, 1980).

In a natural cycle, circulating P remains low throughout the follicular phase until ovulation, specifically at a median of 0.30 ng/mL (0.19 ng/mL – 0.51 ng/mL; Figure 5, p. 33). Prior to ovulation, the main source of P shifts from the adrenal cortex to the ovary (De Geyter, et al., 2002) and, at that same time, P receptor mRNA and proteins become more abundant in the nuclei of the granulosa cells (Hild-Petito, et al., 1988). The LH peak causes a slight but noticeable physiologic late-follicular phase P increase (De Geyter, et al., 2002) which, besides contributing for the timing of ovulation (Couzinet, et al., 1992), may be essential for follicular development (Collins and Hodgen, 1986, Chaffkin, et al., 1992, Bridges, et al., 2006, Stouffer, et al., 2007). Animal experiments have shown that blocking mid-cycle P production is detrimental for oocyte maturation (Borman, et al., 2004), oocyte fertilization competence (Zelinski-Wooten, et al., 1994) and granulosa/theca luteinisation (Hibbert, et al., 1996). Specifically, studies in rodents have shown that the administration of antiprogestins in the periovulatory phase did not allow LH-induced luteinisation of granulosa cells (Natraj and Richards, 1993) and that only granulosa cells with P receptors in the pre-ovulatory phase where capable of luteinizing (Hild-Petito, et al., 1988). Meanwhile, in the endometrium, P receptors reach their maximal concentrations during the mid to late proliferative phase of the menstrual cycle (Barile, et al., 1979, Lessey, et al., 1988, Ravn, et al., 1994) and P directly affects the action of E2 on the endometrium (Ozcakir, et al., 2004). Studies where the anti-progestin RU-486 was administered during final
oocyte maturation showed that it affected endometrial receptivity in mice (Batten, et al., 1988).

While indeed P does seem to play a role in regulation of the last stages of the follicular/proliferative phases, the main concerns around its potential detrimental effect during IVF is that COS may cause an overproduction of P and, thus, alter ER of a fresh ET cycle and, eventually, embryo quality (Venetis, et al., 2013). COS-derived P overproduction has traditionally been considered to be a consequence of CYP17-produced P in the theca cells of the multiple developing follicles (Andersen and Ezcurra, 2014). This P produced in the follicular phase is usually metabolized downstream by the LH-activity in the later stages of follicular development, as LH promotes the conversion of P into androgens which will be then transported and converted into E2 by the granulosa cells (Fleming and Jenkins, 2010, Andersen and Ezcurra, 2014). However, recent in vitro studies have shown that, during COS, P may be produced directly in a dose-dependent manner by FSH activity (Figure 7, p. 38) in the granulosa cells (Oktem, et al., 2017).

**Figure 7 - Proposed mechanism of FSH-dependent P production during COS**

According to the two-cell two gonadotropins theory, P produced by the granulosa cells enter the theca cells to be converted into androgens since CYP17A1 enzymes (cytochrome P450 enzymes 17α-OH and 17,20-lyase) are not responsive to FSH, therefore no further reactions take place downstream in the granulosa cells. When this effect is combined with the direct stimulatory effect of FSH on 3β-HSD and P biosynthesis, accumulated P may leak into systemic circulation. Reprinted from Oktem, et al. (2017) with permission.

endometrial genomic/epigenetic expression in the luteal phase (Labarta, et al., 2011, Van Vaerenbergh, et al., 2011, Xiong, et al., 2017). This has led many centers to change their clinical practice and to measure serum P levels on the day of ovulation triggering and to adopt a freeze-all strategy when the threshold of 1.50 ng/mL is exceeded.

Nonetheless, the everyday use of LFEP as an ART outcome predictor in current clinical practice has proven to be harder than originally expected, owing mostly to the fact that a) P production may not affect pregnancy outcomes linearly (Bosch, et al., 2010, Xu, et al., 2012, Santos-Ribeiro, et al., 2014) and b) it is frequently encountered in good-prognosis women with an otherwise healthy multi-follicular response (Fleming, 2008, Younis, 2011). Furthermore, previous studies have recommended caution in the use of LFEP as an indication for all-embryo cryopreservation in high-responder women (Griesinger, et al., 2013) or the adaptation of the cut-off of LFEP according to ovarian response (Xu, et al., 2012). Meanwhile, others have postulated that LFEP may only affect cleavage stage embryo transfers (Papanikolaou, et al., 2009), a hypothesis that has been refuted more recently by others (Bosch, et al., 2010, Corti, et al., 2013, Yang, et al., 2015). Finally, the detrimental effect of circulating P may be set into motion even when P is below 1.50 ng/mL (Roque, et al., 2015). Altogether, these conflicting studies are a testament of the limitations of using a single P assessment during COS to decide whether to defer embryo transfer or not. Many investigators have attempted to enhance the predictive capacity of this single measurement by using ratios such as P-to-follicle (Shufaro, et al., 2015), P-to-oocyte (Hill, et al., 2017), P-to-estradiol (Lai, et al., 2009, Wu, et al., 2012) ratios. However, such indexes have failed to perform superiorly, thus limiting their routine clinical practice, an observation that has led some to propose that clinicians consider disregarding endocrine monitoring of P during COS altogether (Martinez, et al., 2016).

Interestingly, Huang, et al. (2012) concluded in their retrospective analysis that the duration of LFEP (defined as a universal P threshold of >1.0 ng/mL during 0, 1-2 or ≥3 days prior to ovulation triggering) seemed to play a major role in how circulating P affected pregnancy outcome. This approach is unique in the sense that it was an interesting departure from the previous literature, deviating focus from the question “how high P can reach?” to “how long can the endometrium be exposed to P above a certain universal threshold?”. As P receptors are present in endometrium in all phases of the menstrual cycle, namely during menses and the early follicular phase (Ingamells, et al., 1996, Lau, et al., 1996, Marshburn, et al., 2005), it would be reasonable to assume that P exposure might be detrimental even when occurring days before ovulation triggering. Furthermore, this
concept of a hindering cumulative P exposure is more in line with the current knowledge of how P regulates endometrial development, as it is generally thought that the ovaries need to increase (Fransasiak, et al., 2016) and maintain (Coomarasamy, et al., 2015) the secretion of P above a certain level in order to initiate and sustain the development of a secretory endometrium. However, whether the threshold to trigger these endometrial changes is universal (e.g. 1.50 ng/mL, regardless of the patient’s baseline P exposure) or relative (i.e. a difference between the amount of P produced and the usual baseline circulating P concentration encountered in each patient) is unknown.

While being a potentially useful hypothesis (Venetis, et al., 2013, Sonigo, et al., 2014), the approach performed by Huang, et al. (2012) has been assessed only using CPR as the main outcome measure and has been subject to insufficient validation (Sonigo, et al., 2014, Demir, et al., 2016). Furthermore, failures to completely reproduce their results (Dai, et al., 2015) allude to the possibility that the effect of P exposure during COS on ART pregnancy outcomes may be incompletely captured when applying a universal threshold. Specifically, one of the reasons for these conflicting results may stem from the fact that this study elected a universal arbitrary cut-off of 1.00 ng/mL, which is frequently found at the end of the follicular phase of unstimulated cycles as well (Stricker, et al., 2006). Conversely, one could also hypothesize that evaluating how much P varied relatively to the basal circulating P levels of that particular patient (instead of whether it surpassed a certain universal cut-off or not) could assist physicians in predicting better the detrimental effect of P exposure during COS.

### 4.3.2 Ovulation triggering

Prior to oocyte retrieval, it is necessary to artificially stimulate the final stages of oocyte maturation and ovulation. Traditionally, this has been performed by administering exogenous hCG (Humaidan, et al., 2011), a molecule historically obtained in large quantities by purification of the urine of pregnant women while serving the same purpose of LH via its common site to the LH/hCG receptor. Ovulation is triggered once a sufficient amount of large antral follicles are present. Various trigger criteria have been proposed, with and without considering additionally the serum concentration of E2 – however, none have shown to be superior in terms of the retrieval rate of matures oocytes (Olivennes, et al., 1998, Albano, et al., 2000, Felberbaum, et al., 2000, de Jong, et al., 2001, Fluker, et al., 2001, Garcia-Velasco, et al., 2001, Li, et al., 2014). Following a period of approximately 36 to
38 hours, the oocyte retrieval is performed, usually transvaginally under ultrasound guidance (Bosdou, et al., 2015, Garor, et al., 2015).

While the administration of hCG is essential for the final stages of oocyte competence and to trigger luteinization, some investigators have alluded to the possible relationship between this drug and changes in ER (Kolibianakis, et al., 2002). Specifically, a recent immunohistochemistry and flow cytometry analyses performed on endometrial biopsies deriving from 34 women with recurrent implantation failure postulated that hCG may deregulate T-cell activity (Diao, et al., 2017).

Another major concern around the use of hCG for ovulation triggering is its association with the most notable iatrogenic event in ART: OHSS, which is almost exclusively an ovarian-stimulation-related iatrogenic event that occurs in women who are frequently otherwise healthy (Papanikolaou, et al., 2006). While its general incidence is approximately 2% to 3% per cycle, OHSS can occur in up to a third of all cases of high-risk patients (Acevedo, et al., 2006, Engmann, et al., 2008), namely those with a previous history of OHSS or polycystic ovaries. In its most severe forms, this syndrome has the potential to cause serious morbidity or mortality, mainly due to the increased occurrence of ovarian torsion and thromboembolism (Delvigne and Rozenberg, 2003).

OHSS is an exaggerated response to COS characterized by cystic enlargement of the ovaries, abdominal distention and pain, fluid shift from the intravascular space to the third space, which can result in ascites, pericardial and pleural effusions, and even generalized edema. Although this syndrome has been extensively studied, its pathophysiology remains rather elusive. Until now, there is sufficient evidence to believe that OHSS results from an excessive secretion of vasoactive substances during COS, namely VEGF and factors that derive from the renin–angiotensin system (Vloeberghs, et al., 2009). VEGF is produced by the granulosa cells after stimulation with gonadotropins and increases sharply after the administration of hCG due to a hypersensitivity to this latter hormone (Neulen, et al., 1995). This vasoactive substance over-secretion seems to be almost entirely dependent on the activity of LH, which is present only for a short period in the natural cycle. During IVF, however, final oocyte maturation and ovulation is frequently triggered with hCG, which, in comparison to LH, has a substantially longer half-life (Humaidan, et al., 2011). For this reason, many authors have proposed that women with a high-risk of developing OHSS should perform COS under GnRH antagonist suppression and then use a GnRH agonist for triggering instead (Acevedo, et al., 2006, Griesinger, et al., 2006, Garcia-
Velasco, 2012, Iliodromiti, et al., 2013). When administered, the GnRH agonist will cause the displacement of the GnRH antagonist from the pituitary receptors (Table 1, p. 43), resulting in the induction of a LH flare-up/"surge" that lasts only for approximately 24-36 hours in total (Humaidan, et al., 2011). Although this approach has effectively reduced the incidence of OHSS (Iliodromiti, et al., 2013), it has not eliminated the risk completely (Griesinger, et al., 2011, Fatemi, et al., 2014, Gurbuz, et al., 2014, Ling, et al., 2014, Santos-Ribeiro, et al., 2015). GnRH agonist triggering has been increasingly acknowledged as a worthy strategy to minimize the risk of OHSS (Devroey, et al., 2011) and is currently applied more broadly, namely in cycles with abnormal late-follicular P levels (Roque, et al., 2015) or in oocyte donation programmes (Stoop, et al., 2012). However, while seeming equally efficient in terms of oocyte competence (Herrero, et al., 2011), the generalized use of GnRH agonist triggering has remained thus far limited by the fact that this approach seems to cause an artificially shortened luteal phase with abrupt luteolysis which significantly reduces IVF pregnancy outcomes (Griesinger, et al., 2007, Humaidan, et al., 2010). In light of the reduced CPR associated with GnRH agonist triggering, several possible strategies have been proposed. The first strategy is intensified luteal phase support with either an alternative supplementation of P/E2 (Engmann, et al., 2008) or a low-dose administration of hCG immediately after oocyte retrieval (Humaidan, 2012). This approach seems to significantly increase pregnancy rates after GnRH triggering but may come potentially at the cost of once more increasing the risk of OHSS (Iliodromiti, et al., 2013, Seyhan, et al., 2013, Youssef, et al., 2014). The second strategy (i.e. freeze-all strategy) is to electively cryopreserve all embryos and then replace them in a subsequent artificially-supported cycle (Griesinger, et al., 2007, Devroey, et al., 2011), which, until now, has been shown to be the method most effective at reducing the occurrence of severe OHSS (Bodri, et al., 2010). Although both approaches seem reasonable, there is no consensus on which is the most adequate since no comparative trial has ever been performed (Evers, 2013).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>hCG</th>
<th>GnRH agonist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mechanism</strong></td>
<td>Biologically similar to LH</td>
<td>GnRH antagonist displacement</td>
</tr>
<tr>
<td><strong>FSH surge</strong></td>
<td>No</td>
<td>Yes (flare-up)</td>
</tr>
<tr>
<td><strong>Luteotropic effect</strong></td>
<td>Sustained (up to 6 days after 5000 IU)</td>
<td>Shorter</td>
</tr>
<tr>
<td><strong>Luteal phase steroids</strong></td>
<td>High</td>
<td>Closer to physiologic ranges</td>
</tr>
</tbody>
</table>

Table 1 – Comparison between hCG and GnRH agonist for ovulation triggering
**4.3.3 Fresh embryo transfer and its predictors of success**

Following oocyte retrieval, the oocytes are inseminated either by IVF or ICSI (Van Landuyt, et al., 2005) and, after fertilization, cultured in vitro between 2 to 6 days prior to being transferred to the uterus at either a cleavage or blastocyst stage (De Vos, et al., 2016). During their time spent in the laboratory, the embryos are evaluated frequently in order to assess developmental quality. The aim of this quality assessment and extending embryo culture to day 5 or 6 (when performed) is to select, among the multiple embryos produced, those with increased probability of implantation, namely via a so-called *self-selection* process that might occur at the blastocyst stage (De Vos, et al., 2016, van de Vijver, et al., 2017). Good quality embryos that are not transferred in a fresh transfer attempt are usually cryopreserved for later use.

For a live birth to occur, a viable embryo that has been transferred must successfully implant in a receptive endometrium (Fatemi and Popovic-Todorovic, 2013). This intricate dialogue between the embryo and the endometrium is generally only possible during a four-day timeframe in the luteal phase – the WOI (Achache and Revel, 2006, Diedrich, et al., 2007, Cha, et al., 2012, Fatemi and Popovic-Todorovic, 2013). In a natural cycle, the WOI begins approximately 7 days after ovulation, at a time when the embryo is a blastocyst and at an optimal stage for implantation. This is a seamless example of how the ovaries, pituitary and endometrium are constantly regulating the function of one another to allow perfect synchronicity towards the development of a viable pregnancy.

Despite significant advancements in the fields of ultrasonography (Wang, et al., 2010, Singh, et al., 2011, Zhang, et al., 2016), immunology (Seshadri and Sunkara, 2014) and molecular diagnostics (Ruiz-Alonso, et al., 2013, Koot, et al., 2016) the potential benefit of these novel approaches to evaluate the endometrium are yet to be confirmed, leaving most physicians with only the measurement of EMT during COS to aid in the decision whether to perform a fresh embryo transfer or not. That said, there is also a continuous debate on the predictive value of the measurement of EMT prior to the administration of hCG for ovulation triggering in ART. While some authors have demonstrated a linear correlation between pregnancy rates and EMT (Rinaldi, et al., 1996, Richter, et al., 2007, Al-Ghamdi, et al., 2008, Chen, et al., 2010), others have posited that the chance to conceive may even decline above a thickness of 14 mm (Weissman, et al., 1999), after which miscarriage rates may increase. Moreover, in a retrospective study (Lamanna, et al., 2008) including 606 women undergoing a long-agonist protocol, the investigators noted a parabolic trend in
ART outcomes across EMT categories (with lower pregnancy rates in the EMT extremes below 8 mm and above 14 mm, respectively). In 2014, a meta-analysis of 22 studies concluded that the measurement of EMT was a valuable predictor for CPR, with lower CPR below the frequently mentioned cut-off of 7 mm, which progressively increased until 10 mm of EMT (Kasius, et al., 2014). These results could lead to the conclusion that a thick endometrium may not necessarily predict pregnancy but, conversely, a thin endometrium may be associated with lower chances of conception, possibly owing to a thinner functional layer that exposes the embryo to the higher oxygen concentrations present in the blood from the spiral arteries (Casper, 2011).


4.3.4 Embryo cryopreservation and the “freeze-all strategy”

The early 1980s saw the first cases of live births after the transfer of autologous frozen embryos being reported (Trounson and Mohr, 1983, Zeilmaker, et al., 1984). These preliminary reports were met with reasonable resistance by the scientific community owing to the limited efficacy of earlier cryopreservation methods and concerns regarding the overall safety of embryo cryopreservation. This uncertainty effectively relayed embryo cryopreservation at its genesis to the status of an “adjuvant method” for cycles in which
the number of embryos produced was deemed too excessive for simultaneous replacement during the fresh ET attempt. However, following the advent of more efficient cryopreservation strategies [i.e. vitrification (Balaban, et al., 2008)] and reassuring safety data (Belva, et al., 2008, Belva, et al., 2016), the use of embryo cryopreservation has progressively increased, currently accounting for up to one third of all children born after ART in the US (Doody, 2014). Furthermore, cryopreservation has now become an indispensable tool in everyday clinical practice, providing the necessary means to assure the safe storage of supernumerary embryos which increase CLBRs while minimizing the risks associated with the multiple pregnancies (Pandian, et al., 2013). For this reason, an increasing amount of scientific societies and governments have encouraged (Maheshwari, et al., 2011) or even enforced (Sundstrom and Saldeen, 2009, Peeraer, et al., 2014) elective single ET policies, progressively changing the benchmark of ART from pregnancy rates to Birth Emphasizing a Successful Singleton at Term [BESST, (Min, 2004)]. These considerations have set the stage for a new stance on embryo cryopreservation in modern-day medicine, which is no longer viewed as a simple adjuvant of fresh ET, especially in light of the growing use of the freeze-all strategy (Blockeel, et al., 2016).

A vast number of researchers have alluded to the potential benefit of distancing the transfer of embryos produced during ART away from the ovarian stimulation cycle. The scientific evidence which may justify this recent trend in the use of the so called “freeze-all strategy” includes randomized controlled trials (RCTs) demonstrating an increase in LBR after elective embryo cryopreservation (Chen, et al., 2016) and population studies associating frozen embryo transfers with a lower incidence of prematurity and/or low birthweight (Shih, et al., 2008, Pelkonen, et al., 2010, Pinborg, et al., 2010, Wang, et al., 2010, Maheshwari, et al., 2012, Belva, et al., 2016, Maheshwari, et al., 2016, Zhao, et al., 2016, Vidal, et al., 2017). Conversely, many have raised concerns regarding the increased risk of macrosomia after FET and the inconsistencies in the reporting of the studies evaluating perinatal/neonatal death, which have shown a full-spectrum of possible results, ranging from lower (Maheshwari, et al., 2012), unchanged (Pelkonen, et al., 2010, Vidal, et al., 2017) and increased (Pinborg, et al., 2010) risk. At a time where perinatal care accounts for a total burden which is well over 3.8-3.9 billion euros in developed countries such as the UK and Canada (Mangham, et al., 2009, Johnston, et al., 2014), the possible logistical and financial implications such findings may have has become the focus of much attention.

Physicians are commonly asked by their patients for how long COS may bear any carryover effect on a subsequent treatment (Reichman, et al., 2013) and FETs are
frequently postponed one or more menstrual cycles in an attempt to minimize any conceivable residual effect COS may have on ER (Maas, et al., 2008). However, the previous literature on this matter is rather scarce (Diamond, et al., 2012, Hu, et al., 2014). For this reason, while this empirical decision may be based on the best of intentions, the elective deferral of FETs may unnecessarily frustrate couples who wish to become pregnant as soon as possible.

4.4 Novel methods to predict endometrial receptivity

4.4.1 Molecular diagnostic tools

The knowledge that ovarian stimulation hinders ER has led to multiple efforts to adequately assess the endometrium prior to ET. These research groups stem from many scientific fields, including immunology, histology, endocrinology, microbiology, proteomics and genomics (Fatemi and Popovic-Todorovic, 2013). Amongst these, customized microarrays that analyse the transcriptomic signature of freshly biopsied secretory endometria have recently been developed (Diaz-Gimeno, et al., 2011). By analysing the endometrium’s expression profile, these microarrays can accurately discriminate between receptive and non-receptive uteri. Although this innovative approach has an enormous potential, its use as a decision-making tool during ART has been hampered thus far by two factors: a) criticism stating that the increased pregnancy rates following the diagnostic biopsy may be due to the effect of endometrial scratching, and b) a biopsy during the secretory phase induces endometrial injury which, although temporary, effectively precludes the transfer of an embryo during that same WOI period.

4.4.1.1 The controversy around endometrial scratching

The fate of the endometrium is decided according to the successful implantation of an embryo. In the absence of implantation, the CL will lack positive embryonic feedback and will progressively senesce, reducing P production. This hormone inhibits the pro-inflammatory NF-kB pathway, which in the absence of P will initiate endometrial breakdown (Fritz and Speroff, 2012). Hence, a series of inflammatory mechanisms are responsible for the cyclic endometrial remodeling until the perfect conditions are met and implantation occurs. The similarity between the “physiologic” menstrual cycle and the
mechanisms involved in the repair after endometrial injury raised researchers to question whether artificially induced injury may have any effect on implantation. Endometrial scratching, defined as the intentional injury to the endometrium performed to enhance the reproductive outcome of women desiring pregnancy, was thus attempted. The first experiment in this field was performed in 1907, where Loeb concluded that mechanical injury caused to the endometrium of guinea pigs during the progestational phase caused rapid decidualisation (Loeb, 1907). These results were later confirmed in the mouse model (Humphrey, 1969, Finn and Martin, 1972, Zhang, et al., 2015). Conversely, anti-inflammatory drugs seem to influence the decidualisation in rabbits (Hoos and Hoffman, 1983) and “re-normalize” the NK cell population in women undergoing COS (Junovich, et al., 2011). Although the exact mechanisms involved in this ER enhancement are yet to be identified, these experiments lead to the conclusion that inflammation seems to play a major role in implantation.

In humans undergoing IVF, most systematic reviews published thus far have concluded that endometrial injury is associated with a doubling of CPR and LBR in both the “general IVF” population (El-Toukhy, et al., 2012) and patients with a history of implantation failure (Nastri, et al., 2012, Potdar, et al., 2012). The overwhelming apparent strength of this initial evidence on endometrial scratching led to its widespread use (Lensen, et al., 2016) and extension to other treatment modalities, namely intra-uterine insemination (EL-Khayat, et al., 2015) and frozen embryo transfers (Dunne and Taylor, 2014). However, contesting voices began slowly to rise pointing out weak biological plausibility, heterogeneity in the existing RCTs, the potential for selection bias in the systematic reviews and the neglect to evaluate the risks associated with the procedure (Simon and Bellver, 2014, van Wely, 2014). This criticism was later proven defensible by modern-day RCTs failing to show any pragmatic benefit of endometrial scratching (Yeung, et al., 2014, Gibreel, et al., 2015). However, the topic of endometrial scratching is still a contentious and unresolved matter with multiple research groups attempting to evaluate its benefits in ongoing RCTs (Nastri, et al., 2015, Lensen, et al., 2016, van Hoogenhuijze, et al., 2017).

4.4.1.2 Steps towards same-cycle endometrial receptivity assessment

In an attempt to circumvent the need for endometrial injury to evaluate ER during the same cycle in which the embryo is to be transferred, two technologies currently used in fundamental science may soon have a place in everyday clinical practice. The first
approach involves performing transcriptomic profile analysis during either the late-proliferative phase [the specific timing during which abnormal cellular proliferation seems to be associated with failed implantation (Koot, et al., 2016)] or following in-vitro decidualization [IVD, a technique frequently used during basic laboratorial research for ER (Teklenburg, et al., 2010, Weimar, et al., 2012) which simulates the in vitro of the endometrium up until the theoretical day of ET]. The second analyses the secretomic profile instead, specifically the secretion of the prostaglandins (PG), such as PGE₂ and PGPGF₂α, in the fluid present in the endometrial cavity (Berlanga, et al., 2011). While these strategies seem promising, the leap from the bench to the clinic still lacks extensive validation.

4.4.2 Microbiome of the female genital tract

The uterine immune system is essential for both implantation and pregnancy (Arck and Hecher, 2013) and requires the activation of multiple inflammatory pathways (Dekel, et al., 2014). Such information raises one to question whether extrinsic factors such as the female urogenital tract microbiota may alter these inflammatory factors and hinder female fertility, embryo development and ER (Garcia-Velasco, et al., 2017, Moreno and Franasiak, 2017).

The least variable microbiota of the body is found in the female genital tract, specifically in the vagina. Genomic studies have indicated that the normal vaginal microbiome is dominated by *Lactobacilli*, with the most frequent species being the *L. iners*, followed by *L. crispatus, L. gasseri* and *L. jensenii* (Gajer, et al., 2012). These species maintain their dominance by producing lactic acid, which lowers the vaginal pH and creates an adverse environment for many other potential competing microorganisms. Disturbances of the female genital microbiota has been previously shown to be associated with various female sexual health and obstetrical complications such as an increased risk for sexually transmitted infections and preterm delivery (Gajer, et al., 2012, Hyman, et al., 2014). Furthermore, an accumulating amount of research in recent years has shown that women undergoing IVF are particularly prone to genital dysbiosis, with an abnormal vaginal microbiome being detected in up to 40% of patients (Sirota, et al., 2014). Besides infertility, the first microbiome studies performed specifically in women undergoing IVF has shown that an abnormal microbiota may affect both embryo development (Pelzer, et al., 2013) and ART pregnancy rates (Haahr, et al., 2016, Moreno, et al., 2016).
The general aim of this doctoral thesis was to improve our understanding on how COS affects ER and to what extent one can change clinical practice to enhance ART pregnancy outcome. To this effect, we formulated multiple specific objectives aimed to tackle the potential mechanisms behind COS-related endometrial non-receptivity from several possible angles.

**Specific aim 1**

**Better understand the relationship between ART outcome and circulating P during COS**

We intended to determine whether P during COS affects (C)LBR and to characterize its effect on these pregnancy outcomes, namely the potential for confounding or effect modification caused by ovarian response.

*Our scientific question was “is circulating P detrimental for embryo quality and ER and what is the most clinically useful way to predict its effect?” We hypothesized that P affects (C)LBR non-linearly and that its association may be better explained by the cumulative exposure to abnormal P values.*

For this specific aim, three studies were performed. Firstly, we sought to prove that the effect of abnormal late-follicular P on ART pregnancy rates was reproducible using LBR as the outcome of interest and to evaluate whether the relationship between both variables was linear (Study 1-A). Secondly, we assessed if late-follicular P affected embryo quality and, concomitantly, CLBR (Study 1-B). Finally, we posited whether the variation in P exposure throughout COS could predict ART pregnancy outcome in a fresh cycle better than a single assessment (Study 1-C).
Specific aim 2

Predict and optimize in-cycle ER

The evaluation of ER in a fresh ET cycle has been mostly (and controversially) limited to the assessment of EMT. We proposed with this specific aim to evaluate how EMT performed in the prediction of LBR when accounting for the endocrine milieu present during COS and to pit this measurement against a more novel approach of in-cycle ER assessment, transcriptomic expression analysis. Our theory was that EMT would no longer predict ART pregnancy outcome when the effects of COS were taken into consideration and that endometrial transcriptomic profiling may prove to be a better tool. Furthermore, we formulated an RCT to evaluate prospectively the potential benefit of in-cycle endometrial scratching for same-cycle CPR.

Two studies were designed to tackle this specific aim. In the first study, we retrospectively evaluated the effect of EMT on LBR and neonatal outcomes taking the endocrine profile into account (Study 2-A). For the second study, we initiated an RCT (the REFRESH trial, NCT02061228) comparing the CPR of women who did and did not perform endometrial scratching (Study 2-B). Nested within this study, we also planned to analyse the endometrial transcriptomic expression profile of a matched cohort of women participating according to their pregnancy outcome.

Specific aim 3

Assess the benefits of deferring ETs away from COS

Another objective of this thesis was to compare the early pregnancy outcomes between fresh ET and elective all-embryo vitrification followed by warming and transfer in a subsequent cycle. Our hypothesis was that endometria subject to COS would be associated with worse early pregnancy outcomes (i.e. lower CPR and higher miscarriage, EP and MZT rates) when compared to unstimulated uteri. This was, thus, a pragmatic approach to evaluate the effect of COS on ER.

In order to test this scientific question, we designed an explanatory prospective study (the ICE trial, NCT02148393) evaluating the CPR between high-responder women performing
either a fresh ET or a freeze-all approach (Study 3-A). However, given that both MZT and EP are rare events, two additional retrospective studies were designed to evaluate each of these outcomes more extensively (Studies 3-B and 3-C).

**Specific aim 4**

Determine the temporal extent of the hindering effect of COS on ER

Taking advantage of the large dataset of ART cycles performed yearly at Universitair Ziekenhuis Brussel (UZ Brussel), we were interested in evaluating whether the empirical decision that physicians frequently take to postpone FETs in order to avoid any residual effect of COS served any purpose. Our hypothesis was that COS had no carry-over effect and that the intentional deferral of FETs beyond the menstrual cycle that follows an IVF attempt would play no role in increasing pregnancy rates.

Given that the luteal phase of cycles triggered with hCG and a GnRH agonist vary substantially, two separate studies were conceived to evaluate the differences in terms of ART pregnancy rates between immediate and deferred FETs in both settings (Studies 4-A and 4-B).
6 MATERIALS AND METHODS

With the exception of study 3-C (explained in full in section 6.7.3.3, p. 67), details regarding a) ovarian stimulation and pituitary suppression, b) cycle monitoring and final oocyte maturation triggering, c) insemination, embryo quality assessment and cryopreservation, d) embryo transfer and endometrial preparation for FET, e) P immunoassay and f) main outcome measures were similar among the remaining studies as detailed below from sections 6.1 to 6.6 (p. 55-58). The study population, design and statistical analysis for each study are explained in section 6.7 (p. 59).

6.1 Ovarian stimulation and pituitary suppression

The choice of stimulation protocol was made on a case-by-case basis according to clinician preference and patient characteristics. Women began ovarian stimulation on day 2 of the menstrual cycle with either rFSH (Gonal-F®, Merck Serono Pharmaceuticals, Darmstad, Germany; or Puregon®, Merck Sharp & Dohme, New Jersey, USA) or HP-hMG (Menopur®, Ferring Pharmaceuticals, St. Prex, Switzerland). Pituitary down-regulation, using a GnRH antagonist, was performed with daily administrations of either cetrorelix (Cetrotide®, Merck Serono Pharmaceuticals, Darmstad, Germany) or ganirelix (Orgalutran®, Merck Sharp & Dohme, New Jersey, USA) starting from day 7 of the menstrual cycle onwards.

6.2 Cycle monitoring and final oocyte maturation triggering

Cycles were monitored by means of serial vaginal ultrasound scans and serum determinations of E2 and P. Whenever necessary, dose adjustments of rFSH/HP-hMG were performed according to ovarian response. As soon as three follicles of ≥17 mm were observed, final oocyte maturation was triggered with either highly purified urinary hCG
Materials and methods

[5000 UI or 10000 UI, according to the physicians’ preference and female weight; Pregnyl®, Merck Sharp & Dohme, New Jersey, USA] or 250 UI of recombinant hCG (Ovitrelle®, Merck Serono Pharmaceuticals, Darmstadt, Germany). Cumulus-oocyte complexes were collected by transvaginal aspiration approximately 36 hours after hCG administration.

6.3 Insemination, embryo quality assessment and cryopreservation

The insemination of the collected oocytes was performed via IVF or ICSI. Fertilization was confirmed 16-18 hours after insemination by the presence of two pronuclei and two polar bodies, from then onwards, embryo development was evaluated daily until the transfer or cryopreservation of cleavage stage embryos (Day 3) or blastocysts (Day 5 and 6). Good quality embryos that were not used for the fresh transfer were cryopreserved by means of vitrification with the use of a closed vitrification device with high-security straws (CBS-VIT-HS®; Cryobiosystem) in combination with dimethyl sulfoxide and ethylene glycol as cryo-protectants (Irvine Scientific Freeze Kit®; Irvine Scientific). Embryo quality was classified similar to what is described in a previous study (Van Landuyt, et al., 2005). Day 3 embryos were evaluated on the basis of the number and symmetry of the blastomeres, percentage of fragmentation, vacuolization, granulation, and multinucleation. Based on all these parameters, an embryo quality (EQ) score was assigned to all normally fertilized embryos based on a predefined algorithm, which divides them in four categories: excellent, good, moderate, or poor. On day 3, fresh embryos were considered eligible for transfer if at least 4 blastomeres were present with a maximum of 50% for fragmentation. Blastocysts were scored according to the grading system developed by Gardner and Schoolcraft (1999). Based on a) the expansion stage, b) the number of cells joining the compaction or blastulation, and c) the appearance of the trophectoderm (TE) and inner cell mass (ICM), one of the following blastocyst quality scores was given to all day 5 or day 6 embryos: excellent, good, moderate, or poor. On day 5, fresh embryos were considered eligible for transfer if they reached at least the fully compacted stage. The utilization rate was calculated based on all embryos that were either transferred or cryopreserved.

The following embryos were considered eligible for cryopreservation: day 3 embryos with ≥6 blastomeres and ≤50% fragmentation; day 5 and day 6, full, expanded or hatching blastocysts with a type A/B/C ICM and type A/B for TE. After warming (Irvine Scientific Thaw Kit ®; Irvine Scientific), day 3 embryos were cultured overnight in blastocyst medium.
until they were transferred the following day (day 4). Survival was scored based on the number of surviving blastomeres. If >2 cells degenerated after warming, a surplus embryo was warmed if available. One of the following embryo scores was given on day 4 at the moment of transfer: excellent, good, moderate or poor. Excellent embryos were at least fully compacted; good embryos were compacting embryos or partially compacted embryos/early blastocysts or compacted embryos and early blastocysts with a low cell number or embryos classified as excellent but with >10% vacuoles. Moderate embryos were partially compacting, or compacting embryos, with a low cell number or compacting embryos with >10% vacuoles or cleavage stage embryos that showed at least one division. Poor quality embryos did not develop from day 3 to day 4. Vitrified blastocyst (day 5 and day 6) ETs were performed on the day of warming. The survival rate was evaluated after warming based on the survival of all cells and the re-expansion state of the blastocyst. One of the following embryo scores was given at the moment of transfer: excellent, good, moderate or poor. Excellent embryos were at least fully blastocysts with ICM type A/B and TE type A. Good embryos were fully blastocysts with ICM type A/B and TE type B or fully blastocysts with ICM type C and TE type A or early blastocysts (BL2) or collapsed blastocysts. Poor blastocysts were early blastocysts (BL1) or fully blastocysts with ICM type A/B/C and TE type C or fully blastocysts with ICM type C and TE type B. Poor embryos were classified as compacted embryos or all blastocysts with ICM or TE type D.

6.4 Endometrial preparation for frozen embryo transfer

Both fresh and frozen embryos were transferred under ultrasound guidance using a soft embryo transfer catheter (Guardia Embryo Transfer Catheter®; Cook Medical). The choice to transfer one or more embryos was decided by the clinician depending on the patient age, embryo quality and cycle rank.

The artificial endometrial preparation consisted of a sequential administration of E2 valerate and micronized vaginal P as previously described for patients derived either from the UZ Brussel [studies 4-A and 4-B (van de Vijver, et al., 2014)] or My Duc Hospital [study 4-B, (Lan, et al., 2008), see below section 6.7.4.2 on p. 70 for the description of the study population for this specific study]. Summarily, 2 mg of E2 valerate was administered at least twice a day for 10-14 days, a dose which was later adjusted according to the EMT measured by ultrasound scan. If the EMT was 7 mm or more, vaginal P supplementation
was initiated. If the EMT was below 7 mm, patients continued to take oral E2 until the endometrium reached the necessary threshold, at which point P supplementation commenced. In the delayed FETs (studies 4-A and 4-B), we registered all cases in which progestins were administered during the preceding menstrual cycle.

### 6.5 Progesterone immunoassay

P was assessed throughout COS using a validated electrochemiluminescence immunoassay (Cobas 6000®, Roche, Basel, Switzerland) with a measured sensitivity and total imprecision (% CV) of 0.03 µg/L and <7%, respectively. The same assay was performed during the full duration of these studies and was regularly calibrated to minimise variation of the results associated with time and reagent batch renewal.

### 6.6 Main outcome measures

For pregnancy outcomes, we generally used the definitions provided by the International Committee for Monitoring Assisted Reproductive Technology (ICMART) in 2009 (Zegers-Hochschild, et al., 2009) and then 2017 (Zegers-Hochschild, et al., 2017).

A biochemical pregnancy was diagnosed only by the detection of hCG in serum or urine and that did not develop into a clinical pregnancy. Monozygotic pregnancies were identified when more than one gestational sac was visualized or when the number of fetal poles exceeded the number of gestational sacs. EP was defined as the presence of at least one gestational sac outside of the uterine cavity. In our definition, we also considered heterotopic pregnancies (the simultaneous visualization of both at least one intra-uterine and one extra-uterine gestational sac) as being ectopic. Clinical pregnancy (CP) was defined as the visualization of a gestational sac during transvaginal ultrasound at 7 weeks of gestational age. Live birth (LB) deliveries were defined as such after 24 weeks, with twin or higher order pregnancies being considered only once. Cumulative live birth was defined as the delivery of a live born (>24 weeks of gestation) in the fresh or in one of the subsequent frozen–thawed embryo transfer cycles. Each IVF treatment cycle was considered separately and only the first delivery was included in the analysis.
6.7 Study population, design and statistical analysis

A p-value was considered significant whenever <0.05, adjusting for multiple comparisons when necessary. All statistical analyses were performed in Stata Software versions 12 or 13 (StataCorp, College Station, Texas, USA), or SPSS version 20 (IBM, New York, USA).

6.7.1 Specific aim 1

Better understand the relationship between ART outcome and circulating P during COS

6.7.1.1 Study 1-A

Evaluate if the effect of abnormal late-follicular P on ART pregnancy rates is reproducible on LBR and whether the relationship between both variables is linear

We performed a retrospective, single-centre cohort study between January 2006 and March 2012 including all women between 18-36 years old undergoing their first or second COS for IVF in UZ Brussel. As IVF success is influenced by cycle rank (Malizia, et al., 2009, Luke, et al., 2012), especially after three or more attempts, we excluded all cycles of patients who had more than one previous cycle performed in our centre. To avoid other potential confounders, only patients younger than 36 years old planned for fresh embryo transfer using autologous oocytes and under GnRH antagonist pituitary suppression were included. Furthermore, couples with planned embryo biopsy were excluded.

To avoid bias of the results by assuming that any relationship between P and LBR may be linear (Bosch, et al., 2010), patients were divided into the following six distinct ordinal interval P groups (ng/mL): ≤0.50, 0.51-0.75, 0.76-1.00, 1.01-1.25, 1.26-1.50 and >1.50. These thresholds were selected taking into account the cut-offs applied amongst previous studies (Bosch, et al., 2010, Xu, et al., 2012) and a balance between the use of clinically relevant intervals that would divide our sample homogenously as not to hinder the inter-interval statistical analysis.

LBR per oocyte retrieval were calculated for each P interval, determining odds ratios between each of the P level intervals and the highest P group (>1.50 ng/mL) using logistic regression adjusting the standard errors to allow repeated cycles performed in the same women to be included. Using this approach, we would be able to further analyse the relationship between P and IVF outcomes to eventually determine the existence of a lower
P threshold under which these endpoints were hindered as much as they are already known to be with P levels above 1.50 ng/mL. Potential confounder adjustment for the number of embryos transferred (1 or ≥2), stage of embryo development (day 3 cleavage-stage or day 5 blastocyst), and the continuous variables age, serum E2 levels on the day of hCG administration, total dose of exogenous FSH and number of oocytes retrieved was performed. Other potential differences amongst the groups were assessed using either logistic regression (for binomial variables) or Kruskal-Wallis (for continuous variables, followed by a post-hoc pairwise analysis when significant).

6.7.1.2 Study 1-B

Assess whether late-follicular P affects EQ and, concomitantly, CLBR

This was a retrospective, single-centre cohort study including all women who underwent IVF at our centre between the January 2010 and the December 2015. Approval to retrieve and analyse the data was provided by the Ethics Committee of UZ Brussel. In order to evaluate the oocyte maturation rate and minimize the effects of the known differences in terms of fertilization between conventional IVF and ICSI (van der Westerlaken, et al., 2005, Abdalmageed, et al., 2015), we included only patients who underwent ICSI. Patients who performed both conventional IVF and ICSI within the same cycle were also excluded from the study. Furthermore, women with an unknown LB outcome, who underwent GnRH agonist triggering, who were acceptors of donated oocytes, who performed embryo biopsy for pre-implantation genetic diagnosis or freeze-only cycles were disregarded.

Patients were categorized into three groups according to the following P levels on the day of ovulation triggering: ≤0.50 ng/mL, 0.51-1.49 ng/mL and ≥1.50 ng/mL. These thresholds were selected according to the cut-offs applied amongst previous studies (Bosch, et al., 2010, Santos-Ribeiro, et al., 2014) and a balance between the use of clinically relevant intervals that would divide our sample homogenously so as not to hinder the inter-interval statistical analysis. These groups were compared using multivariable generalized estimating equations (GEE) regression modelling (to account for the eventual clustering of cycles performed by the same women) followed by exploratory pairwise comparisons whenever warranted. Additionally, maturation, fertilization and utilization rates were also compared using mixed-effects multilevel regression analysis, accounting for the potential for correlated outcomes within fresh cycles performed in the same patient and also oocytes deriving from the same ICSI cycle.
Materials and methods

LB outcomes were assessed using multivariable GEE regression analysis including the P levels categories as a predictor along with other potential confounders in each model according to the outcome being evaluated. We chose to adjust for potentially confounding variables previously known to affect pregnancy rates or whenever they varied significantly between the study groups. That said, we adjusted for female age and body mass index (BMI), reasons for infertility (male factor, tubal factor, ovarian factor, endometriosis, uterine factor, other female factor or otherwise unexplained infertility), number of COCs retrieved and the embryo developmental stage at transfer (cleavage or blastocyst stage) in all regression models. Moreover, for fresh LB we included also the total dose of COS, the circulating E2 on the day of triggering, embryo quality and the number of embryos transferred (single embryo versus multiple embryos). For the first frozen-thawed ET cycle, after a failed fresh embryo transfer, we adjusted additionally for number of embryos transferred and the type of FET cycle preparation (natural cycle with spontaneous LH peak, natural cycle with ovulation triggered using hCG, hormone-replaced cycle). Given the potential confounding effect of oocyte retrieval on the effect of P on CLBR, we performed two supplemental analysis. Firstly, we performed a sensitivity analyses evaluating the effect of each P level on CLBR. Then, we repeated the GEE regression analysis for CLBR detailed before, however, this time adding an interaction term between serum P and the number of oocytes retrieved.

6.7.1.3 Study C

Evaluate whether the variation in P exposure throughout COS can predict ART pregnancy outcome in a fresh cycle better than a single measurement

We performed a retrospective, single-centre cohort study between January 2010 and March 2015 including all women undergoing COS for IVF in UZ Brussel. To minimize potential confounding, only patients planned for fresh ET using autologous oocytes and under GnRH antagonist pituitary suppression were included. Furthermore, couples with a planned embryo biopsy, managed natural cycles or in-vitro maturation were excluded from the analysis. Approval and waiver of written informed consent to retrieve and analyse the data was obtained from the Ethical Committee of UZ Brussel.

P variation during COS (PV-OS) was defined as the difference between the serum levels of P on the day of hCG administration and the baseline circulating P concentrations on the day COS was initiated (“hCG-day P” – “COS-start P”). Statistically significant
predictors of PV-OS were determined using multivariable GEE regression modelling (to account for the eventual clustering of cycles performed by the same women) including the following potential confounders: female age, the type, duration and total dose of exogenous COS utilized, and the number of oocytes retrieved.

Our main outcome measure was LB delivery after 24 weeks, with unknown outcomes (i.e. patients lost to follow-up) being considered as negative. LBR per embryo transfer cycle were assessed using multivariable GEE regression analysis including PV-OS and the following known potential confounders for ART pregnancy outcome in the model: female age, number of preceding IVF/ICSI cycles, the type and total dose of exogenous gonadotropin administration, the endocrine profile (P and E₂) on the day of hCG administration, number of oocytes retrieved, stage and number of embryos transferred, and quality of the best embryo transferred. PV-OS was included in the regression model both as a continuous variable and categorized in regular intervals of 0.20 ng/mL, to facilitate the clinical interpretation of the results and to postulate potentially relevant clinical cut-offs of PV-OS for further analysis. To minimize bias by assuming that the relationship between PV-OS and LB may be linear (Templeton, et al., 1996, Bosch, et al., 2010, Sunkara, et al., 2011, Xu, et al., 2012, Santos-Ribeiro, et al., 2014), we pitted all significant PV-OS linear relationships that were found against multiple first and second degree fractional polynomials (FP) using an algorithm developed by Sauerbrei, et al. (2006), with the best-fit being chosen for the final regression model.

Finally, in order to verify whether the association between the duration of LFEP and CPR found by Huang, et al. (2012) was reproducible using LBR as the outcome of interest, we performed an additional analysis in which we divided our sample according to the duration of P elevation (0 days, 1 day, 2 days or ≥3 days). We also repeated the analysis using two cut-off levels of LFEP, specifically 1.00 ng/mL [as used by Huang, et al. (2012)] and 1.50 ng/mL. These models were then compared to each other using the Akaike's and Bayesian information criteria (AIC and BIC, respectively). This statistical approach is a measure of the relative quality of statistical models for a given set of data. Specifically, AIC and BIC estimate the quality of a model relative to each of the other models. The lowest AIC/BIC suggest the best model, with a difference (ΔAIC and ΔBIC, respectively) of ≥10 providing strong evidence of significant superiority of the model with the lowest AIC/BIC (Kass and Raftery, 1995, Burnham and Anderson, 2010).
6.7.2 Specific aim 2
Optimize and predict in-cycle ER

6.7.2.1 Study 2-A
Determine the effect of EMT on LBR and neonatal outcomes taking the endocrine profile into account

This was a retrospective, single centre, cohort study including ART treatment cycles performed at the UZ Brussel between January 2010 and December 2014. Only cycles in which patients underwent a GnRH antagonist down-regulated stimulation protocol followed by a fresh embryo transfer were included. To minimize confounding derived from women with a baseline poor prognosis, we excluded cycles in women aged 40 years old or above and managed natural cycles. The exclusion criteria also included those who performed cycles with known uterine abnormalities [including uterine malformations and intrauterine disease diagnosed during ultrasound or a preceding hysteroscopy (e.g. Asherman’s syndrome, endometrial polyps, submucosal myomas)] and the planned use of either surgically retrieved sperm, donor oocytes, in-vitro maturation or preimplantation genetic diagnosis. Approval and waiver of written informed consent to retrieve and analyse the data was obtained from the Ethical Committee of UZ Brussel.

EMT was measured in millimetres on the day or the day prior to ovulation triggering. We considered the EMT as the maximal anterior-posterior distance between both endometrial layers approximately 1 cm from the uterine fundus, subtracting the thickness of intrauterine fluid in the unlikely event that such was detected. The following variables were assessed as potential predictors of EMT: female age, BMI, total dose of exogenous rFSH/HP-hMG, duration of COS, and late-follicular phase E2 and P. In order to avoid bias by assuming that the relationship between these continuous predictors and EMT was linear, the best-fitting FP of each of these variables were compared against their linear function to assess which better described their association with EMT.

EMT was our main exposure variable for LB. Aware of the fact that the categorization of continuous variables may be of limited value to evaluate the real effect of a predictor (since it simultaneously assumes that values in different intervals have different effects even if very close to each other and that values on the extremes but within the same interval group have the same effect), our main approach was once again to assess the relationship of EMT
as a continuous variable comparing the best fitting FP against the linear function (model 1). Nonetheless, to facilitate the application of the results into everyday clinical practice, EMT was also categorized in the following regular-2-mm-intervalled categories: <7.0 mm, 7.0-8.9 mm, 9.0-10.9 mm, 11.0-12.9 mm and ≥13.0 mm (model 2). These intervals were chosen to provide equal intervals as close as possible to the different threshold values employed across previous studies.

When assessing the effect of EMT on LB, we considered the following variables as potential confounders: female age, BMI, rank of IVF/ICSI treatment cycle attempt, number of useable embryos (transferred and cryopreserved), number of embryos transferred (single versus multiple), embryo stage at transfer (cleavage day-3 versus blastocyst day-5), quality of the best embryo transferred and both late-follicular phase E2 and P levels (determined on the day or day before ovulation triggering).

We also assessed the effect of EMT on neonatal outcomes, specifically gestational age at delivery, preterm birth (<37 weeks), birthweight and low birthweight (<2500 g). Given the non-linear progression of fetal growth as pregnancy develops and differences according to fetal sex (Hadlock, et al., 1985), the recorded birthweights were standardized using z-scores (Niklasson and Albertsson-Wikland, 2008). The z-scores indicate how many standard deviations an observation was above or below the reference population mean, accounting for gestational age and neonatal sex. Furthermore, given the significant influence of multiple pregnancies on fetal development, only singleton live births were evaluated. For this analysis, we considered the following variables as potential confounders: female age, BMI and parity (nulliparous versus multiparous).

Baseline characteristics were compared among the above-mentioned EMT categories for model 2, with categorical variables being presented using relative frequency (%). Continuous variables were presented as means (standard deviation) or medians (interquartile range) according to the normality of the distribution. Comparisons were performed using univariable linear (continuous variables), logistic (dichotomous variables) or ordered logistic (for embryo quality) regression analysis, with robust standard errors to account for the possibility of clustering of more than one cycle deriving from the same couple.

The predictors for EMT, LB and neonatal outcomes were determined using both univariable and multivariable regression analysis with robust standard errors, adjusting
Materials and methods

6.7.2.2 Study 2-B

Study the use of in-cycle endometrial scratching and analysing the endometrial RNA expression profile to predict pregnancy outcome

We propose a two-arm randomised, single-centre, controlled open-label trial. Summarily, women undergoing exogenous COS for ART in an antagonist downregulated cycle will be included in either the control or intervention. Those in the intervention group will additionally undergo an endometrial biopsy during the follicular phase. The full study protocol can be consulted in section 7.2.2 (p. 129).

6.7.3 Specific aim 3

Assess benefits of deferring ET away from COS

6.7.3.1 Study 3-A

Evaluate the CPR in high-responder women performing either a fresh ET or freeze-all

We designed a two-arm randomised, single-centre, open-label clinical trial. Women undergoing COS for ART in a GnRH antagonist down-regulated cycle and at high risk for OHSS will be included in either “fresh ET” or “vitrified-warmed ET” groups. Those in the “fresh ET” group will perform GnRH agonist triggering followed by intensified luteal phase support while the “vitrified-warmed ET” group will electively vitrify all viable embryos after GnRH triggering and perform the ET in a subsequent unstimulated cycle. The full study protocol can be consulted in section 7.3.1 (p. 137).

6.7.3.2 Study 3-B

Compare MZT rates between fresh ET and FET

This study retrospectively reviewed all consecutive fresh or frozen ET cycles with autologous or donated oocytes performed in our centre between January 2004 and December 2013. Approval to retrieve and analyse the data was obtained from the Ethical
Committee of UZ Brussel. Only single ET cycles resulting in a CP documented by ultrasonography were considered for the analysis.

Clinical pregnancy was defined by the presence of an intrauterine gestational sac as visualized by transvaginal ultrasound examination (Zegers-Hochschild, et al., 2009) and later subdivided in the two following groups: MZT and singleton (non-MZT). Monozygotic pregnancies were identified when more than one gestational sac was visualized or when the number of fetal poles exceeded the number of gestational sacs. Chorionicity was established by counting the number of gestational sacs or by taking into account the thickness of the amniotic inter-twin membrane. A MZT pregnancy was considered dichorionic when two separate gestational sacs were visualized during early first-trimester (7 weeks) ultrasound. The thickness of the inter-twin membrane or the presence of the lambda or T signs were also used to confirm chorionicity in the presence of two adjacent gestational sacs or at a later period of the gestation (Finberg, 1992, Sepulveda, et al., 1996). Cases classified as dichorionic-diamniotic twin pregnancy but later recognized as chromosomally discordant were not included in the analysis since the simultaneous occurrence of a spontaneous \textit{in-vivo} pregnancy in the same cycle could not be excluded. Monozygotic pregnancies were considered monochorionic when at least two fetal poles were visualized within one gestational sac with one chorionic area. Amnionicity was established by the presence or absence of a thin inter-twin membrane and by counting the number of yolk sacs or amnions within the sac. Monoamniotic twin pregnancies were characterized by the presence of a single amniotic cavity and yolk sac (Bromley and Benacerraf, 1995).

The following parameters were considered during the analysis: maternal age (or the age of the oocyte donor), use of donor oocytes, zona pelucida manipulation (ICSI, embryo biopsy for pre-implantation genetic diagnosis and assisted hatching), embryo stage at time of ET (cleavage, compaction, early or advanced blastocyst), type of ET (fresh or frozen), and culture media used prior to cryopreservation. During the period of time included in the study, no changes were performed in the standard laboratory culture conditions except for culture media and the introduction of the vitrification method for embryo cryopreservation. The impact of the cryopreservation protocols was not assessed due to the low number of MZT CP obtained after frozen ET.

The baseline sample characteristics were compared amongst the MZT and singleton groups using either the Fisher’s exact (for categorical variables) or t-student (for
Materials and methods

67

continuous variables) tests followed by crude exploratory pairwise comparisons whenever necessary. MZT rates per ET were assessed using multivariable logistic regression to account for the before-mentioned study parameters. Since not all embryos were cultured in the same type of medium before cryopreservation and after thawing, the culture medium could not be included as a possible confounder in the general analysis. We also performed a subgroup analyses dividing the sample according to the type of ET and live birth outcome of the MZT pregnancy. Given the low incidence of MZT (<10%), odds ratios could be considered adequate estimates of risk ratios (Zhang and Yu, 1998).

6.7.3.3 Study 3-C

Compare EP rates between fresh ET and FET

We obtained anonymized data for ART cycles carried out in the UK between 2000 and 2012 from the Human Fertilisation and Embryology Authority (HFEA) database. The HFEA is the statutory regulator of ART in the UK and, following a mandatory reporting policy enforced by law, this institution has kept records of the ART cycles performed in the UK since August 1991. This database has been regularly audited, cross checked and upgraded to ensure the consistency and relevance of the collected data (Human Fertilisation and Embryology Authority, 2007). Furthermore, since March 2002, a double entry system complemented with a thorough verification process including approximately 800 validation rules was set in place to maximise the integrity of the data (Human Fertilisation and Embryology Authority, 2007).

In order to avoid biases related to the quality of the data inserted in the database, we only considered eligible ART cycles carried out in the UK between 2000 and 2012. The rationale for this approach was based on the fact that the collection of information regarding the types of infertility treated only began after the year 2000. In this regard, given that the type of infertility diagnosis has been shown to affect the risk of EP (Strandell, et al., 1999, Farquhar, 2005, Clayton, et al., 2006), we restricted our analysis to cycles performed after 1999 in which a valid early positive pregnancy outcome was registered.

To minimise bias and allow adequate adjustment for potential confounders, we excluded data on cycles in which the treatment details (indication for ART, specific type of treatment, use of ovarian stimulation, type of embryo transfer and number and/or developmental stage of the transferred embryos) were unclear or unknown. Furthermore, we limited our analysis to either intrauterine insemination, or IVF/ICSI cycles, excluding
other less frequent or not currently used techniques (i.e. gamete intrafallopian transfer, zygote intrafallopian transfer or intravaginal insemination). We also excluded cycles in which donor oocytes were used or if the indication for treatment was not infertility, such as IVF for PGD in patients otherwise expected to be fertile. Finally, given the fact that previous reports supported a reduced incidence of EP following the transfer of frozen embryos (Ishihara, et al., 2011, Shapiro, et al., 2012, Huang, et al., 2014, Fang, et al., 2015, Londra, et al., 2015), we did not include cycles having a simultaneous transfer of both fresh and frozen embryos.

The incidence of EP was compared between the two types of studied ART treatments (intrauterine insemination and IVF/ICSI). Given the low incidence of EP (<10%), we considered odds ratios as adequate estimates of risk ratios (Zhang and Yu, 1998).

The statistical significance of the trends in EP rates and relevant potential confounders observed over time were determined using Poisson regression. All graphical depictions of trends were smoothed using moving averages, adjusting for the mean EP rates of the year before and after, in order to minimize spurious variations which could limit the visual interpretation of the temporal trend. We assessed the relationship between potential confounders and EP using multivariable logistic regression. The variables included in the model as potential confounders were year of treatment, female age, use of exogenous ovarian stimulation, infertility diagnosis and type of treatment (intrauterine insemination or IVF/ICSI). For IVF/ICSI cycles, the type of embryo transfer (fresh or frozen/thawed), embryo development stage (days of in-vitro embryo culture until transfer) and number of embryos transferred were also accounted for via a separate logistic regression model.

**6.7.4 Specific aim 4**

**Determine the temporal extent of the hindering effect of COS on ER**

**6.7.4.1 Study 4-A**

**Following a failed fresh ET**

We performed a retrospective cohort study including all women who underwent at least one FET after ovarian stimulation for IVF between January 2010 and November 2014 at UZ Brussel. Approval to retrieve and analyse the data was provided by the Ethics Committee of UZ Brussel.
Only the outcomes of the first FET cycles performed after ovarian stimulation and a failed fresh embryo transfer attempt were assessed. In order to minimize bias, we only included FETs that followed fresh cycles in which a GnRH antagonist and hCG alone were administered for down-regulation and ovulation triggering, respectively. Women who were acceptors of donated oocytes or performed either in-vitro maturation or embryo biopsy for preimplantation genetic diagnosis were excluded from the study. Furthermore, if during the preceding COS cycle either ovulation was triggered with a drug other than hCG (e.g. a GnRH agonist, either alone (Griesinger, et al., 2006) or in combination with hCG (Shapiro, et al., 2008)) or hCG was administered for reasons other than for ovulation triggering (e.g. for late-follicular ovarian stimulation (Blockeel, et al., 2009) or luteal phase support (van der Linden, et al., 2015)), these cycles were also disregarded. Finally, FET cycles performed under GnRH agonist down-regulation or with concomitant exogenous ovarian stimulation were also excluded from the sample.

The “timing” of the first frozen embryo transfer (TF-FET) was defined as the interval between oocyte retrieval and the start of the first FET cycle. We divided our sample in cycles with either an immediate (≤22 days after oocyte retrieval) or delayed (>22 days after oocyte retrieval) start of FET cycle. This cut-off was devised by adding the interval between oocyte retrieval and the first pregnancy test (15 days) to an extra interval of up to 7 days necessary for the patients to have their withdrawal bleeding and begin their first FET cycle. By using these intervals, we essentially divided our sample in a) women who performed an immediate FET and b) women who waited at least one menstrual cycle until having their transfer.

Basic demographic characteristics were compared amongst the women who performed immediate and delayed FET, using the t-test/Mann-Whitney (continuous variables) or χ² (categorical variables) tests. CPR and LB delivery rates per FET were assessed both crudely and using multivariable logistic regression accounting for the following known potential confounders for FET cycle outcome: female age, number of good quality embryos produced, type of FET cycle, stage and number of embryos transferred, and quality of the best embryo transferred. Crude and adjusted odds ratios were estimated, adjusting the standard errors to eventually allow for more than one fresh cycle performed in the same women to be included in the analysis.
6.7.4.2 Study 4-B

Following a “freeze-all protocol”

We performed a retrospective cohort study including the first FET cycle of all women who underwent their first freeze-all protocol between October 2010 and October 2015 at one of the two following hospitals: UZ Brussel (Centre 1, Belgium) and My Duc Hospital (Centre 2, Vietnam). We included only artificially-supplemented FETs. In order to minimize the potential for confounding, women who used donated oocytes or performed either in-vitro maturation or preimplantation genetic diagnosis were excluded from the analysis. Finally, two delayed FET cycles which were preceded by a cancelled attempt of immediate FET due to a persistently thin endometrium were also excluded from the analysis to avoid misleadingly skewing negatively the pregnancy outcome in the delayed FET group.

The TF-FET was defined as the interval between oocyte retrieval and the start of the first FET cycle. We divided our sample in women with either an immediate (following the GnRH agonist withdrawal bleeding) or delayed (by at least one menstrual cycle) start of FET cycle. Basic demographic characteristics were compared amongst these groups using the t-test (for continuous variables) or χ²-test (for categorical variables). CPR/FET were assessed both crudely and using mixed-effects multivariable regression analysis, accounting for the potential for correlated outcomes between cycles performed in the same centre. In this regression analysis, we also accounted for variables which were either unevenly distributed amongst the study groups or presumed as potential confounders, namely: female age (as a continuous variable), indication for freeze-all (high-risk of OHSS, late-follicular P >1.50 ng/mL, late-follicular endometrium <7.0 mm, and patient preference or other reasons), number of good quality embryos produced (as a continuous variable) and number (single versus multiple), developmental stage (cleavage versus blastocyst stage) and quality of the embryos transferred (quality of the best embryo transferred).

Although our sample size would likely be too limited to perform an adequate confounder-adjusted sensitivity analysis, for preliminary exploratory purposes, we crudely assessed whether the main outcome parameter presented obvious disparities according to potential confounders distributed unequally amongst the study groups, the centre where the procedure was performed and the indication for freeze-all, evaluating the CPR/FET according to these subgroups.
7 RESULTS

7.1 Specific aim 1
Better understand the relationship between ART outcome and circulating P during COS

7.1.1 Study 1-A
Evaluate whether the effect of abnormal late-follicular P on ART pregnancy rates is reproducible on LBR and if the relationship between both variables is linear

Full manuscript published
Live birth rates after IVF are reduced by both low and high progesterone levels on the day of human chorionic gonadotrophin administration


1Centre for Reproductive Medicine, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel, Laarbeeklaan 101, Brussels 1090, Belgium
2Department of Obstetrics, Gynaecology and Reproductive Medicine, Hospital Universitario de Santa Maria, Av. Conde Ega Moniz, Lisbon 1649-036, Portugal

Correspondence address: E-mail: samueloribeiro@gmail.com

Submitted on September 15, 2013; resubmitted on May 15, 2014; accepted on May 22, 2014

STUDY QUESTION: Are low serum progesterone levels on the day of human chorionic gonadotrophin (hCG) administration detrimental for live birth delivery rates during in vitro fertilization (IVF)?

SUMMARY ANSWER: Progesterone levels ≤0.5 ng/ml on the day of hCG administration hinder live birth rates.

WHAT IS KNOWN ALREADY: Fundamental research has shown that the presence of late follicular phase progesterone is essential for follicular development, ovulation and endometrial receptivity. However, previous studies in patients undergoing ovarian stimulation have only assessed if progesterone levels in the higher range are detrimental for pregnancy or not. That said, information on the effect of the full range of late follicular progesterone on IVF outcomes is still lacking.

STUDY DESIGN, SIZE, DURATION: This was a retrospective, single-centre cohort study with 2723 cycles performed in patients aged between 19 and 36 and undergoing controlled ovarian stimulation between January 2006 and March 2012 for their first or second attempt of IVF followed by a fresh embryo transfer (ET).

PARTICIPANTS/MATERIALS, SETTING, METHODS: All patients underwent ovarian stimulation using a gonadotrophin-releasing hormone (GnRH) antagonist pituitary down-regulation. Final oocyte maturation was triggered with hCG 36 h before oocyte retrieval. On the day of hCG administration, serum progesterone evaluation was performed. Live birth delivery rates were compared amongst various ordinal and regular progesterone intervals (<0.5, 0.50–0.75, 0.75–1.00, 1.00–1.25, 1.25–1.50, >1.50 ng/ml) using logistic regression.

MAIN RESULTS AND THE ROLE OF CHANCE: The average age of our sample was 30.3 years. Almost 82% of all embryo transfers were of a single embryo and 51.8% were performed with a Day 3 embryo. The average value (+ standard deviation) of progesterone on the day of hCG administration was 1.02 ± 0.50 ng/ml and the live birth rate was 23.4%. The live birth rates (according to the above-described ordinal serum progesterone intervals) were 17.1, 25.1, 26.7, 25.5, 21.9 and 16.6%, respectively. The live birth rates were significantly lower in patients with both low (<0.5 ng/ml) and high (>1.5 ng/ml) late follicular progesterone levels (P < 0.05).

LIMITATIONS, REASONS FOR CAUTION: The main limitation of our study was its retrospective nature. Furthermore, our study was restricted to patients under GnRH antagonist pituitary suppression and requires confirmation in a GnRH agonist setting.

WIDER IMPLICATIONS OF THE FINDINGS: This study comprehensively assessed the relationship between live birth delivery rates and progesterone levels on the day of hCG administration during ovarian stimulation for IVF. Clinically relevant lower (<0.5 ng/ml) and higher (>1.5 ng/ml) progesterone level limits were determined.

STUDY FUNDING/COMPETING INTEREST(S): No funding was received for this study and the authors have no conflicts of interest to declare.

Key words: ovarian stimulation / progesterone / in vitro fertilization / live birth rate
Introduction

Pituitary suppression with gonadotropin-releasing hormone (GnRH) analogues has drastically reduced the incidence of premature luteinizing hormone (LH) surge during controlled ovarian stimulation (COS) to <2% per cycle (Smits et al., 1992; Felderbaum and Diederich, 1999). Nonetheless, suboptimal follicular development (P) elevations unrelated to a premature LH surge still occur in up to 38% of all in vitro fertilization (IVF) cycles regardless of artificial pituitary suppression (Ubaldi et al., 1996; Bosch et al., 2003).

Many research groups have attempted to assess how late follicular high P levels affect IVF. The majority of available trials have demonstrated that high P levels may hinder IVF success (Felderguth et al., 1989; Schoorl et al., 1991; Silverberg et al., 1991; Mio et al., 1992; Cheek et al., 1993; Dimliefeld et al., 1993; Fanchin et al., 1993, 1997a,b; Harada et al., 1995; Yonei et al., 1995; Randell et al., 1996; Shuman et al., 1996; Younis et al., 1998, 2001; Bosch et al., 2003, 2010; Ozcalir et al., 2004a,b; Li et al., 2008; Kikudai et al., 2010; Kolibianakis et al., 2012; Xu et al., 2012). However, these results were initially disputed by other researchers who found late follicular P rises to be either irrelevant (Edelstein et al., 1990; Givens et al., 1994; Bustillo et al., 1995; Abuzed and Sasy, 1996; Hofmann et al., 1996; Huang et al., 1996; Miller et al., 1996; Ubaldi et al., 1996; Fanchin et al., 1997a,b; Moffit et al., 1997; Martinez et al., 2004; Sahil et al., 2009) or even beneficial in selected populations (Legro et al., 1993; Doki et al., 1999). Furthermore, a 2007 meta-analysis (Venets et al., 2007) showed that high serum P levels on the day of human chorionic gonadotropin (hCG) administration did not significantly affect pregnancy rates (PR). However, the heterogeneity of the studies included in this meta-analysis may have biased the conclusions (Bosch, 2008; de Ziegler et al., 2008; Fleming, 2009) and updated meta-analyses by the same group provided contradictory results (Kolibianakis et al., 2012; Venets et al., 2013).

Although the main function of progesterone is to support the endometrium in the luteal phase, fundamental research has shown that there is a physiologic late follicular phase P increase (De Geyter et al., 2002) which, besides contributing to the timing of ovulation (Covens et al., 1992), may be essential for follicular development (Collins and Hodgen, 1986; Chaffin et al., 1992; Bridges et al., 2006; Stoof et al., 2010). And, while the before-mentioned studies seem to conclusively resolve the debate on whether a supra-physiologic P elevation affects the PR following COS or not, they infer very little on the influence of P levels in the lower range on IVF outcomes. Furthermore, animal experiments have shown that blocking mid-cycle P production is detrimental for oocyte maturation (Borman et al., 2004), oocyte fertilization competence (Zalewski-Woesten et al., 1994) and granulosa/thecocytes luteinization (Hibbert et al., 1996). Consequently, it may be hypothesized that low P levels may compromise PR.

Taking into account that none of the previous studies have assessed the effect of low late follicular P levels on IVF live birth rates, our primary objective was to evaluate whether there is a relationship between these two variables. Our secondary objective was to examine which factors could be associated with elevated and reduced late follicular P levels.

Materials and Methods

Study population and design

We performed a retrospective, single-centre cohort study between January 2006 and March 2012 including all women between 18 and 36 years of age undergoing their first or second COS for IVF in our centre. As IVF success is influenced by cycle rank (Malbick et al., 2009; Luke et al., 2012), especially after three or more attempts, we excluded all cycles of patients who had more than one previous cycle performed in our centre. To avoid other potential confounders, only patients younger than the age of 36 years planned for fresh embryo transfer using antagonist protocols and under GnRH antagonist pituitary suppression were included. Furthermore, couples with planned hysteroscopy biopsy were excluded from the analysis.

Ovarian stimulation and pituitary suppression

The choice of stimulation protocol was made on a case-by-case basis according to clinician preference and patient characteristics. Women began ovarian stimulation on Day 2 of the menstrual cycle with either recombinant follicle-stimulating hormone (FSH; Gonal F), Merck Serono Pharmaceuticals, Damstadt, Germany; or Pergonal®, Merck Sharp & Dohme, Whitehouse Station, NJ, USA; or Bionova®, Merck Sharp & Dohme) or highly purified human menopausal gonadotropin (HMG; Menopur®, Ferring Pharmaceuticals, St. Pölten, Austria) or low-dose recombinant FSH (Cetrotide®, Merck Serono Pharmaceuticals) or gonal-f (Olgapulin®, Merck Sharp & Dohme) starting from Day 7 of the menstrual cycle onwards.

Cycle monitoring and final oocyte maturation triggering

Cycles were monitored by means of serial vaginal ultrasound scans and serum determination of estradiol (E2) and P. Whenever necessary, dose adjustments of rFSH/HP-HMG were performed according to ovarian response. As soon as three follicles of ≥17 mm were observed, final oocyte maturation was triggered with either highly purified urinary hCG (5000 IU) or 10,000 IU, according to the physician’s preference (Prempak®, Merck Sharp & Dohme) or 250 IU of recombinant hCG (Cetrotide®, Merck Serono Pharmaceuticals). Cumulus-oocyte complexes (COCs) were collected by transvaginal aspiration ~36 h after hCG administration.

Progestosterone assessment immunoassay

Serum P levels were assessed on the day of hCG administration using a validated electrochemiluminescence immunoassay (Cobas 6000® Roche, Basel, Switzerland) with a measured sensitivity and total imprecision (% coefficient of variation) of 0.03 ng/ml and <7%, respectively. The same assay was performed during the full duration of the study and was regularly calibrated to minimize variation of the results associated with time and reagent batch renewal.

Statistical analysis

To avoid bias of the results by assuming that any relationship between P and PR may be linear (Bosch et al., 2010), patients were divided into the following six distinct ordinal intervals (ng/ml): <0.5, 0.51–0.75, 0.76–1, 1–1.25, 1.26–1.50 and >1.50. These thresholds were selected taking into account the cut-offs applied amongst previous studies (Bosch et al., 2010, Xu et al., 2012) and a balance between the use of clinically relevant intervals that would divide our sample homogeneously so as not to hinder the inter-interval statistical analysis.

Live birth delivery rates per oocyte retrieval were calculated for each P interval, determining odds ratios (ORs) between each of the P level intervals and the highest P group (>1.5 ng/ml) using binary regression and adjusting the standard errors to allow repeated cycles performed in the same women to be included. Using this approach, we would be able to further analyse the relationship between P and IVF outcomes to eventually determine the existence of a lower P threshold under which these end-points were hindered as much as they are already known to be in P levels >1.5 ng/ml.
Results

Patient demographics and general characteristics of the treatment protocol

The baseline and treatment characteristics of the 2723 cycles (performed in 2157 patients) included in the study are detailed in Table I (Supplementary data). Tables SI and SI1 compare these results according to IVF outcomes and P levels, respectively. The average age of our sample was 30.5 years (range from 19.0 to 36.0 years). Almost 82% of all embryo transfers (ET) were of a single embryo and 51.8% were performed with a Day 5 embryo. The average value (± standard deviation) of P on the day of hCG administration was 1.02 ± 0.50 ng/ml and the live birth rate was 23.4%.

Table I Patient demographics, treatment protocol, pregnancy outcome and endocrine profile on the day of hCG administration (n= 2723).

<table>
<thead>
<tr>
<th>Patient demographics</th>
<th>Age (mean ± SD)</th>
<th>BMI (mean ± SD)</th>
<th>Primary cause for infertility (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30.5 ± 3.6 years</td>
<td>24.0 ± 4.6 kg/m²</td>
<td>Male factor 57.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tubal factor 9.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ovarian factor 8.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Endometriosis 4.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pelvic adhesions 28.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Genetic 0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Uterine anomaly 0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment protocol and pregnancy outcome (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hCG/IVF/IVF versus ICSI</td>
</tr>
<tr>
<td>rFSH/HMG</td>
</tr>
<tr>
<td>Day 3 ET/Day 5 ET</td>
</tr>
<tr>
<td>Single ET/double ET/triple ET</td>
</tr>
<tr>
<td>Clinical PR</td>
</tr>
<tr>
<td>Live birth delivery rate</td>
</tr>
<tr>
<td>Endocrine profile on the day of hCG administration (mean ± SD)</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>E₂</td>
</tr>
</tbody>
</table>

SD, standard deviation; BMI, body mass index; ICSI, intra-cytoplasmic sperm injection; IVF, in vitro fertilization; rFSH, recombinant follicle-stimulating hormone; HMG, human menopausal gonadotropin; ET, embryo transfer; hCG, human chorionic gonadotropin; P, progesterone; E₂, estradiol.

Our statistical analysis was performed using the Statistical software version 12 (StataCorp, College Station, Texas, USA) and SPSS version 20 (IBM, New York, USA).

Relationship between P levels on the day of hCG administration and live birth delivery rates

Figure 1 illustrates the live birth delivery rates for the various intervals of P levels on the day of hCG administration, while Fig. 2 shows the ORs (with a 95% confidence interval) for each interval when compared with the highest P groups (>1.5 ng/ml). Adjusted live birth rates were significantly higher for all groups with a P level <1.5 ng/ml except for the group with a P level ≤0.5 ng/ml. These results show that the live birth rates in groups with serum P levels >0.5 ng/ml on the day of hCG administration do not vary significantly from the hindered live birth rates in the group with P >1.5 ng/ml (P = 0.24).

To further assess for potential confounding factors and the effect of serum P levels on live follicular development, we applied the previously determined Power and upper limits to divide our sample in three groups: patients with low P (<0.5 ng/ml), normal P (0.5–1.5 ng/ml) and high P (>1.5 ng/ml) levels (Table II). As hypothesized before, patient age, the total dose of exogenous FSH administered, number of COCs retrieved, percentage of Day 3 embryo transfers and E₂ levels on the day of hCG administration were different among all these groups. Nonetheless, our analysis still showed that, when compared with the normal P group, patients with both low and high P had significantly lower live birth rates, even after adjusting for potential confounding variables (Table II). Conversely, maturation and fertilization rates did not vary significantly amongst the groups (Table III).

Discussion

This study comprehensively assessed the relationship between live birth delivery rates and P levels on the day of hCG administration during COS in a GnRH antagonist protocol. Furthermore, clinically relevant lower (<0.5 ng/ml) and higher (>1.5 ng/ml) P level limits between which live birth rates seem to be optimal were determined.

High progesterone levels and live birth rates

A previous study performed with a similar number of patients using GnRH antagonist down-regulation had already determined that P levels >1.5 ng/ml were detrimental for IVF pregnancy outcomes [n = 2855 (Bosch et al., 2010)]. However, these results were obtained using an unselected population and ongoing PR as the study outcome. The strength of our results come from the fact that they are less prone to bias, as the inclusion and exclusion criteria that were applied limited the possibility for unverified confounding factors. Furthermore, they...
Results

Figure 1: Crude (A) and adjusted (B) live birth delivery rates according to serum P levels on the day of hCG administration. Data are presented in percentage (95% CI), adjusted for patient age, number of COC retrieved, serum E2 level on the day of hCG administration, total dose of exogenous FSH administered, number of embryos transferred, and stage of embryo development on the day of embryo transfer; P: progesterone; hCG, human chorionic gonadotropin; CI: confidence interval; COC, cumulus—oocyte complexes; E2, estradiol; FSH, follicle-stimulating hormone.

Figure 2: Forest plot of crude (A) and adjusted (B) live birth rates according to serum P level on the day of hCG administration. Data are presented in OR (95% CI) of the comparison of the odds between each of the P levels with the highest P level interval (>=1.5 ng/ml), adjusted for patient age, number of COC retrieved, serum E2 level on the day of hCG administration, total dose of exogenous FSH administered, number of embryos transferred and stage of embryo development on the day of embryo transfer; *P < 0.05; P: progesterone; OR, odds ratio; CI, confidence interval; COC, cumulus—oocyte complexes; hCG, human chorionic gonadotropin; E2, estradiol; FSH, follicle-stimulating hormone.

are of more clinical relevance, since live birth delivery was the evaluated outcome.

The exact mechanisms that cause this P increase (still often misleadingly referred to as ‘premature luteinization’) remain a largely debated topic amongst researchers (Vaneechoutte et al., 2011; Bosch et al., 2010; Al-Azemi et al., 2012). Some believe that the excessive ovarian steroidogenic activity resulting from the stimulation of multiple follicles with exogenous FSH is the main source of late follicular P (Bosch et al., 2010; Elbashir, 2010). Nonetheless, alternative causes, namely increased LH production or an altered LH receptor sensitivity have been postulated for specific populations such as low responders (Fanchin et al., 1997a,b; Younis et al., 1998, 2001; Elbashir, 2010; Kildag et al., 2010; Bosch, 2011; Younis, 2011). In our study, patients with high P levels were younger (0.8 years), had a higher consumption of exogenous FSH, had more COC recovered and had higher E2 levels on the day of hCG administration (Table II). However, even after controlling for these factors, the live birth rates of the patients with high P levels remained significantly lower compared to the normal P group (Table III).

We found no significant difference with regard to oocyte maturation or fertilization rates amongst the high and normal P level groups. Previous
Table II: Comparison between low P (<0.5 ng/ml), normal P (0.5–1.5 ng/ml) and high P (>1.5 ng/ml) groups.

<table>
<thead>
<tr>
<th>Patient demographics</th>
<th>Total</th>
<th>Low P</th>
<th>Normal P</th>
<th>High P</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>2723</td>
<td>245</td>
<td>2129</td>
<td>344</td>
<td>NA</td>
</tr>
<tr>
<td>Age (mean ± SD)</td>
<td>30.5 ± 3.6</td>
<td>31.3 ± 3.4</td>
<td>30.6 ± 3.6</td>
<td>29.8 ± 3.6</td>
<td>&lt;0.001 ^1</td>
</tr>
<tr>
<td>BMI (mean ± SD)</td>
<td>24.0 ± 4.6</td>
<td>24.3 ± 4.6</td>
<td>23.9 ± 4.5</td>
<td>24.3 ± 4.7</td>
<td>0.186</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment protocol and outcomes</th>
<th>Total</th>
<th>Low P</th>
<th>Normal P</th>
<th>High P</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total FSH (IU)</td>
<td>169.4 ± 667.6</td>
<td>159.8 ± 617.5</td>
<td>187.6 ± 667.6</td>
<td>187.2 ± 672.6</td>
<td>&lt;0.001 ^1</td>
</tr>
<tr>
<td>Use of HP-HMG (%)</td>
<td>12.9</td>
<td>18.0</td>
<td>13.3</td>
<td>8.8</td>
<td>0.049 / &lt;0.01 ^2</td>
</tr>
<tr>
<td>COC retrieved</td>
<td>11.0 ± 6.7</td>
<td>7.4 ± 6.8</td>
<td>10.7 ± 6.4</td>
<td>14.9 ± 7.8</td>
<td>&lt;0.001 ^3</td>
</tr>
<tr>
<td>Day 3 embryo (%)</td>
<td>48.4</td>
<td>58.8</td>
<td>48.5</td>
<td>41.3</td>
<td>0.005 / 0.01 ^4</td>
</tr>
<tr>
<td>SET (%)</td>
<td>81.7</td>
<td>80.6</td>
<td>81.6</td>
<td>83.2</td>
<td>0.707 / 0.490 ^5</td>
</tr>
<tr>
<td>Maturity rate</td>
<td>79.2</td>
<td>79.2</td>
<td>79.5</td>
<td>77.8</td>
<td>0.160</td>
</tr>
<tr>
<td>Fertilization rate</td>
<td>73.7</td>
<td>74.4</td>
<td>73.3</td>
<td>75.6</td>
<td>0.364</td>
</tr>
<tr>
<td>Live birth rates</td>
<td>23.4</td>
<td>17.1</td>
<td>25.3</td>
<td>16.6</td>
<td>0.005 / 0.001 ^6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Endocrine profile on the day of HCG administration (mean ± SD)</th>
<th>E2</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1858.0 ± 1122.2</td>
<td>1223.2 ± 695.6</td>
<td>1304.6 ± 1054.8</td>
<td>2471.6 ± 1489.0</td>
<td>&lt;0.001 ^7</td>
<td></td>
</tr>
</tbody>
</table>

P, progesterone; NA, not applicable; SD, standard deviation; BMI, body mass index; IU, international units; HP-HMG, highly purified human menopausal gonadotrophin; COC, cumulative oocyte complexes; SET, single embryo transfer; E2, estradiol; FSH, follicle-stimulating hormone.

*P values for pairwise comparisons of group means in Table II were calculated using the Tukey-Kramer method.

Studies have also shown that oocyte and embryo quality seem to be comparable between high and normal P patients, while the endometrium significantly changes its gene expression profile (Labarta et al., 2011) and may have decreased receptivity (Fanchin et al., 1996; Chetkowski et al., 1997a; Fanchin et al., 1997b). Furthermore, studies performed in acceptors of donated oocytes found no difference in PR between oocytes deriving from stimulated cycles with P levels above or below 1.0–1.2 ng/ml (Check et al., 1994; Mielo et al., 2006), indicating that the high P levels may affect PR by altering endometrial receptivity.

**Low progesterone levels and live birth rates**

Until now, no prior trial has attempted to evaluate whether low P levels on the day of HCG administration may impair pregnancy following IVF or not. By comparing the live birth rates of patients with high P levels (>1.5 ng/ml) with the various ordinal regular groups of patients with decreasing values of P, we were able to conclude that patients with P levels <0.5 ng/ml were as detrimental to live birth as high P. To the best of our knowledge, the only trial with similar results (lower PR in the low P range) was performed by Levy et al. in 245 patients undergoing COS with GnRH agonist pituitary suppression (Levy et al., 1995). In this small study, clinical PRs were significantly lower in patients with P levels <0.7 ng/ml and over 0.8 ng/ml.

Patients in our low P group (<0.5 ng/ml) were older (0.7 years), had less COC retrieved and lower E2 levels on the day of HCG triggering. Nevertheless, low P levels were still associated with lower live birth rates even after adjusting for these and other confounding variables (Table II). More interestingly, patients were administered similar amounts of exogenous FSH units as the normal P group, were triggered with hCG using the same criteria and had unaffected Maturity and Fertilization rates (Table II). Thus, the lower live birth rates observed do not appear to be related to inadequate stimulation nor to poorer oocyte maturation or fertilization competence. These results differ from previous animal experiments using rhesus monkeys in which late follicular P seems to be essential for adequate oocyte maturation and fertilization (Zelinski-Weitoft et al., 1994; Hibbert et al., 1996; Borman et al., 2004).

As oocyte quality in patients with low P levels seems to be unaltered, one can postulate that the reduced live birth rate may be due to either decreased luteinization, altered endometrial receptivity or both. Basic science research on the menstrual cycle has shown that small amounts of P are extremely important during the late follicular phase (Hibbert et al., 1996; De Geyter et al., 2002; Borman et al., 2004). Prior to ovulation, the main source of P shifts from the adrenal cortex to the ovary (De Geyter et al., 2002) and, at that same time, P receptor mRNA and proteins become more abundant in the nuclei of the granulosa cells (Hibbert et al., 1998). Studies in rodents have shown that the administration of anti-progesterins in the periovulatory phase did not allow LH-induced luteinization of granulosa cells (Nattaj and Richards, 1993) and that only granulosa cells with P receptors in the pre-ovulatory phase were capable of luteinizing (Hibbert et al., 1998). Furthermore, endometrial P receptors reach their maximal concentrations during the mid- to late proliferative phase of the menstrual cycle (Barile et al., 1999;
Results

Lessey et al. (1988; Ravn et al., 1994) and P directly affects the action of E2 on the endometrium (Ozakli et al., 2004a, b). Studies that administered the anti-progesterin RU-486 during final oocyte maturation in mice showed that it affected endometrial receptivity and not embryo development (Batten et al., 1988). Furthermore, a prospective trial on women undergoing IVF showed that an insufficient endogenous P production can be an early predictor of poor pregnancy outcome which cannot be rescued by exogenous supplementation (Tian and others, 2005). Trials designed to specifically assess the endometrial development could shed further light on the mechanisms behind the decreased PR in patients with low serum P concentrations on the day of hCG administration.

Finally, the low P levels may also be a confounding factor of another mechanism that hinders simultaneously late follicular P production and pregnancy after IVF. Therefore, a randomized controlled trial in patients with low P levels (≤0.5 ng/ml) when at least three follicles of 17 mm are present compared immediate hCG administration with delayed triggering and serial P assessment might further explain the clinical importance of our findings.

Strengths, limitations and clinical impact of the study

To our knowledge, this is the first trial to comprehensively assess the effect of the full spectrum of serum P levels on the most important outcome of IVF, i.e., live birth delivery. Furthermore, this was performed in a large and homogenous sample.

The main limitation of our study is its retrospective nature, an issue also encountered in previously mentioned high P studies (Fanck et al., 1997; Zich et al., 2009; Kildsgaard et al., 2010; Wu et al., 2012; Xu et al., 2012). We attempted to diminish the importance of this weakness by adjusting for many potential confounders, but there could still be important variables which we might have failed to consider. Furthermore, our results are limited to patients under GnRH antagonist pituitary suppression and require confirmation in a GnRH agonist setting.

Regardless of its limitations, our research offers robust evidence that low late follicular P levels may be as detrimental as high P to live birth following IVF. Hence, the most important conclusion that can be extrapolated is that serum evaluation of late follicular P should not be ignored. Since 22% of all cycles included in our trial had either low or high P levels, a considerable number of cycles seemed hampered even before embryo transfer.

The knowledge that a patient has abnormal late follicular P could lead one to try to intervene to optimize IVF outcomes (Brashear, 2010). To that extent, limiting the total dose of FSH administered could be a beneficial way to address the issue of high trigger/day P levels (Bosch et al., 2010; Xu et al., 2012). For example, in two small previous randomized trials, late follicular replacement of daily FSH with low-dose hCG showed comparable PR (Ficici et al., 2005; Blockeel et al., 2009) without the detrimental late follicular P elevation (Ficici et al., 2005). Alternatively, close P monitoring and hCG triggering before reaching excessively high P levels (i.e., P levels >1.0 ng/ml) has also resulted in higher implantation rates in a small retrospective study (Harada et al., 1996).

Although other authors have suggested the use of HP-HMG to reduce the incidence of high late follicular P (Bosch et al., 2010), we would use this advice with caution, as the results of our study lead us to conclude that the exclusive use of HP-HMG could cause a higher incidence of similarly detrimental late follicular low P levels.

On the other hand, low hCG trigger/day P values also require close monitoring and, eventually, an FSH step-up or hCG trigger postponement to give the necessary stimulus to reach the minimal level of steroidogenesis. Furthermore, as a luteinization defect could be the cause for lower live birth rates in this population, a more intense or earlier luteal phase support could counterbalance the negative influence of low P. Finally, a randomized controlled trial comparing fresh embryo transfer and cryopreservation-thawing embryo transfer protocols could definitively obviate the need for further attempts to rescue what seems to be suboptimal fresh cycles (with both low- and high late follicular P levels), especially in an era where vitrification cryopreservation systems are becoming increasingly widespread (Bosch et al., 2010; Xu et al., 2012).

In conclusion, our study demonstrated, for the first time, that low P levels on the day of hCG administration also significantly impair live birth rates in women undergoing IVF with GnRH antagonist co-treatment followed by fresh embryo transfer. This should be carefully considered when treating infertile patients in order to aim towards P levels between 0.5 and 1.5 ng/ml in an attempt to maximize the reproductive outcomes in this population.

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Authors’ roles

S.S., N.P.P. and C.B. were responsible for the study conception, article drafting, statistical analysis and critical review. P.H. contributed during the study conception and aided with the statistical analysis. J.S., M.C. and H.T. participated in the study conception and review of the paper. C.B. was the research organizer.

Funding

No funding was received for this study.

Conflict of interest

The authors have no conflict of interest to declare.

References


Bosch E. Reply: elevated P levels on the day of hCG administration is related to FSH dose is it the whole truth? Hum Reprod 2011;26:499–500.


De Geyter C, De Geyter M, Huber PR, Neelelag E, Hodgrevve W. Progesterone serum levels during the follicular phase of the menstrual cycle originate from the crosstalk between the ovaries and the adrenal cortex. Hum Reprod 2002;17:933–939.

de Ziegler D, Bajou G, Chapron C. Pre-hCG elevation of plasma progesterone: good, bad or otherwise. Hum Reprod Update 2008;14:393.


Results

Study 1

79


Legro RS, Ary BA, Poulton RJ, Stanczyk FZ, Sauvage MV. Premature luteinization as detected by elevated serum progesterone is associated with a higher pregnancy rate in donor oocyte in vitro fertilization. Hum Reprod 1993;8:1505–1511.


Moffatt DI, Queenan JT, Jr, Shaw R, Musater SJ. Progesterone levels on the day of human chorionic gonadotropin do not predict pregnancy outcome from the transfer of fresh or cryopreserved embryos from the same cohort. Fertil Steril 1997;67:396–391.


Younis J. Increased P level on the day of hCG administration is related to FSH closer: is it the whole truth? Hum Reprod 2011;26:498–499.


7.1.2 Study I-B

Assess whether late-follicular P affects embryo quality and, concomitantly, CLBR

Racca A, Santos-Ribeiro S, De Munck N, Mackens S, Camus M, Verheyen G, Tournaye H, Blockeel C. Impact of late follicular phase elevated serum progesterone on cumulative live birth rates: is the endometrium the only responsible factor?

Proof of full manuscript submitted for peer review, accepted following revision in Hum Reprod

Abstract

Study question: Does elevated late follicular phase serum progesterone (LFEP) have a deleterious impact on embryo quality?

Summary answer: High serum progesterone (P) levels at the end of the follicular phase are detrimental for embryo quality, as well as for embryo utilization rate. Cumulative live birth rates (CLBR) are lower in women with high serum P levels regardless of the number of cumulus-oocyte complexes (COC) retrieved.

What is known already: Exogenous ovarian stimulation promotes the excessive production of P during the follicular phase, which adversely affects live birth rates (LBRs) during in-vitro fertilization (IVF), presumably by hindering endometrial development. While one can postulate that a freeze-all strategy would circumvent this issue, this hypothesis has been insufficiently validated.

Study design, size, duration: A single-centre retrospective cohort analysis of all gonadotropin-releasing hormone (GnRH) antagonist down-regulated cycles performing ICSI as an insemination method and followed by a fresh embryo transfer between 2010 and 2015 was done. The sample was stratified according to the following P levels on the day of ovulation triggering: ≤0.50 ng/mL, 0.51-1.49 ng/mL and ≥1.50 ng/mL. Our primary outcomes were embryo development, LBR and cumulative LBR after the fresh and subsequent frozen embryo transfers.

Participants/materials, setting, methods: 3400 cycles with GnRH antagonist suppression and fresh embryo transfer using ICSI were included. We performed multivariable generalized estimating equation regression analysis to account for potential confounding.

Main results and the role of chance: Between-group comparisons based on P level (≤0.50 ng/mL, 0.51-1.49 ng/mL and ≥1.50 ng/mL) showed that late follicular E2 levels and the
number of COCs retrieved increased significantly with increasing trigger P serum values. A significantly higher maturation failure was found in the low P level group when compared with the medium P (0.51-1.49 ng/mL) group. The embryo utilization rate, defined as the total number of embryos transferred and cryopreserved per fertilized oocyte, decreased on both day 3 (72.3%, 63.0% and 45.4%, respectively) and day 5 (48.8%, 47.8% and 38.8%, respectively) as serum P levels increased. Furthermore, although the overall CLBRs were between 50% and 60%, the CLBR increased at a slower pace with increasing P levels, thus curtailing the benefit of the increasing number of oocytes on CLBR.

Limitations, reasons for caution: The main limitation of our study was its retrospective nature. Furthermore, it was restricted to patients under GnRH antagonist suppression.

Wider implications of the findings: The impact of LF EP raises the question whether the freeze only strategy is enough to mitigate the detrimental effect on the implantation potential. We believe that further studies are required in order to find possible solutions to minimize the risk and consequence of premature P rise such as lowering the stimulation dose or applying a step-down protocol.

Study funding/competing interest(s): No funding was received for this study and the authors have no conflicts of interest to declare.

Introduction

Although significant improvements have been achieved in assisted reproductive technologies (ART), delivery rates remain around 30% per embryo transfer (Castello, et al., 2016, Kupka, et al., 2016). Thus, as multiple treatment attempts are generally required to achieve a live-birth [LB, (Malizia, et al., 2009)], increasing efforts have been made to report in-vitro fertilization (IVF) success rates as cumulative live-birth rates (CLBR) to provide patients with better prognostic information and more meaningful counselling (Viardot-Foucault, et al., 2015).

The question whether the supra-physiological increase of serum progesterone (P) during ovarian stimulation (OS) may influence the success of ART has been a matter of debate over the last 30 years. While a large number of clinical studies suggest that there is no association between P levels and pregnancy rates (Edelstein, et al., 1990, Silverberg, et al., 1991, Check, et al., 1994, Givens, et al., 1994, Bustillo, et al., 1995, Levy, et al., 1995, Ubaldi, et

The possibility to predict live birth delivery rates after the fresh cycle, the first frozen embryo transfer (FET) and CLBR (fresh plus all eventual FETs until the first live-birth) has two clinically important consequences: first, physicians could estimate in the short and long term the likelihood of success of ART, providing more precise information to couples; secondly, it would be possible to determine better to what extent EP influences both ER and EQ. Therefore, the aim of the current study was to investigate the influence of high P levels on embryo delivery potential by evaluating its effect on LBR after a fresh embryo transfer cycle, the first FET and CLBR, controlling especially for ovarian response.

Materials and methods

Study Population and Design

This was a retrospective, single-centre cohort study including all women who underwent IVF at our centre between the January 1st 2010 and the December 31st 2015. Approval to retrieve and analyse the data was provided by the Ethics Committee of Brussels University
Hospital (BUN 143201733383). We included only patients who performed OS under gonadotropin-releasing hormone (GnRH) antagonist suppression. In order to evaluate the oocyte maturation rate and minimize the effect of the known differences in terms of fertilization between conventional IVF and intra-cytoplasmatic sperm injection (ICSI) (van der Westerlaken, et al., 2005, Abdalmageed, et al., 2015), we included only patients who underwent ICSI. Patients who performed both conventional IVF and ICSI within the same cycle were also excluded from the study. Furthermore, women with an unknown LB outcome, who underwent GnRH agonist triggering, who were acceptors of donated oocytes, who performed embryo biopsy for pre-implantation genetic diagnosis or freeze-only cycles were disregarded.

Ovarian stimulation, pituitary suppression, cycle monitoring and ovulation triggering

As previously described (Santos-Ribeiro, et al., 2014), ovarian stimulation was initiated on day 2 of the menstrual cycle with either recombinant FSH (rFSH; Gonal-F®, Puregon®, or Elonva®) or highly purified human menopausal gonadotrophin (hp-hMG; Menopur®). Pituitary down-regulation was performed with daily administration of a GnRH antagonist [either cetrorelix (Cetrotide®) or ganirelix (Orgalutran®)] starting from day 7 of the menstrual cycle onwards. Whenever necessary, dose adjustments of rFSH/hp-hMG were performed according to ovarian response. As soon as three follicles ≥ 17 mm were observed during pelvic ultrasound, final oocyte maturation was triggered with either highly purified urinary (5000 IU or 10000 IU, according to body weight and the physician preference; Pregnyl®) or 250 IU of recombinant hCG (Ovitrelle®). Cumulus-oocyte complexes (COCs) were collected by transvaginal aspiration 36 h after the hCG administration. Intravaginal P (Utrogestan®) was administered daily (3 times 200 mg per day) for luteal phase support.

Insemination, Embryo Quality assessment and cryopreservation

The insemination of the collected oocytes was performed via ICSI. Fertilization was assessed 16-18 h after injection by the presence of two pronuclei and, from then onwards, embryo development was evaluated daily until the transfer or cryopreservation of cleavage stage embryos (day 3) or blastocysts (day 5 and 6). Supernumerary embryos of good quality were cryopreserved by means of vitrification using a closed vitrification...
device with high-security straws (CBS-ViT-HS®; Cryobiosystems) in combination with dimethyl sulfoxide and ethylene glycol as cryoprotectants (Irvine Scientific Freeze Kit®; Irvine Scientific). EQ was classified similar to what is described in a previous study performed by De Munck, et al. (2015), with a minor update in the classification (good quality embryos included up to <50% fragmentation). Day 3 embryos were evaluated on the basis of the number and symmetry of the blastomeres, percentage of fragmentation, vacuolization, granulation, and multinucleation. Based on all these parameters, an EQ score was assigned to all normally fertilized embryos using a predefined algorithm, which divided them into four categories: excellent, good, moderate, or poor. On day 3, fresh embryos were considered eligible for transfer if at least four blastomeres were present with a maximum of 50% fragmentation. Blastocysts were scored according to the grading system developed by Gardner and Schoolcraft (1999) based on (i) the expansion stage, (ii) the number of cells joining the compaction or blastulation, and (iii) the appearance of the trophectoderm (TE) and inner cell mass (ICM). One of the following blastocyst quality scores was given to all day 5 or day 6 embryos: excellent, good, moderate, or poor. On day 5, fresh embryos were considered eligible for transfer if they reached at least the fully compacted stage. The utilization rate was calculated based on all embryos that were transferred or cryopreserved.

The following embryos were considered eligible for cryopreservation: day 3 embryos with ≥6 blastomeres and ≤50% fragmentation; day 5 and day 6, fully expanded or hatching blastocysts with a type A/B/C ICM and type A/B for TE. After warming (Irvine Scientific Thaw Kit®), day 3 embryos were cultured overnight in blastocyst medium until they were transferred the following day (day 4). Survival was scored based on the number of surviving blastomeres. If >2 cells degenerated after warming, a surplus embryo was
warmed if available. An embryo score was given on day 4 at the moment of transfer: excellent, good, moderate or poor. Vitrified blastocyst (day 5 or day 6) embryo transfers were performed on the day of warming. The survival rate was evaluated after warming based on the survival of all cells and the re-expansion of the blastocyst. An embryo score was given at the moment of transfer: excellent, good, moderate or poor (Table 2, p. 84).

**Endometrial preparation for the FET**

The FET took place in either a natural or an artificial cycle. In a natural cycle (NC), ovulation occurred either spontaneously or artificially triggered (with the use of hCG). In artificially supplemented cycles (HRT), preparation of the endometrium consisted of sequential administration of E₂ valerate (Progynova®) and micronized vaginal P [Utrogestan®, (van de Vijver, et al., 2014)].

**Progesterone assessment immunoassay**

The serum P levels were assessed on the day of hCG administration using a validated electrochemilluminescence immunoassay (Cobas 6000®, Roche) with a measured sensitivity and total imprecision (coefficient of variation) of 0.03 μg/l and <7%, respectively. The assay was regularly calibrated to minimize variation of the results associated with time and reagent batch renewal.

**Main outcome measures**

Our main outcome parameters were embryo utilization rate, live birth delivery rate for the fresh cycle, LBR for the first FET after a failed fresh ET and CLBR.

Embryo utilization rate was defined as the total number of embryos transferred or cryopreserved per number of fertilized oocytes. LBR was defined as the delivery of a live born (>24 weeks of gestation) according to the criteria defined by International Committee for Monitoring Assisted Reproductive Technology [ICMART, (Zegers-Hochschild, et al., 2017)]. LBR after the first failed fresh embryo transfer was defined as the delivery of a live born (>24 weeks of gestation) in the first FET cycle and CLBR was defined as the delivery of a live born (>24 weeks of gestation) in either the fresh or one of the subsequent frozen–
thawed cycles using solely embryos derived from one OS cycle. For the CLBR each ovarian stimulation cycle was considered separately and only the first delivery was considered for the analysis.

Statistical analysis

Patients were categorized into three groups according to the following P levels on the day of ovulation triggering: \( \leq 0.5 \text{ ng/mL} \), \( 0.51-1.49 \text{ ng/mL} \) and \( \geq 1.5 \text{ ng/mL} \). These thresholds were selected according to the cut-offs applied amongst previous studies (Bosch, et al., 2010, Xu, et al., 2012, Santos-Ribeiro, et al., 2014) and a balance between the use of clinically relevant intervals that would divide our sample homogenously so as not to hinder the inter-interval statistical analysis. These groups were compared using multivariable generalized...
estimating equation (GEE) regression modelling (to account for the eventual clustering of cycles performed by the same patient) followed by exploratory pairwise comparisons whenever warranted. Additionally, maturation, fertilization and utilization rates were also compared using mixed-effects multilevel regression analysis, accounting for correlated outcomes within fresh cycles performed in the same patient and also oocytes deriving from the same ICSI cycle.

LB outcomes were assessed using multivariable GEE regression analysis including the P levels categories as a predictor along with other potential confounders in each model according to the outcome being evaluated. We chose to adjust for potentially confounding variables whenever they previously known to affect pregnancy rates or if they varied significantly between the study groups. That said, we adjusted for female age and body mass index (BMI), reasons for infertility (male factor, tubal factor, ovarian factor, endometriosis, uterine factor, other female factor or otherwise unexplained infertility), number of COCs retrieved and the embryo developmental stage at transfer (cleavage or blastocyst stage) in all regression models. Moreover, for fresh LB we included also the total dose of ovarian stimulation, the circulating E2 on the day of ovulation triggering, EQ (EQ1, EQ2, EQ3-4) and the number of embryos transferred (single embryo versus multiple embryos). For the first frozen-thawed embryo transfer (ET) cycle, after a failed fresh embryo transfer, we adjusted additionally for number of embryos transferred and the type of FET cycle preparation (NC with spontaneous LH peak, NC with ovulation triggered using hCG, HRT).

Given the potential moderating effect of the number of oocytes retrieved on the effect of EP on CLBR, we also did two supplemental analyses. Firstly, we performed sensitivity analyses evaluating the effect of each P level on CLBR. Then, we repeated the GEE regression analysis for CLBR detailed before, however, this time adding an interaction term between serum P and the number of oocytes retrieved (Venetis, et al., 2015).

**Results**

A total of 3400 autologous ICSI cycles (performed in 2492 patientso were included in the analysis (Figure 8, p. 86). Patients’ baseline characteristics are summarized in Table 3 (p. 88) while information regarding cycle and crude pregnancy outcomes are reported in Table 4 (p. 89).
Results (Study 1–B)

P level influence on live birth delivery rate and cumulative live birth rates

Women in the high P level group were generally younger and had better ovarian response, compared to medium and low P level groups (Table 3, p. 88). Furthermore, between-group comparisons based on P level (≤0.5 ng/mL, 0.51-1.49 ng/mL and ≥1.5 ng/mL) showed that
late follicular E2 levels (1049 pg/mL, 1509 pg/mL and 1976 pg/mL respectively, p<0.001) and number of oocytes retrieved (5, 7, 10 respectively, p<0.001) increased with increasing trigger P levels. Furthermore, even though the maturation rate was comparable among the three groups, the rate of cycles with no mature oocyte was significantly higher in the low P group compared to the medium one. With regard to the embryo development, there were less embryos on day 3 for the low and high P groups. In terms of the utilization rates, while these seemed to decrease linearly as P increased on day 3 (72.3%, 63.0% and 45.4%, respectively), only the high progesterone group remained associated with a significant decrease in utilization rates on day 5 (48.8%, 47.8% and 38.8%, respectively).

Table 4 (p. 89) summarizes the crude pregnancy outcomes of our sample. In the unadjusted analysis, LBRs for the fresh cycle seemed to be significantly lower for the low P group when compared to both the medium and high P groups (18.2%, 28.7% and 24.9% per started cycle, respectively). Furthermore, the unadjusted CLBR also were negatively affected by low P levels when compared with medium and high P (24.3%, 36.5% and 35.1%, respectively). However, the adjusted analysis revealed that the high P group was associated with both hindered fresh LBR and CLBR. Figure 9 (p. 90) illustrates the ORs

<table>
<thead>
<tr>
<th>Fresh cycle</th>
<th>50.50 (n=478)</th>
<th>0.51-1.49 (n=2657)</th>
<th>21.50 (n=265)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single embryo transfer</td>
<td>192/404 (47.5%)</td>
<td>1174/2430 (48.3%)</td>
<td>122/232 (52.6%)</td>
<td>0.382</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Live birth rates</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Per started cycle</td>
<td>87/478 (18.2%)</td>
<td>762/2657 (28.7%)</td>
<td>66/265 (24.9%)</td>
<td>&lt;0.001*†</td>
</tr>
<tr>
<td>Per transfer</td>
<td>87/404 (21.5%)</td>
<td>762/2430 (31.4%)</td>
<td>66/232 (28.5%)</td>
<td>&lt;0.001*†</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>First FET cycle after a failed fresh embryo transfer</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No embryo remaining</td>
<td>201/317 (64.4%)</td>
<td>949/1668 (56.9%)</td>
<td>82/166 (49.4%)</td>
<td>0.014*†</td>
</tr>
<tr>
<td>No embryo surviving thawing</td>
<td>7/116 (6.0%)</td>
<td>34/719 (4.7%)</td>
<td>6/84 (7.1%)</td>
<td>0.578</td>
</tr>
<tr>
<td>Single warmed embryo transfer</td>
<td>59/109 (54.1%)</td>
<td>379/685 (55.3%)</td>
<td>43/78 (55.1%)</td>
<td>0.967</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Live birth delivery rates</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Per started FET</td>
<td>20/116 (17.2%)</td>
<td>146/719 (20.3%)</td>
<td>18/84 (21.4%)</td>
<td>0.685</td>
</tr>
<tr>
<td>Per transfer</td>
<td>20/109 (18.4%)</td>
<td>146/685 (21.3%)</td>
<td>18/78 (23.1%)</td>
<td>0.683</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cumulative live birth rates (fresh and all subsequent FETs until first live birth)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Live birth delivery rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per started cycle</td>
<td>116/428 (24.3%)</td>
<td>970/2657 (36.5%)</td>
<td>93/265 (35.1%)</td>
<td>&lt;0.001*†</td>
</tr>
</tbody>
</table>

Table 4 - Embryo transfer and LBR for fresh cycles, first frozen-thawed cycles and CLBR
p< 0.05 for the following pairwise comparisons: *50.50 vs 0.51-1.49, †50.50 vs 21.50 or 30.51-1.49 vs 21.50.

Table 4 (p. 89) summarizes the crude pregnancy outcomes of our sample. In the unadjusted analysis, LBRs for the fresh cycle seemed to be significantly lower for the low P group when compared to both the medium and high P groups (18.2%, 28.7% and 24.9% per started cycle, respectively). Furthermore, the unadjusted CLBR also were negatively affected by low P levels when compared with medium and high P (24.3%, 36.5% and 35.1%, respectively). However, the adjusted analysis revealed that the high P group was associated with both hindered fresh LBR and CLBR. Figure 9 (p. 90) illustrates the ORs
Results (Study 1 - B)

90 (with 95% confidence intervals) for the evaluated pregnancy outcomes for low and high levels of P compared to the medium group.

A sensitivity analysis was initially performed according to the number of oocytes retrieved (Figure 10, p. 91). In this analysis, we generally noted that the slope of increase in pregnancy outcomes tapered as P levels raised. Specifically, while the low P level pregnancy outcomes improved steeply as the number of COCs increased, this positive effect of oocyte retrieval rate was not as pronounced for the normal and then high P groups. This observation of the unadjusted trends in CLBR generated the hypothesis that the detrimental effect of P on CLBR might be moderated by the number of oocytes retrieved. In order to confirm this hypothesis, we performed a multivariable regression analysis adjusting for late-follicular P, number of COC retrieved, female age and day of embryo transfer, adding the interaction of late-follicular P and number of COC retrieved in the model as well (Venetis, et al., 2015). This analysis confirmed that EP seems to curtail the benefit of an increasing number of oocytes on CLBR (interaction coefficient -0.06 with 95% confidence interval from -0.09 to -0.03; p<0.001).

Figure 9 – ORs for late-follicular P and pregnancy outcomes, (unadjusted OR on the left and adjusted OR on the right)

LBR was assessed using multivariable GEE regression analysis including the P levels categories as a predictor along with other potential confounders in each model according to the outcome being evaluated. Specifically, for fresh LB we included also the embryo stage at transfer (cleavage or blastocyst stage), EQ (EQ1, EQ2, EQ3-4), number of embryos transferred (single embryo versus multiple embryos) and the continuous variables female age and number of COCs retrieved. For the first FET after a failed fresh embryo transfer we adjusted female age, number of COC retrieved, number of embryos transferred and the type of FET cycle preparation (NC with spontaneous LH peak, NC with ovulation triggered using hCG, artificial). Finally, for CLBR, we controlled for the following variables: female age, number of COC retrieved and stage of the embryos transferred.

P level influence on live birth delivery rate and cumulative live birth rates when controlling for ovarian response

Figure 9 – ORs for late-follicular P and pregnancy outcomes, (unadjusted OR on the left and adjusted OR on the right)
Discussion

The study clearly demonstrated that P level rises were associated with a decrease in embryo utilization rate, for both cleavage and blastocyst stage embryos. Furthermore, ovarian response correlated with high serum P levels, which theoretically would be associated with better CLBR (Drakopoulos, et al., 2016) as noted at first glance in our unadjusted analysis. However, a subsequent multivariable regression analysis confirmed that CLBR was lower in the high P level group regardless of the number of COCs retrieved. To the best of our knowledge, this is the first study evaluating simultaneously the influence of EP on both EQ and CLBR.

Late follicular P elevation during ovarian stimulation is a complex event with an unclear underlying mechanism. Several researchers have eloquently detailed the potential explanations for this phenomenon, such as the compounding effect of otherwise normal production of P during late-follicular development (Venetis, et al., 2007) or the supraphysiologic exposure of mature granulosa to FSH (Bosch, et al., 2003). More recently (Oktem, et al., 2017), provided a molecular explanation why serum P levels may be elevated prior to ovulation trigger in stimulated IVF cycles. Summarily, they described that the FSH activity promotes P synthesis in-vitro directly from granulosa cells without luteinization by up-regulating the expression and increasing enzymatic activity of 3β-hydroxysteroid dehydrogenase, the enzyme that converts pregnenolone to P in the human granulosa cells.
Our data showed that the high P level group had the lowest embryo utilization rate and the highest ovarian response, compared to medium and low P level groups. These contradictory trends indicated a poorer EQ in the high P level group. This finding is consistent with the recent study of Vanni, et al. (2017) which demonstrated that EP on the day of triggering is associated with decreased top quality blastocyst formation rate. Similarly, Huang, et al. (2016), demonstrated that EP has a negative impact on top-quality embryo rate in IVF cycles. Nevertheless, these previous analyses did not evaluate the embryo live birth delivery potential. In the present study, a step further has been performed by including both cleavage and blastocyst stage embryos in the analyses; furthermore, we did not exclude patients based on EQ or ovarian response, and we evaluated LBR and CLBR, thus increasing the clinical usefulness of our results in terms of patient counselling. We evaluated the first FET specifically with the aim of looking at the differences between the effects of P on ER and EQ. We confirmed that LBRs were lower in the high P level group for fresh cycles, while failing to find any difference between the groups in the first FET. These results are in line with the hypothesis that freezing might help in reducing the embryo-endometrial asynchrony. Indeed, as previously described by Healy, et al. (2017), the deleterious effect of high P on the fresh cycle increased in the day 6 embryos when compared to day 5; however, this difference did not persist in the subsequent FET cycles.

Thus, we have come to observe that, despite having on average a better ovarian response and a younger age, the high P group still fails to reach the best LBR and CLBR. Until now, most investigators evaluated the influence of high P level focused on fresh embryo transfer cycles (Santos-Ribeiro, et al., 2014, Venetis, et al., 2015) with only one study exploring the influence of serum P elevation on CLBR (Bu, et al., 2014). In fact, Venetis, et al. (2015), established that the number of oocytes retrieved and late-follicular P levels affect IVF outcomes independently, by failing to find any effect modifying interaction using multivariable logistic regression. However, given that CLBR accounts for the net-effect of P and ovarian response over multiple embryo transfers, we were especially interested to see whether this hypothesis held true using CLBR as the outcome of interest. Contrary to the analysis by Venetis, et al. (2015) for LBR, our analysis suggested that EP seems to curtail the benefit of an increasing number of oocytes on CLBR, by an effect modifying interaction. Specifically, in our unadjusted analysis we showed a superior ovarian response in women with EP, nonetheless, taking into account that this group was favoured by the advantage of the young age, the overall CLBR outcomes end up being non-superior.
to the medium P level group. However, when we performed the adjusted analysis we found a significant decrease in the CLBR for the high P level group, compared with the medium P level, potentially owed to the detrimental effects of EP on the ER and EQ.

The main weakness of the study is its retrospective nature that might lead to bias the interpretation of the results. In order to circumvent as much as possible this weakness, we adjusted for multiple potential confounders known to affect IVF pregnancy outcome. Furthermore, despite being suggested as the most suitable benchmark for an IVF programme, the use of CLBR as a pregnancy outcome is challenging given the lack of universal agreement on which numerators and denominators should be used to calculate IVF pregnancy rates (Maheshwari, et al., 2015). We chose to evaluate only the first LB of all embryos deriving from a single oocyte retrieval cycle, given that we wanted to assess how P production during the ovarian stimulation impacted on that cycle specifically. That said, our results should not be extrapolated to subsequent cycles. Another limitation of the study is that in our database we did not have information concerning the smoking status of the patients, which can have an important influence on the IVF outcome. Lastly, we only included patients under GnRH antagonist suppression.

Could the freeze only strategy circumvent the negative impact of elevated serum progesterone on the implantation potential?

Premature serum P level rise in stimulated IVF cycles adversely affect implantation and ART outcomes by impairing ER (Gellersen and Brosens, 2003, Labarta, et al., 2011, Liu, et al., 2017, Xiong, et al., 2017), leading one to consider that this issue may be amenable by elective embryo cryopreservation. However, the negative influence that high P levels may have on EQ and CLBR, despite the possible higher number of oocytes retrieved, may put this generalized assumption into question. In light of the above, the multidirectional impact of EP levels on IVF may deem the freeze-all strategy insufficient to solve the problem at hand. We believe that further studies are required in order to find possible solutions to actually prevent premature P rise.

Previous authors have proposed a number of potential solutions to both avoid premature P rise and the consequences of EP on ER. Specifically, a first potential solution to reduce P production could be to lower the total dose of ovarian stimulation (Bosch, et al., 2010, Venetis, et al., 2016). Although this strategy would likely reduce P production, the
possibility that this approach could also reduce the number of oocytes retrieved should also be taken into consideration (Alper and Fauser, 2017). In fact, although the previous evidence has been inconsistent (Out, et al., 2000, Hoomans and Mulder, 2002, Klinkert, et al., 2005, Tan, et al., 2005, Sterrenburg, et al., 2011), some studies have shown a direct association between the gonadotropin dose and the number of oocytes retrieved. If this relationship is to be confirmed, while one could make the argument that a small decrease in ovarian stimulation and oocyte retrieval rates might have little direct effect in the immediate fresh embryo transfer (Sunkara, et al., 2011), it may still hinder long term CLBR, as recent studies have demonstrated (Ji, et al., 2013, Drakopoulos, et al., 2016, Vaughan, et al., 2017). In this context, it is possible that a step-down approach would be a better alternative, given the fact that both P’s production and function seem to be concentrated on the late-follicular phase (Massin, 2017). However, this hypothesis requires further confirmation. Alternatively, another approach to prospectively evaluate the effect of the total dose of stimulation on EQ and ER would be to perform a cross-over study in which oocyte donors would undergo two stimulation cycles with different ovulation triggering points (e.g. when at least 3 follicles of 17 mm versus 20 mm are present). This would effectively prolong the period of OS associated with the largest amount of P production and evaluate its effect on ER (by performing endometrial biopsies in both scenarios) and EQ (by comparing the utilization rate and, potentially, pregnancy outcomes of the oocyte recipients), separately. This study may assist physicians to finally establish the balance between the highest possible number of oocytes retrieved at the lowest possible level of circulating serum P.

In conclusion, our analysis offers evidence that high serum P levels at the end of the follicular phase may be associated with decreased EQ, as well as a lower embryo utilization rate.
7.1.3 Study 1-C

Evaluate whether the variation in P exposure throughout COS can predict ART pregnancy outcome in a fresh cycle better than a single measurement


Proof of full manuscript submitted for peer review

Abstract

Objective: to evaluate whether the variation in P exposure during ovarian stimulation (OS) may predict live birth rates (LBR) following a fresh in-vitro fertilization (IVF) transfer cycle.

Design: retrospective cohort study.

Setting: university-based tertiary referral centre.

Patient(s): patients who underwent OS between 2010 and 2015 for IVF followed by a fresh embryo transfer. All cycles (n=4312) were performed using a GnRH antagonist for pituitary suppression. Final oocyte maturation and ovulation were triggered with exogenous human chorionic gonadotropin (hCG) approximately 36 hours before oocyte retrieval. We calculated the relative progesterone variation during OS (PV-OS, defined as the subtraction between the level of P on the day of hCG administration and at the start of OS) and the duration of late-follicular elevated P (LFEP) >1.00 ng/mL and >1.50ng/mL. Using multivariable regression, we assessed the predictive value of both PV-OS and the duration of LFEP for LBR after IVF.

Intervention(s): none.

Main Outcome Measure(s): LBRs.

Results: the mean (±standard deviation) PV-OS was 0.29±0.37 ng/mL and the overall LBR was 25.2%. LBR decreased linearly with increasing PV-OS, with cycles in which PV-OS was >0.40 ng/mL performing significantly worse. This relative increase of P exposure was especially relevant for those women below the conventionally applied hCG-day P thresholds, between whom PV-OS revealed an important subgroup of women with a
Results

(detrimentally elevated P exposure which would have remained otherwise unnoticed if only the hCG-day P levels were to be assessed. Conversely, a universal cut-off for the duration LFEP was a poor predictor of IVF pregnancy outcomes when LBR was the outcome of interest.

Conclusion(s): our results suggest that evaluating PV-OS may reveal a detrimental effect of elevated follicular-phase P exposure where hCG-day P levels alone may not, adding crucial insight into the current limitations of isolated evaluations of P during OS to predict IVF outcome.

Introduction

An increasing number of researchers have alluded to the potential benefit of distancing the transfer of embryos produced during assisted reproductive technologies (ART) away from the ovarian stimulation (OS) cycle (Blockeel, et al., 2016, Coutifaris, 2016). The scientific evidence frequently used to justify this recent trend of the so-called elective “freeze-all strategy” includes randomized controlled trials (RCTs) showing an increase in live birth rates (LBR) after elective frozen embryo transfers (FETs) (Shapiro, et al., 2011, Roque, et al., 2013, Chen, et al., 2016) and large population studies associating FETs with a lower incidence of prematurity and low birth weight (Pelkonen, et al., 2010, Kalra, et al., 2011, Wennerholm, et al., 2013, Ishihara, et al., 2014).

Many authors have postulated that the supraphysiologic milieu of hormones produced during OS may affect endometrial receptivity (ER) and hinder both embryo implantation and neonatal outcomes (Martinez-Conejero, et al., 2007, Horcajadas, et al., 2008, Pinborg, et al., 2010, Labarta, et al., 2011, Imudia, et al., 2012, Kalra, 2012, Imudia, et al., 2013, Evans, et al., 2014, Haouzi, et al., 2014, Coutifaris, 2016). Of all the candidate biomarkers evaluated during the monitoring of OS, the abnormal production of progesterone (P) during the later stages of the follicular phase has been, thus far, proposed most frequently as the best surrogate for ER during a fresh cycle (Bosch, et al., 2010, Venetis, et al., 2013, Hill, et al., 2017). Specifically, late-follicular elevated progesterone (LFEP, most frequently defined as a P >1.50 ng/mL) has been linked to lower pregnancy rates in a fresh embryo transfer (Venetis, et al., 2013). Although the mechanism behind this hindered outcome remains controversial (Melo, et al., 2006, Healy, et al., 2017), some consider that this issue may be, at least in part, amenable by elective embryo cryopreservation (Yang, et al., 2015, Healy, et
al., 2016, Vanni, et al., 2016), given that LFEP has been associated with and abnormal endometrial gene expression in the luteal phase (Labarta, et al., 2011, Van Vaerenbergh, et al., 2011). This has led many fertility centres to change their clinical practice and to measure serum P levels on the day of human chorionic gonadotropin (hCG) administration, adopting a freeze-all strategy whenever the threshold of 1.5 ng/mL is exceeded.

Nonetheless, the everyday use of LFEP as an ART outcome predictor in current clinical practice has proven to be harder than originally expected, owing mostly to the fact that a) P production may not affect pregnancy outcomes linearly (Bosch, et al., 2010, Xu, et al., 2012, Santos-Ribeiro, et al., 2014) and b) it is frequently encountered in good-prognosis women with an otherwise healthy multi-follicular response (Fleming, 2008, Younis, 2011). Furthermore, previous studies have postulated that the detrimental effect of circulating P may be set into motion even when P is below 1.50 ng/mL (Roque, et al., 2015). Altogether, these conflicting studies are a testament of the limitations of using a single P assessment during OS to decide whether to defer embryo transfer or not. Many investigators have attempted to enhance the predictive capacity of this single measurement by using ratios such as P-to-follicle (Shufaro, et al., 2015), P-to-oocyte (Hill, et al., 2017), P-to-estradiol (Lai, et al., 2009, Wu, et al., 2012) ratios. However, such indexes have failed to perform superiorly, thus limiting their routine clinical practice, an observation that has led some to propose that clinicians consider disregarding endocrine monitoring of P during OS altogether (Martinez, et al., 2016).

Interestingly, Huang, et al. (2012) concluded in their retrospective analysis that the duration of LFEP (defined as a universal P threshold of >1.0 ng/mL during 0, 1-2 or ≥3 days prior to ovulation triggering) seemed to play a major role in how circulating P affected pregnancy outcome. This approach is unique in the sense that it was an interesting departure from the previous literature, deviating focus from the question “how high P can reach?” to “how long can the endometrium be exposed to P above a certain universal threshold?”. As P receptors are present in endometrium in all phases of the menstrual cycle, namely during menses and the early follicular phase (Ingamells, et al., 1996, Lau, et al., 1996, Marshburn, et al., 2005), it would be reasonable to assume that P exposure might be detrimental even when occurring days before ovulation triggering. Furthermore, this concept of a hindering cumulative P exposure is more in line with the current knowledge of how P regulates endometrial development, as it is generally thought that the ovaries need to increase (Franasiak, et al., 2016) and maintain (Coomarasamy, et al., 2015) the
secretion of P above a certain level in order initiate and sustain the development of a secretory endometrium. However, whether the threshold to trigger these endometrial changes is universal (e.g. 1.50 ng/mL, regardless of the patient’s baseline P exposure) or relative (i.e. a difference between the amount of P produced and the usual baseline circulating P concentration encountered in in each patient) is unknown.

While being a potentially useful hypothesis (Venetis, et al., 2013, Sonigo, et al., 2014), the approach performed by Huang, et al. (2012) has been assessed only using clinical pregnancy rates (CPR) as the main outcome measure and has been subject to insufficient validation (Sonigo, et al., 2014, Demir, et al., 2016). Furthermore, failures to completely reproduce their results (Dai, et al., 2015) allude to the possibility that the effect of P exposure during OS on ART pregnancy outcomes may be incompletely captured when applying a universal threshold. Specifically, one of the reasons for these conflicting results may stem from the fact that this study elected a universal arbitrary cut-off of 1.00 ng/mL, which is frequently found at the end of the follicular phase of unstimulated cycles as well (Stricker, et al., 2006). Conversely, one could also hypothesize that evaluating how much P varied relatively to the basal circulating P levels of that particular patient (instead of whether it surpassed a certain universal cut-off or not) could assist physicians in predicting better the detrimental effect of P exposure during OS.

Our centre has systematically evaluated serum P throughout OS (i.e. from start to the end of OS) for over 10 years, mostly in line with previous research also linking higher serum P in the beginning of OS to hindered ART outcomes (Hamdine, et al., 2014). Taking advantage of this fact, we sought to assess whether the variation in P exposure during OS (using both relative and universal thresholds) could predict ART pregnancy outcome in a fresh cycle.

**Material and methods**

**Study population and design**

We performed a retrospective, single-centre cohort study between January 2010 and March 2015 including all women undergoing OS for in-vitro fertilization (IVF) in our centre. To minimize potential confounding, only patients planned for fresh embryo transfer using autologous oocytes and under GnRH antagonist pituitary suppression were included. Furthermore, couples with a planned embryo biopsy, managed natural cycles
or in-vitro maturation were excluded from the analysis. Approval and waiver of written informed consent to retrieve and analyse the data was obtained from the Ethical Committee of Universitair Ziekenhuis Brussel.

**Ovarian stimulation, pituitary suppression and cycle monitoring**

The choice of stimulation protocol was made according to clinician preference and patient characteristics, as described before (Santos-Ribeiro, et al., 2014). Of relevance, women began ovarian stimulation on day 2 of the menstrual cycle with either daily recombinant follicle-stimulating hormone (rFSH; Gonal-F®, Merck Pharmaceuticals, Darmstad, Germany; or Puregon®, Merck Sharp & Dohme, New Jersey, USA) or highly purified human menopausal gonadotrophin (HP-HMG; Menopur®, Ferring Pharmaceuticals, St. Prex, Switzerland). Pituitary down-regulation, using a GnRH antagonist, was performed with a daily administration of either cetrorelix (Cetrotide®, Merck Pharmaceuticals, Darmstad, Germany) or ganirelix (Orgalutran®, Merck Sharp & Dohme, New Jersey, USA) starting from day 7 of the menstrual cycle (or day 6 of exogenous OS) onwards. As the routine policy in our centre is not to perform any hormonal treatment (e.g. oral contraceptive pill or estrogens) prior to starting an ART cycle, the hormonal levels sampled prior to the start of OS reflect the patient’s basal endogenous endocrine profile. Cycles were monitored by means of serial vaginal ultrasound scans and serum determination of estradiol (E₂) and P. As soon as three follicles of ≥ 17 mm were observed, final oocyte maturation and ovulation were triggered with either highly purified urinary hCG (5000 UI or 10000 UI, according to the physicians’ preference and female patient weight; Pregnyl®, Merck Sharp & Dohme, New Jersey, USA) or 250 UI of recombinant hCG (Ovitrelle®, Merck Pharmaceuticals, Darmstad, Germany). Cumulus-oocyte complexes were collected by transvaginal aspiration approximately 36 hours after hCG administration.

**Progesterone assessment immunoassay and main measure of progesterone variation**

P was assessed throughout OS using a validated electrochemiluminescence immunoassay (Cobas 6000®, Roche, Basel, Switzerland) with a measured sensitivity and total imprecision of 0.03 ng/mL and <7 %, respectively. The same assay was performed during
the full duration of the study and was regularly calibrated to minimize variation of the results associated with time and reagent batch renewal. All samples were analysed at the same department.

P variation during OS (PV-OS) was defined as the difference between the serum levels of P on the day of hCG administration and the baseline circulating P concentrations on the day OS was initiated (“hCG-day P” – “OS-start P”). Statistically significant predictors of PV-OS were determined using multivariable generalized estimating equation (GEE) regression modelling (to account for the eventual clustering of cycles performed by the same women) including the following potential confounders: female age, the type, duration and total dose of exogenous OS utilized, and the number of oocytes retrieved.

Main outcome measures and statistical analysis

Our main outcome measure was live-birth (LB) delivery after 24 weeks, with unknown outcomes (i.e. patients lost to follow-up) being considered as negative. LBR per embryo transfer cycle were assessed using multivariable GEE regression analysis including PV-OS and the following known potential confounders for ART pregnancy outcome in the model: female age, number of preceding IVF/ICSI cycles, the type and total dose of exogenous gonadotropin administration, the endocrine profile (P and E₂) on the day of hCG administration, number of oocytes retrieved, stage and number of embryos transferred, and quality of the best embryo transferred. PV-OS was included in the regression model both as a continuous variable and categorized in regular intervals of 0.20 ng/mL, to facilitate the clinical interpretation of the results and to postulate potentially relevant clinical cut-offs of PV-OS for further analysis. To minimize bias by assuming that the relationship between PV-OS and LB may be linear (Templeton, et al., 1996, Bosch, et al., 2010, Sunkara, et al., 2011, Xu, et al., 2012, Santos-Ribeiro, et al., 2014), we pitted all significant PV-OS linear relationships that were found against multiple first and second degree fractional polynomials (FP) using an algorithm developed by Sauerbrei, et al. (2006), with the best-fit being chosen for the final regression model.

Finally, in order to verify whether the association between the duration of LFEP and CPR found by Huang, et al. (2012) was reproducible using LBR as the outcome of interest, we performed an additional analysis in which we divided our sample according to the duration of P elevation (0 days, 1 day, 2 days or ≥3 days). We repeated this analysis using
two cut-off levels of LFEP, specifically 1.00 ng/mL [as used by Huang, et al. (2012)] and 1.50 ng/mL. These models were then compared to each other using the Akaike's (AIC) and Bayesian information criteria (AIC and BIC, respectively). These statistical approaches are measure of the relative quality of statistical models for a given set of data. Specifically, AIC and BIC estimate the quality of a model relative to each of the other models. The lowest AIC/BIC suggest the best model, with a difference (ΔAIC and ΔBIC, respectively) of ≥10 providing strong evidence of significant superiority of the model with the lowest AIC/BIC (Kass and Raftery, 1995, Burnham and Anderson, 2010).

A p-value was considered significant whenever <0.05, with Bonferroni-adjusted pairwise comparisons whenever applicable. For the statistical analysis, we used Stata Software version 13.1 (StataCorp, College Station, Texas, USA).

Results

Patient demographics and treatment characteristics

A cohort of 4312 cycles (performed in 3151 women) were included in the study (Table 5, p. 102). The mean (±standard deviation, SD) female age of the cohort was 34.0±5.0 years old and 41.3% had no prior IVF/ICSI cycle attempt. The mean number of oocytes retrieved was 9.3±5.7 and 64.7% of the fresh embryo transfers were performed at cleavage stage (specifically, at day 3 of embryo development), with a single embryo being transferred in 51.0% of all cases. Of reference, the mean PV-OS was 0.29±0.37 ng/mL and the overall LB rate was 25.2%.

PV-OS as an overall predictor of LB

Table 6 (p. 104) presents the multivariable GEE regression analyses performed to assess the relationship between LB and PV-OS, while Figure 11 (p. 103) depicts the estimated LBR according to PV-OS levels. In the overall sample, an increase in PV-OS was associated with a linear and statistically significant decrease in LBR, with LBR being significantly lower whenever PV-OS >0.40 ng/mL (17.4%). Specifically, predicted LBR ranged from below 15% and above 30%, according to the PV-OS of the cycle (Figure 11, p. 103). This relationship remained significant even though the level of P on the day of hCG administration was already accounted for in the model (Table 6, p. 104).
How much does PV-OS add to the prediction of LBR?

In order to evaluate whether PV-OS had an independent predictive value and that it was not a surrogate of P on the day of hCG administration, we performed two separate sensitivity analyses: a crude and confounder-adjusted approach. For both approaches, PV-OS was subdivided in more clinically-useful subgroups, specifically non-elevated (≤0.40 ng/mL) and elevated (>0.40 ng/mL) PV-OS. The baseline characteristics of these subgroups are detailed in Table 5 (p. 102).

Table 5 – Patient and cycles characteristics (n=4312)

*The sample was also sub-classified in non-elevated PV-OS (≤0.40 ng/mL) and elevated PV-OS (>0.40 ng/mL) subgroups following the results of the regression analysis presented later in Table 6; SD, standard deviation; ART, assisted reproductive technologies; OS, ovarian stimulation; E2, estradiol; P, progesterone; HP-HMG, highly purified human menopausal gonadotrophin; IU, international units; hCG, human chorionic gonadotropin; PV-OS, progesterone variation during ovarian stimulation.
First, we sub-grouped the sample within the boundaries of the two commonly applied (Bosch, et al., 2010, Huang, et al., 2012) thresholds adopted for LFEP (1.00 ng/mL and 1.50 ng/mL). Then, we evaluated how LBR varied within each subgroup according to whether they had a non-elevated or elevated PV-OS (Table 7, p. 105). LBR did not vary significantly between those with hCG-day P of <1.00 ng/mL and 1.00-1.50 ng/mL (26.1% versus 25.2%, respectively), but were significantly lower in cycles with P levels >1.50 ng/mL (17.4%). Furthermore, discriminating between women with non-elevated and elevated PV-OS did not seem to have an additional predictive value in the hCG-day P >1.50 ng/mL group (LBR 17.9% versus 17.4%, respectively; p=0.936). However, among those with hCG-day P <1.00 ng/mL and between 1.00-1.50 ng/mL, the non-elevated PV-OS cycles (27.0% and 28.8%, respectively) had significantly better LBR when compared to those with elevated PV-OS (19.5% and 22.9%, respectively; p=0.007 and p=0.024, respectively).

Second, to confirm whether PV-OS maintained its predictive use in addition to hCG-day P after adjusting for potential confounding, we repeated our multivariable GEE regression analysis using all the potential confounders mentioned in Table 6 (p. 104, including hCG-day P levels), but in a sample restricted according to their hCG-day P levels. Elevated PV-
Results

OS continued to be independently predictive of LB even when we restricted our sample to those with a hCG-day P of <1.50 ng/mL (aOR 0.71, 95% CI 0.57-0.87) and <1.00 ng/mL (aOR 0.61, 95% CI 0.44-0.84). Conversely, for cycles in which the P level on the day of hCG administration was >1.50 ng/mL, PV-OS had no added predictive value (aOR 0.92, 95% CI 0.32-0.67). In terms of the relative frequency of elevated PV-OS among the overall

<table>
<thead>
<tr>
<th>Basic demographic characteristics</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female age</td>
<td>0.97 (0.95-0.98)</td>
</tr>
<tr>
<td>Previous ART cycles</td>
<td>0.92 (0.87-0.96)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exogenous gonadotropin administration</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of HP-HMG</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>Total dose</td>
<td>1.04 (0.87-1.23)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Endocrine profile on day of hCG administration</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E2</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td>P</td>
<td>1.21 (0.93-1.58)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PV-OS (ng/mL)*†</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>1.02 (0.82-1.26)</td>
</tr>
<tr>
<td>0.00-0.20</td>
<td>Reference</td>
</tr>
<tr>
<td>0.20-0.40</td>
<td>0.86 (0.70-1.05)</td>
</tr>
<tr>
<td>0.40-0.60</td>
<td>0.75 (0.58-0.96)</td>
</tr>
<tr>
<td>0.60-0.80</td>
<td>0.59 (0.42-0.83)</td>
</tr>
<tr>
<td>≥0.80</td>
<td>0.42 (0.28-0.63)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ovarian response</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Oocytes retrieved</td>
<td>1.03 (1.01-1.04)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Embryo transfer</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of embryos transferred</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Reference</td>
</tr>
<tr>
<td>2</td>
<td>1.14 (0.96-1.36)</td>
</tr>
<tr>
<td>≥3</td>
<td>0.93 (0.64-1.34)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Embryo developmental stage</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleavage</td>
<td>Reference</td>
</tr>
<tr>
<td>Blastocyst</td>
<td>1.71 (1.45-2.01)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quality of best embryo transferred</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reference</td>
</tr>
<tr>
<td>2</td>
<td>0.76 (0.64-0.90)</td>
</tr>
<tr>
<td>3-4</td>
<td>0.31 (0.20-0.49)</td>
</tr>
</tbody>
</table>

Table 6 – Predictors for LB (n=4312)
aOR (95% CI) obtained following multivariable GEE regression analysis; *The predicted LBR are plotted in Figure 12; †PV-OS remained a significant and an independent predictor of LBR (aOR 0.50, 95% CI 0.37-0.68) also when added to the model as a continuous (instead of ordinal categorical) variable. Furthermore, when evaluating the linearity of the relationship between PV-OS and LB (Sauerbrei, et al., 2006), the best-fit for PV-OS in the model was as a linear continuous variable; aOR, adjusted odds-ratio; CI, confidence interval; ART, assisted reproductive technologies; HP-HMG, highly purified human menopausal gonadotrophin; hCG, human chorionic gonadotropin; E2, estradiol; P, Progesterone; PV-OS, progesterone variation during ovarian stimulation; LB, live birth; GEE, generalized estimating equations; LBR, live birth rates.
sample, disregarding this measure of the relative increase of P during OS would mean that up to 25.0% of the cycles (n=1052, 344 cycles with hCG-day P <1.00 ng/mL and 708 cycles with hCG-day 1.00-1.50 ng/mL) would be potentially considered as having an unaffected likelihood of LB following a fresh embryo transfer if only the absolute threshold of hCG-day P were to be considered.

Pitting PV-OS against the duration of LFEP

As we were particularly interested in verifying whether the association between the duration of LFEP and CPR found by Huang, et al. (2012) was reproducible using LBR as the outcome of interest, we subdivided our sample according to the duration of LFEP (0 days, 1 day, 2 days or ≥3 days). We evaluated the relationship between LFEP and LBR using two different LFEP cut-offs, specifically 1.00 ng/mL [as used by Huang, et al. (2012)] and 1.50 ng/mL, as depicted in Figure 12 (p. 106). Overall, we failed to reproduce the results of Huang, et al. (2012) using LBR as the outcome measure (p=0.259). Furthermore, our analysis revealed that the duration of LFEP only affected LBR significantly for the cut-off >1.50 ng/mL (p<0.001). These results remained the same even after performing multivariable regression analysis for each cut-off value adjusting for potential confounding (for LFEP >1.0 ng/mL: aOR 1.05, 95% CI 0.96-1.15; Table 8, p. 107).
although the LFEP >1.50 ng/mL threshold did indeed predict LBR independently (aOR 0.69, 95% CI 0.53-0.90), the relative frequency of having such LFEP levels for 2 days (1.2%) or ≥3 days (0.2%) was exceedingly rare (Figure 12, p. 106).

![Figure 12 – LBR frequency of LFEP according to varying cut-offs for LFEP](image)

The lines represent the crude LBRs that occurred in the sample according to the duration of serum LFEP at two specific cut-offs to define LFEP (>1.00 ng/mL and >1.50 ng/mL). The bars depict the relative frequency within the sample of cycles with LFEP according to the duration of the LFEP and the cut-off used. Mantel-Haenszel trend analysis p-values for the cut-offs 1.00 ng/mL and 1.50 ng/mL: 0.259 and <0.001, respectively. LBR, live birth rate; LFEP, late-follicular elevated progesterone.

Finally, given that hCG-day P levels are the only widely used measurement in clinical practice to estimate the effect of P exposure on LBRs following ART, we were interested in estimating whether adding either universal and relative cumulative P exposure measures would be of any additional benefit (Table 8, p. 107). The results revealed that knowing the PV-OS improved the ability of physicians to predict LBR significantly (Kass and Raftery, 1995, Burnham and Anderson, 2010). Conversely, the duration LFEP had little effect towards understanding better how P hinders ER and LBR.

What may be done to avoid an increase in PV-OS?

Table 9 (p. 108) presents the multivariable regression models assessing the potential predictors of PV-OS. Ovarian response (as assessed by the number of oocytes retrieved) was associated with a significant increase in PV-OS (0.020 ng/mL increase in PV-OS per
Results

Sequentially, each extra 1000 IU of exogenous gonadotropin were associated with an increase in 0.100 ng/mL in PV-OS. Conversely, whilst statistically significant, the type and duration of gonadotropin utilized during OS also seemed to play a very modest role in determining PV-OS levels. Specifically, the use of HP-HMG was associated, on average, with an overall decrease of only 0.077 ng/mL in PV-OS. Meanwhile, each extra day of OS was associated with a mere 0.007 ng/mL increase in PV-OS.

Discussion

The evaluation of circulating P on the day of hCG administration is widely used as a surrogate marker of ER. The rationale for conducting the present study was the hypothesis that this single assessment of P during OS could be insufficient to provide the complete picture of the cumulative effect of follicular phase P exposure on the endometrium Huang et al. (2012). Our results seem to confirm that this could indeed be the case and that knowing the relative variation of exposure to P during OS (i.e. PV-OS) may reveal an
important subgroup of women who would be originally expected to have unaltered chances of conception if only hCG-day P levels were to be measured. Furthermore, we showed that a universal cut-off for cumulative P exposure (more specifically, the duration of LFEP) failed to deliver the same amount of insight on the effect that cumulative P during OS may have on each individual woman on LBR. Such information is important as it may allow physicians to tailor better the ideal timing for embryo transfer to each patient (i.e. during the fresh cycle versus a subsequent unstimulated cycle).

<table>
<thead>
<tr>
<th>Basic demographic characteristics</th>
<th>Adjusted regression coefficients (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female age (per year-old)</td>
<td>0.001 (-0.002; 0.003)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exogenous gonadotropin administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of HP-HMG</td>
</tr>
<tr>
<td>Total duration (per day)</td>
</tr>
<tr>
<td>Total dose (per 1000 IU)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ovarian response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oocytes retrieved (per oocyte)</td>
</tr>
</tbody>
</table>

**Table 9 – Predictors of PV-OS (n=4312)**

Adjusted regression coefficients (95% CI) obtained following multivariable GEE regression analysis; CI, confidence interval; HP-HMG, highly purified human menopausal gonadotrophin; IU, international units; PV-OS, progesterone variation during ovarian stimulation; GEE, generalized estimating equations.

**Potential added-value of PV-OS (versus a universal hCG-day LFEP measurement only)**

Difficulties in the interpretation of LFEP is a challenge that is still seen in daily clinical practice and that led to a nearly 20-year controversy around the usefulness of such a marker (Bosch, et al., 2010, Venetis, et al., 2013). Although the study of Bosch, et al. (2010) was an important landmark in the field, by acknowledging for the first time that the relationship of P and pregnancy outcome may be non-linear, such a conclusion limits the ease-of-use of the 1.50 ng/mL threshold these authors established in everyday clinical practice. Various other universal LFEP cut-offs have been proposed, ranging from 0.40 to 3.00 ng/mL. In their meta-analysis, Venetis, et al. (2013) concluded that thresholds >0.80 ng/mL may be clinically relevant. However, the moderate to high levels (as measured by the I² statistic) of heterogeneity among the studies and the overwhelming amount of negative studies using cut-offs <1.50 ng/mL have deterred most clinicians from using these lower universal thresholds. Recently, Martinez, et al. (2016) presented a pragmatic approach to evaluate the usefulness of a single P measurement to predict LB altogether, regardless of any threshold. These researchers failed once more to associate P levels on
the day of hCG administration to ART outcomes, alluding to the frustrations encountered with a universal LFEP cut-off for all women. To this extent, PV-OS may be a more interesting alternative, especially in women below the usual hCG-day P thresholds (e.g. <1.50 ng/mL and even <1.00 ng/mL).

**Understanding the potential advantages of PV-OS over LFEP**

Following endometrial priming by E2, P regulates the intricate sequence of molecular events in preparation for implantation (Sonigo, et al., 2014, Lawrenz and Fatemi, 2017). Although P receptors reach their maximal concentrations during the mid to late proliferative phase of the menstrual cycle (Barile, et al., 1979, Lessey, et al., 1988, Ravn, et al., 1994), they are present throughout the menstrual cycle and are responsive to P (Ingamells, et al., 1996, Lau, et al., 1996, Marshburn, et al., 2005). This raises the concern that the presence of these higher concentrations of exogenous FSH-induced (Oktem, et al., 2017) circulating P much earlier than expected in natural cycles may cause endometrial asynchrony during OS.

The potential importance of the duration of elevated P exposure was first presented by Huang, et al. (2012), who concluded that the women with longer-withstanding LFEP>1.00 ng/mL levels seemed to have worse pregnancy outcomes. However, we failed to reproduce their results using LBR as the outcome of interest, showing that this sort of evaluation was only statistically significant for LBR when a threshold of >1.50 ng/mL was used. We argue that our results should be considered more robust and that there may be at least three reasons which may justify the differences in findings. First, the P assay used in our study was more sensitive and subject to less variation. Second, the authors of the previous study grouped LFEP lasting 1 and 2 days together (accentuating the overall effect of LFEP), while we maintained them separate and accounted for the potential clustering of cycles performed by the same patient. Last, but most importantly, using a universal LFEP threshold may fail to adequately account for the higher baseline P levels that the endometrium of younger and otherwise healthy patients will usually be exposed to even prior to commencing OS.

Furthermore, while the duration of LFEP >1.50 ng/mL was a significant and independent predictor of LBR in our results, the application of this finding into daily clinical practice is of limited use and rather challenging, given that a) an LFEP>1.50 ng/mL of ≥2 days was
exceedingly rare and b) it requires that patients perform daily P evaluations in the later stages of OS. For this reason, we consider that PV-OS may not only deliver more information on the effect of follicular phase P exposure during OS, but may also be a more standardized and clinically-applicable approach (the measurement of serum P at the start and end of OS).

Questions that remain unresolved within this study

While PV-OS may be of added benefit to clinicians, some questions remain unanswered. Most notably, a potential limitation of this study is how well one may extrapolate our results to daily clinical practice. Firstly, although we consider that the fact that we do not routinely administer an oral contraceptive pill or estradiol prior to initiating a treatment cycle may have provided an advantage to the analysis, the question whether women performing hormonal pre-treatment for cycle scheduling may have reproducible outcomes remains unanswered. Secondly, our results also cannot be extrapolated to women with P levels in the beginning of OS >1.50 ng/mL, given that such values are considered a contraindication to initiate OS in our centre.

Furthermore, while this study does include a large sample set and adjusts for multiple potential confounding factors, the results are limited by its retrospective nature and by the inability to include all relevant potential confounders (e.g. female smoking habits) in the analysis. A better extrapolation of these results could be obtained if they were to be validated in future and preferably prospective studies, with larger subsets of women not encompassing the group of those originally considered to have normal circulating P levels.

Finally, as other researchers have previously pointed out (de Ziegler, et al., 2017, Healy, et al., 2017), most of the currently available P assays may not have the required accuracy to determine subtle changes in P occurring during the follicular phase, given that they were developed to be used in the luteal phase. These less precise P measurements may justify simultaneously why evaluating P only on the day of hCG administration have a limited predictive value and why multiple P assessments (such as PV-OS) may potentially deliver more clinically relevant predictions. Such a hypothesis should be confirmed using more sensitive assays.
Clinical relevance of these results

Overall, this study takes a different approach in attempting to understand why FETs seem to deliver better LB rates, improving our ability to determine which patients would benefit from delaying their embryo transfer to obtain a higher success rate. Specifically, PV-OS was especially relevant for those women below the conventionally applied hCG-day P thresholds (e.g. <1.50 ng/mL and even <1.00 ng/mL), a result that seems to justify why Roque, et al. (2015), found that even women with hCG-day P levels <1.5 ng/mL benefited from a freeze-all approach.

Most of the literature thus far have proposed two alternative solutions in an attempt to circumvent the need for a systematic freeze-all strategy following LFEP, specifically a) to replace the type of exogenous stimulation with one having luteinizing hormone (LH) activity to stimulate the theca-cell conversion of P to androgens (Smitz, et al., 2007, Bosch, et al., 2010, Arce and Smitz, 2013) and b) lower the total dose of stimulation (Bosch, et al., 2010, Venetis, et al., 2016). However, with the renewed information derived from this study, some of these potential solutions may be of questionable value. Firstly, there seemed to be a very limited benefit of using drugs with LH-activity to prevent LFEP in our study. These results are corroborated by fundamental studies published just recently pointing towards a FSH-derived production of P (Oktem, et al., 2017). This may also be one of the reasons why randomized controlled trials thus far have unanimously failed to provide any convincing evidence of benefit of any OS regimen over the other in terms of LBR (van Wely, et al., 2003, Andersen, et al., 2006, Devroey, et al., 2009). Secondly, although we did confirm that a decrease in the total dose of exogenous gonadotropins effectively reduces the excessive production of P (Bosch, et al., 2010, Venetis, et al., 2015), substantial decreases (specifically, 1000 IU) are required to obtain a modest decrease of 0.10 ng/mL in PV-OS. Such steep reductions, while potentially benefiting ART outcomes in the short-term [i.e. LBR] (Sunkara, et al., 2011)], could later affect the long-term outcome of ART (i.e. cumulative LBR), given the expected decrease in the number of oocytes retrieved (Drakopoulos, et al., 2016).

In conclusion, PV-OS was associated with a linear decrease in LBR even among women with late-follicular P levels below the universal thresholds of LFEP usually applied in clinical practice.
7.2 Specific aim 2
Optimize and predict in-cycle ER

7.2.1 Study 2-A
Evaluate the effect of EMT on LBR and neonatal outcomes taking the endocrine profile into account


Abstract
The evaluation of endometrial thickness (EMT) is still part of standard cycle monitoring during in-vitro fertilization (IVF) despite the lack of robust evidence thus far of any value of this measurement to predict live birth (LB) rates. Furthermore, this surrogate of endometrial receptivity has seen very little revalidation in contemporary medical practice, where other tools such as endocrine profile monitoring have become increasingly popular. The objective of this study was to reassess whether EMT may affect the outcome of a fresh embryo transfer in modern-day medicine using a retrospective, single centre cohort of 3350 IVF cycles (2827 women) performed between 2010 and 2014. EMT was non-linearly associated with LB in the multivariate regression analysis, with LB rates being the lowest with an EMT <7.0 mm and then between 7.0 mm and 9.0 mm. An EMT <7.0 mm was also associated with a decrease in neonatal birthweight z-scores. In conclusion, these results reaffirm the use of EMT as a potential prognostic tool for LB rates and neonatal birthweight in contemporary IVF, namely when taken into consideration together with other OS monitoring methods such as the late-follicular endocrine profile.

Introduction
Throughout the years much has been published regarding potential sonographic markers for endometrial receptivity. And, although it remains a controversial issue, EMT is the
most widely used prognostic factor for endometrial receptivity during Assisted Reproductive Technology [ART, (Kasius, et al., 2014)].

Several mechanisms are responsible for the modifications caused to the morphology and histology of the endometrium prior to embryo implantation. Specifically, previous studies have shown that endometrial proliferation is dependent on reproductive age, the hormonal levels of estradiol (E₂) as well as the expression of endometrial receptors (Zhang, et al., 2005, Paulson, 2011).

There is a continuous debate on the predictive value of the measurement of EMT prior to the administration of human chorionic gonadotropin (hCG) for ovulation triggering in ART. While some authors have demonstrated a linear correlation between pregnancy rates (PR) and EMT (Rinaldi, et al., 1996, Richter, et al., 2007, Al-Ghamdi, et al., 2008, Chen, et al., 2010), others have posited that PR may even decline above a thickness of 14 mm (Weissman, et al., 1999), after which miscarriage rates may increase. Moreover, in a retrospective study (Lamanna, et al., 2008) including 606 women undergoing a long-agonist protocol, the investigators noted a parabolic trend in PR across EMT categories (with lower PR in EMT extremes below 8 mm and above 14 mm, respectively). In 2014, a meta-analysis of 22 studies concluded that the measurement of EMT was a valuable predictor for clinical pregnancy (CP), with lower CP rates below the frequently mentioned cut-off of 7 mm, which progressively increased until 10 mm of EMT (Kasius, et al., 2014). These results could lead to the conclusion that a thick endometrium may not necessarily predict pregnancy but, conversely, a thin endometrium may be associated with lower PRs, possibly owing to a thinner functional layer that exposes the embryo to the higher oxygen concentrations of the blood from the spiral arteries during implantation (Casper, 2011).

Adequate endometrial development seems to be of paramount importance for placentation, given that previous studies have shown an association between abnormal glandular or vascular development and defective-placentation disorders, including placental abruption, low birth-weight (LBW), fetal growth restriction, pregnancy-related hypertensive disorders and pregnancy loss (Toal, et al., 2007, Pelinck, et al., 2010, Rombauts, et al., 2014, Palatnik, et al., 2016). Most of the previously-mentioned studies relate EMT to PR with no mention of the potential impact on neonatal morbidity (Yuan, et al., 2016, Holden, et al., 2017, Ma, et al., 2017) despite the common knowledge that ART is associated with preterm birth and LBW (Schieve, et al., 2002, Jackson, et al., 2004, Poikkeus, et al., 2007, Declercq, et al., 2015). The contributing factors of LBW following
ART are immense with little agreement on the main underlying causes. Multiple studies have pointed to either certain relevant baseline characteristics of the ART-seeking population or specificities in the stimulation protocols and laboratory procedures (Doyle, et al., 1992, Schieve, et al., 2002, Putterman, et al., 2003, Helmerhorst, et al., 2004, Jackson, et al., 2004, Bower and Hansen, 2005, Wang, et al., 2005, Ludwig, et al., 2006). Specifically, subfertility itself is a risk factor for LBW (Bergh, et al., 1999, Draper, et al., 1999, Pandian, et al., 2001, Basso and Baird, 2003, Axmon and Hagmar, 2005) with conflicting results on whether specific etiologies of infertility pose a higher risk than others (Doyle, et al., 1992, Wang, et al., 2005). There is also evidence suggesting a detrimental effect of the hyperestrogenic milieu on neonatal outcomes, given that neonates resulting from minimal stimulation IVF may have higher birthweights when compared to conventional IVF newborns (Pelinck, et al., 2010). Furthermore, others have associated EMT<10 mm with an increased risk of adverse perinatal outcomes, including preterm delivery, LBW and fetal demise (Chung, et al., 2006), an association which may be explained by a reduced selective capacity of thinner endometria (Oron, et al., 2016).

Monitoring of both the endometrial and ovarian responses to ovarian stimulation (OS) with transvaginal ultrasound has become an important predictor of ART success (McWilliams and Frattarelli, 2007). In addition, while not being unanimously applied (Murad, 1998, Vandekerckhove, et al., 2014), many agree that a concomitant hormonal assessment may also be of benefit to predict ART outcome (Hardiman, et al., 1990, Rizk and Smitz, 1992, Loumaye, et al., 1997), as supraphysiologic hormone levels during OS seem to be the underlying cause of a so-called “endometrium-embryo asynchrony” (Kyrour, et al., 2009, Shapiro, et al., 2011, Al-Azemi, et al., 2012, Roque, et al., 2013). More specifically, it has been reported that abnormal serum progesterone (P) levels may be associated with both lower ongoing PR and LB rates (Kolibianakis, et al., 2002, Bosch, et al., 2010, Santos-Ribeiro, et al., 2014).

The main goal of this study was to estimate the predictive value of EMT in LB and the neonatal outcomes of fresh embryo transfers in contemporary medicine, accounting specifically for the endocrine profile of the patient during the late-follicular phase.
Material and Methods

Study design

This was a retrospective, single centre, cohort study including ART treatment cycles performed at the Universitair Ziekenhuis Brussel between January 2010 and December 2014. Only cycles in which patients underwent a gonadotropin-releasing hormone (GnRH) antagonist down-regulated stimulation protocol followed by a fresh embryo transfer were included. To minimize confounding derived from women with a baseline poor prognosis, we excluded cycles in women age 40 years old or above and managed natural cycles. The exclusion criteria also included those who performed cycles with known uterine abnormalities [including uterine malformations and intrauterine disease diagnosed during ultrasound or a preceding hysteroscopy (e.g. Asherman’s syndrome, endometrial polyps, submucosal myomas)] and the planned use of either surgically retrieved sperm, donor oocytes, in-vitro maturation or preimplantation genetic diagnosis. Approval and waiver of written informed consent to retrieve and analyse the data was obtained from the Ethical Committee of Universitair Ziekenhuis Brussel.

ART protocol

OS was initiated on day two of the menstrual cycle with 50-450 IU/day of recombinant follicle-stimulating hormone (rFSH: Gonal-F®, Merck Pharmaceuticals, Darmstadt, Germany; Puregon®, Merck Sharp & Dohme, Whitehouse Station, NJ, USA or Elonva®, Merck Sharp & Dohme) or highly purified human menopausal gonadotrophin (HP-HMG; Menopur®, Ferring Pharmaceuticals, St. Prex, Switzerland). Pituitary down-regulation was achieved with a daily GnRH antagonist injection 0.25 mg of either cetrorelix (Cetrotide®, Merck Pharmaceuticals) or ganirelix (Orgalutran®, Merck Sharp & Dohme) starting on day seven of the menstrual cycle.

Cycle monitoring was performed with periodic transvaginal ultrasound, and serum E₂/P concentration assessments. When at least three follicles with ≥17mm mean diameter were visible, triggering of final oocyte maturation was performed using either highly purified urinary hCG (5000 UI or 10 000 UI, according to patient weight; Pregnyl®, Merck Sharp & Dohme) or 250 UI of recombinant hCG (Ovitrelle®, Merck Pharmaceuticals). Oocyte retrieval was performed approximately 36 hours after hCG administration and followed by fertilization by either conventional IVF or intracytoplasmic sperm injection (ICSI).
Main outcomes measures

EMT was measured in millimetres on the day or the day prior to ovulation triggering. We considered the EMT as the maximal anterior-posterior distance between both endometrial layers approximately 1 cm from the uterine fundus, subtracting the thickness of intrauterine fluid in the unlikely event that such was detected.

LB was defined as the number of deliveries that resulted in at least one live born neonate beyond 24 weeks of gestational age, with twin or higher order pregnancies being considered only once (Zegers-Hochschild, et al., 2009).

Factors potentially associated with EMT

The following variables were assessed as potential predictors of EMT: female age, body mass index (BMI), total dose of exogenous FSH/HP-HMG, duration of ovarian stimulation, and late-follicular E2/P. In order to avoid bias by assuming that the relationship between these continuous predictors and EMT was linear, the best-fitting fractional polynomial of each of these variables were compared against their linear function to assess which better described their association with EMT. This recognized statistical technique is a widely-utilized method that allows one to accurately assess what type of relationship better explains the association between a continuous variable and any given continuous or dichotomous outcome (Templeton, et al., 1996, Sauerbrei, et al., 2006, Sunkara, et al., 2011).

Evaluation of the relationship between EMT, live birth and neonatal outcomes

EMT was our main exposure variable for LB. Aware of the fact that the categorization of continuous variables may be of limited value to evaluate the real effect of a predictor (since it simultaneously assumes that values in different intervals have different effects even if very close to each other and that values on the extremes but within the same interval group have the same effect), our main approach was once again to assess the relationship of EMT as a continuous variable comparing the best fitting fractional polynomial against the linear function (model 1). Nonetheless, to facilitate the application of the results into everyday clinical practice, EMT was also categorized in the following regular-2-mm-intervalled categories: <7.0 mm, 7.0-8.9 mm, 9.0-10.9 mm, 11.0-12.9 mm and ≥13.0 mm.
These intervals were chosen to provide equal intervals as close as possible to the different lower and upper threshold values employed across previous studies (Richter, et al., 2007, Lamanna, et al., 2008, Kasius, et al., 2014, Holden, et al., 2017).

When assessing the effect of EMT on LB, we considered the following variables as potential confounders: female age, BMI, rank of IVF/ICSI treatment cycle attempt, number of useable embryos (transferred and cryopreserved), number of embryos transferred (single versus multiple), embryo stage at transfer (cleavage day-3 versus blastocyst day-5), quality of the best embryo transferred [1, 2 or 3-4, sub-classified as detailed previously (Montagut, et al., 2016)] and both late-follicular phase E2 and P levels (determined on the day or day before ovulation triggering).

We also assessed the effect of EMT on neonatal outcomes, specifically gestational age at delivery, preterm birth (<37 weeks), birth weight and low birth weight (<2500 g). Given the non-linear progression of fetal growth as pregnancy develops and differences according to fetal sex (Hadlock, et al., 1985), the recorded birthweights were standardized using z-scores (Niklasson and Albertsson-Wikland, 2008). The z-scores indicate how many standard deviations an observation was above or below the reference population mean, accounting for gestational age and neonatal sex. Furthermore, given the significant influence of multiple pregnancies on fetal development, only singleton live births were evaluated. For this analysis, we considered the following variables as potential confounders: female age, BMI, parity (nulliparous versus multiparous) and both the late-follicular phase E2 and P levels.

Other considerations of the statistical analysis

Baseline characteristics were compared among the above-mentioned EMT categories for model 2, with categorical variables being presented with relative frequency (%). Continuous variables were presented as means (standard deviation) or medians (interquartile range) according to the normality of the distribution. Comparisons were performed using generalized estimating equation (GEE) regression analysis, to account for the possibility of clustering of more than one cycle deriving from the same couple.

The predictors for EMT, LB and neonatal outcomes were determined using both univariable and multivariable GEE regression analysis, adjusting for the potential
confounders previously mentioned. In model 2, the median values of each variable in the sample set were chosen as reference values.

A p-value was considered significant whenever <0.05. The statistical analysis was performed in Stata Software version 13.1 (StataCorp, College Station, Texas, USA).

Results

Patient baseline demographics and general characteristics of the treatment cycle

A total of 3350 cycles (performed in 2827 women) were included in the analysis. The baseline demographics and main cycle characteristics according to EMT are presented in Table 10 (p. 118) and Table 11 (p. 119). The distribution of the following characteristics prior to oocyte retrieval varied significantly among the many EMT categories: BMI, total dose of exogenous gonadotropins, duration of ovarian stimulation and late-follicular E₂. Conversely, the number of oocytes retrieved, useable embryos produced and characteristics of the embryo(s) transferred (specifically, number, developmental stage and quality) did not vary significantly among the groups.
Table 12 (p. 120) presents the results of the multivariable regression models for the prediction of EMT. While increases in BMI were associated with a very modest increase in EMT, increases in late-follicular P were inversely related with EMT. Specifically, each 1 kg/m² increase in BMI was linearly associated with a 0.07 mm increase in EMT and each 1.0 ng/mL increase in P was linearly associated with a 0.25 mm decrease in EMT.
The duration of ovarian stimulation and late-follicular E2 were independently and non-linearly associated with an increase in EMT (p=0.001 and p<0.001, respectively), as shown in Figure 13 (p. 120). However, their effects were modest and mostly visible only in the lower limits of these variables. Specifically, the mean EMT seemed to stabilize once a minimum of 7 days of ovarian stimulation and concentration of 1000 pg/mL of E2 were reached.

The effect of EMT on LB

In order to assess the effect of EMT on LB, we performed two logistic regression models, differing only in how the continuous variables were introduced into the model [either as
Table 13 - Multivariable regression estimates for LB with continuous variables added either as best-fitting FP (model 1) or categorized (model 2)

Multivariable GEE regression (including simultaneously all variable shown in the model); for model 1, the FPs of the variables female age, embryos produced and EMT were scaled at either as best-fitting FP (model 1) or categorized (model 2).

Fractional polynomials (model 1) or ordinal categories (model 2), respectively. Table 13 (p. 121) presents the details of these multivariable regression models. EMT was a non-linear significant predictor of LB, affecting LB rates as plotted in Figure 14 (p. 122). Furthermore, when specifically using the EMT category that included the median EMT of the sample set (9.6 mm) as the reference value (the 9.0-10.9 mm category from model 2), EMTs <7.0 mm and between 7.0-8.9 mm were both associated with a significant decrease in live-birth rates (Figure 15, p. 122).

Figure 16 (p. 123) depicts the collective effect of EMT and the late-follicular endocrine profile on LB. While late-follicular phase E2 had very little effect on LB rates, the increase of P contributed significantly to a reduction in LB rates within the same EMT measurement. For instance, for a fixed EMT of 10.0 mm, increasing late-follicular P levels

<table>
<thead>
<tr>
<th>Female age</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FP (-2; 1)</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>14.195 (-21.4052; 6.9858)</td>
<td>0.016 (-0.0534; 0.0197)</td>
<td>&lt;0.001</td>
<td>14.075 (-0.8170; 0.0126)</td>
<td>0.284</td>
</tr>
<tr>
<td>26.30</td>
<td>0.0072 (-0.2250; 0.0353)</td>
<td>0.699</td>
<td>26.199 (-0.4700; 0.0315)</td>
<td>0.628</td>
</tr>
<tr>
<td>31-35</td>
<td>Reference</td>
<td>-</td>
<td>Reference</td>
<td>-</td>
</tr>
<tr>
<td>36-39</td>
<td>-0.3730 (-0.7777; -0.1683)</td>
<td>&lt;0.001</td>
<td>-0.3730 (-0.7777; -0.1683)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BMI</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>0.0058 (-0.0097; 0.0251)</td>
<td>0.465</td>
<td></td>
<td>0.0058 (-0.0097; 0.0251)</td>
<td>0.465</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank of treatment cycle attempt</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>FP1 (-1)</td>
<td>-0.0758 (-0.1310; -0.0166)</td>
<td>0.011</td>
<td>-0.0758 (-0.1310; -0.0166)</td>
<td>0.011</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Late-follicular phase E2</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>FP1 (-1)</td>
<td>-0.7277 (-1.0202; -0.4352)</td>
<td>&lt;0.001</td>
<td>-0.7277 (-1.0202; -0.4352)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Late-follicular phase P</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>FP1 (-1)</td>
<td>-0.0826 (-0.1140; -0.0511)</td>
<td>&lt;0.001</td>
<td>-0.0826 (-0.1140; -0.0511)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Embryos produced</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>FP1 (-1)</td>
<td>-0.0226 (-0.0410; -0.0030)</td>
<td>0.019</td>
<td>-0.0226 (-0.0410; -0.0030)</td>
<td>0.019</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of embryos transferred</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>FP1 (-1)</td>
<td>-0.0191 (-0.0373; -0.0096)</td>
<td>0.039</td>
<td>-0.0191 (-0.0373; -0.0096)</td>
<td>0.039</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Embryo developmental stage at transfer</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>FP1 (-1)</td>
<td>-0.0163 (-0.0251; -0.0074)</td>
<td>0.005</td>
<td>-0.0163 (-0.0251; -0.0074)</td>
<td>0.005</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quality of best embryo transferred</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
<td>Adjusted coefficient (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td>FP1 (-1)</td>
<td>-0.0163 (-0.0251; -0.0074)</td>
<td>0.005</td>
<td>-0.0163 (-0.0251; -0.0074)</td>
<td>0.005</td>
</tr>
</tbody>
</table>
Results (Study 2 - A)

122 of 1.0 ng/mL, 1.5 ng/mL or 2.0 ng/mL were associated with decreasing LB rates of 34%, 30% and 27%, respectively.

EMT and neonatal outcomes

Table 14 (p. 124) summarizes the multivariable EMT regression estimates for gestational age, preterm birth, birth weight z-score and low birthweight (n=939). Although the risk of LBW seemed unaltered by EMT in both models, birthweight z-scores varied significantly.
according to EMT, as depicted in Figure 17 (p. 124). Conversely, gestational age and the risk of preterm birth seemed to be unaffected by EMT.

Discussion

Despite significant advancements in the fields of ultrasonography (Wang, et al., 2010, Singh, et al., 2011, Zhang, et al., 2016), immunology (Seshadri and Sunkara, 2014) and molecular diagnostics (Ruiz-Alonso, et al., 2013, Koot, et al., 2016), the potential benefit of these novel approaches are yet to be confirmed, leaving most physicians with only the measurement of EMT to aid with the decision whether to perform a fresh embryo transfer or not. The main question of the present study was to investigate whether measuring EMT still played a role in modern-day medicine. To that extent, our results provided renewed evidence that EMT may affect both LB rates and neonatal outcomes.

In their meta-analysis, Kasius, et al. (2014) concluded that the frequently reported cut-off of 7 mm only occurred in 16 of the 1989 cycles (0.8%) included in the 3 studies reporting either ongoing pregnancy or LB rates. Furthermore, although EMT <7 mm seemed to be associated with a reduction in LBR (OR 0.38), this difference was not statistically significant (95% CI 0.03-1.54). In our study, an EMT <7.0 mm occurred in 8.5% (n=284) of all cycles and was associated with a decrease in LBR, results which are more in line with the conclusions of at least three recently published studies with a higher incidence of EMT <7 mm as well (Yuan, et al., 2016, Holden, et al., 2017, Ma, et al., 2017). Another potential explanation which may justify the differences between the results by Kasius, et al. (2014) and the studies that followed (beyond having a 6- to 10-fold increase in incidence of EMT
<7 mm) may be the fact that ongoing pregnancy rates and LBR were considered together in this meta-analysis, a decision which could have produced significant residual confounding. For this reason, an updated meta-analysis integrating these most recent results is currently underway in an attempt to further understand the discrepancies among the pooled and un-pooled data thus far.
The most common causes for a thin endometrium are either an iatrogenic event (e.g. Asherman’s syndrome), infection (e.g. with secondary adhesion development) or exogenous hormonal therapy [e.g. use of oral combined contraceptives, prolonged P therapy and use of clomiphene citrate (Senturk and Erel, 2008, Lebovitz and Orvieto, 2014)]. However, a thin endometrium can still occur in other cases for reasons which are less well understood [e.g. high blood flow impedance of the uterine radial arteries (Miwa, et al., 2009)]. Early studies demonstrated that, following ovulation, the spiral arteries constrict preemptively to diminish blood flow to the functional layer (Rossman and Bartelmez, 1957). Nevertheless, a thinned or absent functional layer may subject the embryo to higher vascularity from the basal endometrium which might explain the reduction of implantation due to an elevated oxygen tension and the production of detrimental reactive oxygen species (Yang, et al., 1998, Catt and Henman, 2000).

Given the mounting evidence of a detrimental effect of a thin endometrial lining during IVF, multiple researchers have proposed potential treatment alternatives, including alternative routes for E2 administration (Tourgeman, et al., 2001) and adjuvant sildenafil (Check, et al., 2004, Takasaki, et al., 2010), pentoxifylline (Letur-Konirsch and Delanian, 2003, Ledee-Bataille, et al., 2004) and granulocyte colony-stimulating factor treatment (Xu, et al., 2015, Kunicki, et al., 2017, Li, et al., 2017), all with controversial results. More recently, stem-cell transplantation has also been shown as a promising alternative (Xu, et al., 2015, Kunicki, et al., 2017, Li, et al., 2017). Future prospective trials may assist physicians in understanding whether postponing the embryo transfer to a subsequent frozen embryo transfer cycle in order to perform further investigations (e.g. a hysteroscopy) and apply one of these treatments in a later stage may ultimately optimize ART outcome.

In contradiction with previous studies (Gurbuz, et al., 2004, Amir, et al., 2007), we did not find an association between female age and EMT. This difference in results may be justified by the fact that our study included only women below 40 years-old, who were the only group in which the effect of age seemed to become relevant for EMT in the before-mentioned studies. For this reason, we would recommend against extrapolating our results beyond the age group that was included.

Our study also evaluated the potential effect of BMI on EMT. Previous literature has already associated obesity to a hyperinsulinemic state which may promote the ovarian production of androgens and peripheral conversion to estrogens in adipocytes (Rachon and Teede, 2010). Furthermore, a relationship between obesity and abnormal endometrial
receptivity has also been described (Bellver, et al., 2007, Dessolle, et al., 2009, DeUgarte, et al., 2010), positing that obese women, despite having higher EMTs, might also have impaired implantation. Corroborating previous data (Souter, et al., 2011) we also found that BMI was linearly associated with EMT (Table 12, p. 120). However, this discrete effect of BMI on EMT did not seem to translate into a change LB rates. In fact, BMI was not associated with LB rates in this sample of GnRH antagonist suppressed cycles, an observation that was also confirmed by a recent clinical trial (Toftager, et al., 2017).

Late-follicular phase E₂ levels were also shown to have an independent impact on EMT in our study. Although previous researchers have also established a positive linear correlation between E₂ levels and EMT (Zhang, et al., 2005), this finding was contradicted by a study including 2339 cycles using various stimulation protocols, which mentioned that E₂ levels above 1000 pg/ml had no impact on EMT (Amir, et al., 2007). We report a similar finding to this latter study, but in a larger sample set. In addition, late-follicular E₂ >3000 pg/mL seemed also to result in a discrete but statistically significant negative effect on LB (Table 13, p. 121; Figure 16, p. 123) which is also in accordance with previous research (Marchini, et al., 1991, Kolibianakis and Devroey, 2002). Specifically, these authors have speculated that high E₂ levels may alter the expression of P receptors, potentially triggering an advancement in endometrial maturation despite the presence of otherwise normal pre-ovulatory circulating P levels.

There has been a longstanding debate on the importance of P in fresh embryo transfer cycles. According to studies in women undergoing oocyte donation, P plays a crucial role in endometrium receptivity, but not in oocyte quality (Hofmann, et al., 1993, Legro, et al., 1993, Check, et al., 1994, Shulman, et al., 1996, Melo, et al., 2006, Check, et al., 2010). To the best of our knowledge, this is the first study relating circulating late-follicular levels of P with EMT. Elevated P (mostly defined as a hCG-day P>1.5 ng/mL) may decrease the thickness of the endometrium by inducing early secretory endometrial transformation, which may anticipate the window of implantation and have a negative impact on LB rates (Bosch, et al., 2010, Santos-Ribeiro, et al., 2014, Venetis, et al., 2016, Healy, et al., 2017, Hill, et al., 2017).

Regarding the clinical usefulness of the cut-offs presented in model 2 (Table 13, p. 121), we would still advise caution in the application of these limits in all instances given the risks of residual confounding following ordinal categorization of a continuous variable.
However, given the relatively large size (>200 subjects) of all our study subgroups, one would expect this risk to be relatively small.

Regarding neonatal outcomes, EMT was associated with a statistically significant decrease in birthweight z-scores, albeit not clinically translatable in terms of an increase in cases of LBW, a contradicting finding which may be potentially due to the limited size of the singleton LB sample subset (Table 14, p. 124). As z-scores account for the differences in gestational age and new-born gender, this more robust statistical approach results in a more accurate assessment of the association between EMT and birthweight. This finding is also relatively unexplored, with Chung, et al. (2006) also concluding in their small subgroup analysis of singletons that an EMT <10mm had an almost threefold increased risk of LBW. Both studies highlight the potential impact of a thin endometrium beyond LB rates, with birthweight being hindered possibly due to abnormal trophoblastic invasion (Chung, et al., 2006). These results require, however, further confirmation in larger studies.

**Strengths and limitations**

The main strengths of our study are the large sample size and the inclusion of information regarding the late-follicular endocrine profile in an attempt to better understand the relationship between EMT, LB and neonatal outcomes. This allowed for a better estimation of the potential value of EMT in modern-day medicine, in which combined transvaginal ultrasound and serum hormonal monitoring during OS have prevailed in most centres. However, there are also limitations in this study beyond its retrospective nature that need to be addressed. Firstly, given the previous evidence of a potential hindering effect of thin EMT, we can assume that a number of patients may have been proposed to defer their embryo transfer in order to ensue further investigation. In a previous study from our centre, a EMT below 7 mm was the indication for the use of the “freeze-all” strategy in approximately 10% of all cases (Santos-Ribeiro, et al., 2016). This presents a risk for selection bias which may have potentially underestimated the incidence of thin EMT and the effect of thin EMT on the studied outcomes. Furthermore, while we did exclude all women with intrauterine abnormalities visible on either ultrasound or hysteroscopy, one could argue that some patients could still have undiagnosed endometrial disease, given that we do not routinely perform hysteroscopies prior to IVF.
Nonetheless, the likelihood that such could play a major role is minimal, given that two recent randomized controlled trials showed that routine hysteroscopy prior to IVF does not affect pregnancy outcome (El-Toukhy, et al., 2016, Smit, et al., 2016).

Regarding the analysis of neonatal outcomes, the relatively small subset of cases included resulted in large confidence intervals. Furthermore, as we could not adjust for other potential confounders for preterm delivery (such as history of preterm delivery) or low birth weight (namely, female smoking), our results should be interpreted with caution and require confirmation in future studies.

Implications for clinical practice and conclusion

This study adds information on the value of EMT as a non-invasive parameter to infer on endometrial receptivity, a matter that is not yet universally acknowledged. Taken together, our findings suggest that the prognostic value of EMT should still be considered in clinical practice for both LB and neonatal birthweight. Tailoring the ideal timing of embryo transfer (i.e. in the same fresh cycle or in a subsequent frozen embryo transfer cycle) according to EMT could be further optimized when the late follicular hormonal profile is also taken into account.
7.2.2 Study 2-B

Assess the benefit of in-cycle endometrial scratching and analysing the endometrial RNA expression profile to predict pregnancy outcome

Santos-Ribeiro S, Mackens S, Tournaye H; Blockeel C, Stoop D. *Endometrial receptivity enhancement through induced injury and repair during ovarian stimulation in an antagonist down-regulated cycle: the Receptivity Enhancement by Follicular-phase Renewal after Endometrial Scratching (REFRESH) trial protocol*

*Full protocol published*
Endometrial receptivity enhancement through induced injury and repair during ovarian stimulation: the Receptivity Enhancement by Follicular-phase Renewal after Endometrial Scratching (REFRESH) trial protocol

Samuel Santos-Ribeiro 1,2,*, Shari Mackens 1, Herman Tournaye 1, Christophe Blockeel 1,3, and Dominic Stoop 1

1Centre for Reproductive Medicine, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel, Brussels 1090, Belgium
2Department of Obstetrics, Gynecology and Reproductive Medicine, St Ana Marija University Hospital, Avenida Professor Ego Meno, Lisbon 1649-035, Portugal
3Department of Obstetrics & Gynecology, School of Medicine, University of Zagreb, Croatia

*Correspondence address. Centre for Reproductive Medicine, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel, Luizeplein 101, 1090 Brussels, Belgium. Tel: +322 477 66 40; Fax: +322 477 66 49; E-mail: samueldorbeiro@gmail.com; orcid.org/0000-0003-2476-7858

Submitted on September 18, 2017; resubmitted on October 13, 2017; editorial decision on October 18, 2017; accepted on October 19, 2017.

STUDY QUESTION: Does intentional endometrial injury (i.e. endometrial scratching) during ART enhance pregnancy rates?

SUMMARY ANSWER: We propose a randomized controlled clinical trial in women performing ART in which the intervention group will undergo an additional endometrial biopsy during exogenous ovarian stimulation.

WHAT IS KNOWN ALREADY: Although endometrial receptivity has been extensively studied, the mechanisms behind the implantation of an embryo remain largely a mystery. Intentional endometrial injury has been put forward by many researchers as an inexpensive clinical tool capable of enhancing endometrial receptivity. However, despite its widespread use, the benefit of endometrial scratching is still a contentious and unresolved issue.

STUDY DESIGN, SIZE, DURATION: Pragmatic two-arm randomized, single-centre, controlled open-label trial in women undergoing exogenous gonadotropin ovarian stimulation for ART followed by a fresh embryo transfer in a gonadotropin-releasing hormone antagonist suppressed cycle. The trial will include 360 women in total with a 1:1 allocation ratio and an expected total duration of up to 45 months.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Subjects in the intervention group will undergo an endometrial biopsy during the follicular phase, on the sixth to eigth day of exogenous stimulation. Furthermore, nested within this clinical trial, we will also evaluate whether the transcriptomic signatures of the material collected during the biopsy may accurately distinguish women who become pregnant from those who do not. These endometrial transcriptomic signatures will be assessed both immediately after the biopsy and following invitro decidualization.

MAIN RESULTS AND THE ROLE OF CHANCE: Our primary objective is to assess the effect of endometrial injury during exogenous gonadotropin ovarian stimulation on clinical pregnancy rates after ART. Secondary efficacy and safety outcomes include: live birth delivery after 24 weeks, the endometrial transcriptomic profile among women in the intervention group, short-term safety (e.g. procedure intolerance due to pain, post-procedure bleeding) and long-term safety (e.g. cancelled transfers, miscarriage) outcomes.
Introduction and Rationale

To obtain a live birth following ART, an intricate series of steps have to successfully occur (Huang et al., 2011). Amongst these is implantation, when the blastocyst is embedded in the endometrial stroma. This process takes place during a period of the luteal phase known as the window on implantation (Granot et al., 2012). Although endometrial receptivity (ER) has been extensively studied, the mechanisms involved remain largely a mystery.

During the luteal phase, progesterone (P) produced by the corpus luteum stimulates a pool of endometrial cells to differentiate into decidual cells and allow implantation. Various cytokines, adhesion molecules and growth/transcription factors regulate this process and an imbalance in one or more of these modulators may affect ER (Granot et al., 2012). Previous studies have shown that there is a disruption of the natural endocrine function during ART caused by exogenous ovarian stimulation, hindering endometrial decidualization, function and receptivity (Fasemi and Popovic-Todorovic, 2013). Specifically, the super-physiologic sexual steroid concentrations present during ovarian stimulation seem to cause not only an abnormal proliferation and advancement in endometrial development but also the endocrine luteal phase defect present in ART (Fasemi and Popovic-Todorovic, 2013; Koot et al., 2016).

The fate of the endometrium is decided according to the success of embryo implantation. In the absence of implantation, the corpus luteum will lack the positive embryonic feedback and will progressively senesce, reducing P production. This hormone inhibits the pro-inflammatory nuclear factor kappalight-chain-enhancer of activated B cells (NF-kB) pathway, which, in the absence of P, will initiate endometrial breakdown. Hence, a series of inflammatory mechanisms are responsible for the cyclic endometrial remodeling that occurs until the ideal conditions for implantation are met.

The similarity between the “physiologic” menstrual cycle and the mechanisms involved in the repair of endometrial injury raises the question whether artificially induced injury may have any effect on implantation. The first experiment in this field was performed in 1907, where Loeb (1907) concluded that mechanical injury caused to the endometrium of guinea pigs during the progestational phase caused rapid decidualization. These results were later confirmed in the mouse model (Humphrey, 1969; Finn and Martin, 1972; Zhang et al., 2015). Conversely, anti-inflammatory drugs seem to influence decidualization in rabbits (Hoos and Hoffman, 1983). While the exact mechanisms involved in this ER enhancement are yet to be identified, these experiments lead to the conclusion that inflammation seems to play a major role.

In humans undergoing IVF, most systematic reviews published thus far have concluded that endometrial injury is associated with a doubling of clinical pregnancy rates (CPR) and live birth rates (LBR) in both the general IVF population (El-Toukhy et al., 2012) and patients with a history of implantation failure (Nastri et al., 2012; Potdar et al., 2012). The overwhelming apparent strength of this initial evidence on “endometrial scratching” led to its widespread use (Lensen et al., 2016) and extension to other treatment modalities, namely IUI (El-Khayat et al., 2015) and frozen embryo transfers (Dunne and Taylor, 2014), both with negative results. However, opposing views pointing out the weak biological plausibility, the heterogeneity in the existing randomized controlled trials (RCT), the potential for selection bias in the systematic reviews and the lack of adequate assessment of any potential risks associated with the procedure have also been put forward (Simon and Belker, 2014; van Wely, 2014). These criticisms were later strengthened by more recent RCTs that failed to show any pragmatic benefit of endometrial scratching (Yung et al., 2014; Gibree et al., 2015). However, the topic of endometrial scratching remains a contentious and unresolved matter with multiple research groups attempting to evaluate its benefits in ongoing RCTs (Nastri et al., 2015; Lensen et al., 2016; van Hoogenhuijse et al., 2017).

The knowledge that ovarian stimulation hinders ER has led to multiple efforts to adequately assess ER prior to embryo transfer. These research groups stem from many scientific fields, including immunology, histology, endocrinology, proteomics and genomics (Fasemi and Popovic-Todorovic, 2013). Amongst these, customized microarrays that analyse the transcriptomic signature of freshly biopsied
secretory endometria have recently been developed (Diaz-Gimeno et al., 2011). By analysing the transcriptional expression profile of the endometrium, these microarrays can accurately discriminate between receptive and non-receptive uteri. Although this innovative approach has an enormous potential, its use as a decision-making tool during ART has been hampered thus far by an important limitation: a biopsy during the secretory phase induces endometrial injury which, although temporary, effectively precludes the transfer of an embryo during that window of implantation.

**Objectives**

**Primary efficacy outcome**
The primary objective of this study is to assess the effect of endometrial injury during exogenous gonadotropin ovarian stimulation on CPR during an antagonist suppressed IVF cycle. Clinical pregnancy (CP) was defined in accordance with the recommendations of the International Committee for Monitoring Assisted Reproductive Technology (ICMART), specifically as the visualization of a gestational sac during transvaginal ultrasound (Zegers-Hochschild et al., 2009).

**Secondary efficacy and safety outcomes**
The secondary efficacy endpoint is live birth delivery after 24 weeks. Furthermore, we intend to evaluate the following secondary safety outcomes:

- Short-term (e.g. procedure intolerance due to pain, post-procedure bleeding) and long-term (e.g. cancelled transfers, miscarriage) complications associated with endometrial injury.
- Compare the pathological/immunohistochemical and transcriptional profile among women who eventually become and do not become pregnant following endometrial injury.

**Other efficacy and safety outcomes**
Furthermore, we intend to assess:

- The endocrine profile present during ovarian stimulation (namely, the circulating levels of estradiol (E2) and P).
- The occurrence of adverse events (AE) and serious adverse events (SAE) until the first pregnancy test, in accordance to United States Food and Drug Administration Centre for Devices and Radiological Health (2016).

**Material and Methods**

**Study design**
We propose a pragmatic two-arm randomized, single-centre, controlled open-label trial. In summary, women undergoing exogenous gonadotropin ovarian stimulation for ART in an antagonist downregulated cycle will be included in either the control or intervention groups (Fig. 1). Subjects in the intervention group will additionally undergo an endometrial biopsy on the sixth to eighth day of exogenous stimulation.

**Study population and eligibility criteria**
Women undergoing ovarian stimulation for ART in an GnRH antagonist suppressed cycle in the Centrum voor Reproductive Geneeskunde (CRG), Universitair Ziekenhuis Brussel (UZ Brussel) will be screened and invited to participate in the clinical trial. The criteria for inclusion and exclusion are listed in Table I.

**Patient recruitment and randomization**
All patients starting IVF are presented in the monitoring visit held every week day. Patients eligible for recruitment will be screened and then contacted by telephone during the first 3 days of stimulation to be provided with oral and written patient information on the trial. Patients who are interested in participating will be booked a counselling study visit between Days 6 and 8 of ovarian stimulation. Those who consent to the trial following this counselling visit will be requested to sign a consent form. Following patient consent, all participants will be randomized using a computer-generated randomization list developed by a trusted partner from the study nurse department of our centre. Each entry of the list is individually sealed in a sequentially numbered opaque envelope and allocated in that order to patients. Participating physicians will not have access to the randomization list.

---

**Figure 1** Study design and flowchart. SD: 1-8; stimulation Days 1-8; GnRH, gonadotropin-releasing hormone; hCG, human chorionic gonadotropin; OR, oocyte retrieval; ET, embryo transfer; CP, clinical pregnancy.
### Table I Study inclusion and exclusion criteria.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Women aged ≥18 and &lt;40 years</td>
<td>- Other known reasons for impaired implantation (i.e. hydrops algae, fibroid distorted the endometrial cavity, Asherman's syndrome, adenomyosis or endometrial tuberculosis)</td>
</tr>
<tr>
<td>- Fresh ART cycle</td>
<td>- Oocyte donation accepted</td>
</tr>
<tr>
<td>- GnRH antagonist down-regulation</td>
<td>- Frozen egg transfers</td>
</tr>
<tr>
<td>- Signed informed consent</td>
<td>- Embryos planned to undergo embryo biopsy</td>
</tr>
<tr>
<td></td>
<td>- Body mass index &gt;35 or &lt;18</td>
</tr>
<tr>
<td></td>
<td>- Women already recruited for another trial on medically assisted procreation during the same cycle</td>
</tr>
<tr>
<td></td>
<td>- Women who have previously enrolled in the trial</td>
</tr>
<tr>
<td></td>
<td>- Those unable to comprehend the investigational nature of the proposed study</td>
</tr>
</tbody>
</table>

#### Study restrictions

During the study, patients will be required to refrain from continuous use of non-steroidal anti-inflammatory drugs or any other type of medication that may interfere with ovarian stimulation, embryo cryopreservation or early pregnancy.

#### Patient withdrawal, protocol violations and cycle cancellation

Patients may withdraw from the study at any time. Eligible patients who gave adequate consent to the study and later withdraw following randomization will not be replaced.

Whenever a patient does not follow the planned protocol, this will be considered as a protocol violation. This includes, for example, cancelled embryo transfers due to complications related to ovarian hypostimulation, patient intolerance to the intervention or failure to comply with the study restrictions. The nature of the protocol violation will be recorded in the electronic case report form (eCRF) and will be accounted for during the per protocol (PP) analysis.

ART cycle cancellation is defined as any interruption of the ART process that occurs before fresh embryo transfer. Cycle cancellation will occur (i) on patient request, (ii) whenever inadequate follicular development occurs or (iii) if no embryo is available for transfer.

#### Financial incentives

Under the current Belgian Health Care System, eligible subjects for this study are entitled to the reimbursement of up to six ART treatment cycles. Beyond this provision, already available to all eligible Belgian citizens, no other financial incentive will be provided to participating subjects.

#### Treatment common to both study arms

Owing to the pragmatic design of the study, ovarian stimulation, ultrasound and hormonal monitoring, ovulation induction, oocyte retrieval, embryo procedure, IVF and luteal support will be performed according to the standard protocols of our centre (Fig. 1). Specifically, all women included will undergo exogenous ovarian stimulation using GnRH antagonist suppression with daily injections of either ganirelix (Olustran®) or cetorelix (Cetorelix®). Treating physicians will decide on which exogenous gonadotropins should be used according to the patient's profile and preference and can include either recombinant FSH or highly purified urinary FSH. Ovarian stimulation will commence after it is confirmed that the patient is not pregnant and has basal levels of E2, P, FSH and LH. The stimulation will be monitored simultaneously by pelvic ultrasound and hormonal analyses (E2, P, FSH and LH), starting on Days 6–8 of stimulation and then every 1–3 days, according to the individual endocrine profile and follicular development.

#### Differences between the control and intervention arms

Women in the intervention group will undergo an endometrial biopsy on the sixth to eighth day of ovarian stimulation using a Pipelle de Corier® (CCD International, Paris, France). This class I individually and sterile-packaged medical device complies with Directive 93/42/EEC and is routinely used in our centre for endometrial sampling. It is comprised of a flexible disposable polypropylene suction cannula with an outer diameter of 3.1 mm and a 2.4 mm diameter opening on the distal end. An inner plunger creates a vacuum essential for the blind endometrial biopsy.

After the introduction of the Pipelle into the uterine cavity, it will be rotated 360° and moved up and down four times after withdrawing the piston. The procedure is usually painless and otherwise ineffective, requiring no pre- or post-medication. Slight uterine bleeding that subsides spontaneously is rare but can be seen after endometrial biopsy.

#### Endometrial samples collection, analysis and storage

During the conception of this trial, the possibility of performing ER analysis on the samples collected was considered to be an interesting nested study. For this reason, participating subjects will additionally be asked to specifically consent whether they authorize their biopsy material be assessed further or not. The endometrium collected during this RCT will be subdivided in four parts: (i) one for pathological and immunohistochemical evaluation (for eventual quality control, if needed), and (ii) three parts which will be snap-frozen either immediately after the biopsy (two fragments) or after stromal cell culture (1 fragment). These samples will be stored in the UZ Brussels for up to 5 years. We will evaluate the
transcriptomic signatures of the material collected, searching for variations between the women who eventually became pregnant and non-pregnant after ART treatment. These endometrial transcriptomic signatures will be assessed both immediately after the biopsy and following in vitro decidualization (IVD), a technique frequently used during basic laboratory research for ER (Teledenburg et al., 2010; Wiesner et al., 2012) which simulates in vitro the development of the endometrium until the theoretical day of embryo transfer. We hypothesize that at least one of these transcriptomic signatures will vary significantly between pregnant and non-pregnant women.

The analysis of the endometrial biopsies will be divided into two sequential steps. During the first step, we will determine and compare the endometrial transcriptomic signatures with and without IVD, while in the second step we will attempt to validate the predictive accuracy of these tools to ART outcome. Studies using transcriptomic ER profiles, thus far, have directly analysed RNA expression of the biopsy samples (Diaz-Gimeno et al., 2011; Koot et al., 2016). This is the approach we propose to use for the unpaired late-proliferative phase samples [the specific testing during which abnormal cellular proliferation seems to be associated with failed implantation (Koot et al., 2016)]. Although such samples are easier to obtain and quicker to analyze, the fact that they contain multiple cell types (e.g. epithelial, stromal, immune) may bias the expression profile, and issues potentially resolved by the culture and IVD of stromal cells alone.

In other words, our objective is to compare the predictive value of transcriptomic ER analysis in two distinct settings: a pragmatic low-labor intensive approach, and a 'purist', but more labor-intensive scenario.

During the first step of this sample analysis, we will evaluate the transcriptomic signatures of the material collected before and after IVD using Next Generation RNA Sequencing (RNA-seq). This technology analyses coding RNA, non-coding RNA, specific alleles (i.e. imprinting) and splicing variants. The treatment outcome (pregnant versus non-pregnant) will be considered as the outcome variable. We will compare the different gene expression profiles of the entire human genome between receptive and the non-receptive endometrium. While the late-proliferative phase biopsies will be analysed by RNA-seq without further manipulation besides cryopreservation and thawing, the post-IVD samples will be further pre-treated as described below.

For the culture of stromal cells, the endometrial biopsy will be minced and digested (for 1 h with collagenase and DNase) to obtain a single cell suspension. The suspension will be centrifuged and the cell pellet used for inoculation in a culture flask overnight. The following day, the medium will be changed and adherent endometrial stromal cells allowed to grow for further 3 to 4 weeks. These cultures will be cryopreserved and stored for decidualization experiments. IVD will be performed after thawing the confluent cultures. We will decidualize the cells with cyclic AMP and medroxyprogesterone following the protocols extensively tested at Warwick University, mimicking a luteal phase out of a proliferative biopsy. Although this technique has rarely been applied in a clinical setting, it is frequently used during basic laboratorial research to assess ER. At different time points (Days 4, 6 and 8), the cultured cells and respective supernatants will be harvested. The supernatants will be assayed for 45 cytokines, chemokine and growth factors using a multiplex suspension bead immunoassay (Meso Scale Discovery, Meso Scale Discovery, Rockville, MD). These directed expression analyses of well-known factors related to implantation will be performed to determine the optimal time point to perform the RNA-seq (the one revealing the highest differential expression between the pregnant and the non-pregnant groups).

Analysis of the RNA endometrial biopsies
RNA sequencing will be performed by the Brussels Interuniversity Genomics High Throughput Core (BrightCore). A total RNA library of molecules with known strand origin will be prepared using the Illumina TruSeq® stranded mRNA sample preparation kit. This library will be used for cluster generation on the Illumina cBOT® machine and, subsequently, paired-end sequencing (2 × 125 bp) will be performed on an Illumina HiSeq 1500®. A minimal amount of 30 million reads is expected for each sample. After obtaining the raw reads, a demultiplexing and adapter/quality trimming step will be done. Read quality will be evaluated using the readily available online tool FASTQC (http://www.bioinformatics. babraham.ac.uk/projects/fastqc/). Next, the reads will be mapped to the Human Genome (hg19). Alignment (exon-exon junction reads) will be performed using the open-source tool STAR (Dobin et al., 2013), with the quality of this alignment being evaluated with multiple tools [Samtools Stats (Li et al., 2009), Qualimap (Kišek et al., 2016) and DeepCovels (Ramírez et al., 2014)]. The mapped reads will then be translated into a quantitative measure of gene expression using HTSeq (Anders et al., 2015). The differential gene expression between the pregnant and non-pregnant samples will be analyzed with DESeq2 package, which fits generalized linear models for each gene and compares the logarithmic fold changes adjusting for multiple testing as described in detail elsewhere (Love et al., 2014).

Statistical analysis
Sample size calculation and feasibility
We performed a sample size calculation using PASS® version 11.0 (NCSS). As previously stated, most of the trials thus far have associated endometrial injury with an approximate doubling of the CPR. Depending on each trial, this meant a difference in CPR ranging from 9.9 to 54.7%. Using our centre’s database, we retrospectively calculated a 32% CPR for the population with the same inclusion/exclusion criteria. Using a conservative approach, we proceeded with the calculation of the adequate sample size needed to detect an increase of 10% (from 32 to 42%) in the intervention group using a two-side Fischer’s exact test with a significance level (alpha) of 0.05. To achieve an 80% power using a 1:1 randomization ratio, each group would need at least 180 subjects, adding up to a total of 360 subjects required for this trial.

We also performed an additional sample size calculation for the transcriptomic expression analysis to confirm that the trial would be adequately powered for this secondary outcome. We estimated that endometrium exogenously stimulated would have a differential gene expression (>2 fold variation) after a conservative evaluation of at least 20 million reads/corresponding cycle. Using the algorithm and maximal human biologic variation estimates proposed by Hurs et al. (2013), we concluded that at least 36 patients would be required to achieve a 90% power to detect the expected difference using a false-positive detection rate of 0.01 and coefficient of variation of 0.74.

The centre undertaking this clinical trial performs over 5000 oocyte retrievals per year, with ~20 patients per month being eligible for the study. Assuming a conservative participation rate of 40–50%, we estimate that we will require between 36 and 43 months to conclude this trial.

Analysis of clinical outcomes
Continuous baseline patient and cycle characteristics will be detailed using descriptive measures of centre and spread. Specifically, normally distributed data will be presented using mean and standard deviation, while non-normal continuous data will be summarized using median and interquartile range. Categorical data will be presented using absolute and relative within-group frequencies. The reporting of data will be done in accordance to CONSORT guidance (Schulz et al., 2010).

All primary and secondary dichotomous outcomes will be compared among the treatment groups using the χ² test. All continuous outcomes will be compared with either the t-test or Mann-Whitney test depending
on the normality of their distribution. The primary analysis will be performed according to the intention-to-treat (ITT) principle. However, a PP analysis, as mentioned before, will also be considered.

All tests will be two-sided with a P-value being considered significant whenever below 0.05. For the statistical analysis, we will use Stata Software® version 13.1 (StataCorp).

Missing outcome data
Missing efficacy and safety outcomes will be considered as negative, regardless of the cause, which may justify the lack of said information (e.g., cycle cancelled due to the development of no embryo or loss to follow up). These patients will be included in the analysis following the ITT principle.

Data management, monitoring and dissemination of results
Data will be collected in a secure and encrypted eCRF created specifically for the trial using Formaker Pro® version 13 (Formaker Inc.) hosted on a dedicated server at the CRG, UZ Brussel. The database has built-in validation procedures to ensure the correct introduction of data and avoid cases of missing information where such is not applicable. The doctors, study nurses and research assistants collaborating in the trial will be responsible for the data collection. Data will be stored for at least 20 years.

No specific data safety monitoring board will be established for this trial. However, the UZ Brussel has both (i) a formal structure for the systematic reporting and auditing of adverse events and (ii) a Clinical Trial Centre, responsible for the regular monitoring and auditing of ongoing trials. Quality assurance measures have been accredited by the Joint Commission International (JCI) and have EN ISO 15224:2012 certification.

The results of the study will be publicly disseminated following submission in peer-reviewed scientific journals.

Ethics and quality assurance
This study has received ethical approval by the Ethical Committee of the UZ Brussel.

The center performing this clinical trial is fully accredited by the Association for the Accreditation of Human Research Protection Programme (AAHRPP).

Discussion

Despite the weak biological plausibility and lack of knowledge regarding the potential risks associated with endometrial scratching (Simon and Belker, 2014), this procedure is still widely applied in current clinical practice (Lensen et al., 2016). This clinical trial aims to pragmatically assess the potential benefits and harms of a generalized use of this strategy. Furthermore, we will evaluate whether we are able to adequately predict ER while circumventing the need for endometrial injury during the period of the window of implantation currently required for the ER testing. Since, with the strategy applied in this clinical trial, the sampling procedure is performed during the proliferative phase, ER can be assessed while the uterine lining recovers from the short-term injury, prior to the window of implantation (Zhou et al., 2008) and within the same cycle as embryo transfer. This novel approach has great clinical significance, since it may allow physicians, for the first time, to adequately assess ER during the same treatment cycle and better tailor the timing of embryo transfer. This would contribute substantially to the reduction of the “rule of chance” in ART and may finally eliminate the inevitable blind embryo transfer of top-quality embryos into non-receptive uteri, ultimately resulting in higher pregnancy rates per embryo transfer. As the prevalence of delayed childbearing increases, such a development would be of significance for the over 1 million couples annually spending up to 50% of their annual income on ART in an attempt to become parents (Connolly et al., 2010).

Despite the strengths of this study, there are two potential limitations that need to be addressed. First, our power calculation was based on CPR, given the fact that this was the most studied outcome in published data at the time of the design of the trial (Nastrini et al., 2012). In order to tackle this potential limitation, we decided to calculate our sample size more conservatively, using a 10% expected difference in terms of CPR while maintaining the follow-up of our sample until live birth. Since then, an updated meta-analysis (Nastrini et al., 2015) has reported more robust live-birth outcomes and estimated a potential benefit of endometrial scratching in terms of LBR of ~12% (from 26 to 34%). If confirmed in our RCT, this will imply that our sample will have at least 80% power to detect this estimated difference as well. Furthermore, given the lack of previous robust information on the potential adverse events related to intentional endometrial injury (Simon and Belker, 2014), we were unable to adequately calculate the necessary sample for these secondary outcomes. For this reason, it is possible that this trial may eventually be underpowered to adequately detect differences in terms of adverse events.

Supplementary data

Supplementary data are available at Human Reproduction Open online.

Authors’ roles

S.S.R. and D.S. are the principal investigators of the trial. D.S. is the research team head. S.S.R., D.S. and S.M. participated in the initial conception and design of the clinical trial. S.S.R. and C.B. wrote the first draft of the protocol. S.S.R., H.T., S.M. and C.B. contributed to the editing of the final version of the protocol. S.M. is responsible for the processing and analysis of all biopsy samples collected. All authors are involved in the subject recruitment process. All authors read and approved the final version of the protocol.

Funding

Research Foundation—Flanders (FWO, 1524417N). This organization has no further role in the study, namely with regards to protocol development, study conduction and evaluation of results.

Conflict of interest

The authors have no conflicts of interest to declare.

References

Results


Food and Drug Administration Center for Devices and Radiological Health; Guidance for Industry and Food and Drug Administration Staff; Medical Device Reporting for Manufacturers. Maryland: Food and Drug Administration, 2016.


7.3 Specific aim 3
Assess the benefits of deferring ETs away from COS

7.3.1 Study 3-A
Evaluate the CPR in high-responder women performing either a fresh ET or freeze-all
Santos-Ribeiro S, Tournaye H, Blockeel C. Elective blastocyst vitrification for endometrial receptivity enhancement in high-responders: the Implantation enhancement by elective Cryopreservation of all viable Embryos (ICE) trial protocol
Proof of full manuscript submitted for peer review

Abstract

Background: the safest and most efficient clinical approach for women who have an excessive response when undergoing ovarian stimulation (OS) for assisted reproductive technologies (ART) is unknown. While some centres perform exogenous gonadotropin-releasing hormone (GnRH) ovulation triggering followed by enhanced luteal phase support and a fresh embryo transfer (ET), others will propose a “freeze-all” strategy to these women. However, these approaches have never been subject to a prospective head-to-head comparison.

Methods/design: We propose a two-arm randomized, single-centre, open-label clinical trial. Summarily, women undergoing exogenous gonadotropin OS for ART in a GnRH antagonist down-regulated cycle and at high risk for ovarian hyperstimulation syndrome (OHSS) will be included in either the “fresh ET” or “vitrified-warmed ET” study arm. Women in the fresh ET group will undergo GnRH agonist triggering followed by intensified luteal phase support (with a single administration of 1500 IU of exogenous human chorionic gonadotropin approximately one hour after oocyte retrieval followed by daily combined administration of progesterone and estradiol valerate) while the intervention group will electively vitrify all viable embryos after GnRH triggering and perform the ET in a subsequent unstimulated artificial vitrified-warmed embryo cycle.

Discussion: this clinical trial is of utmost importance in order to adequately evaluate how the two currently most used approaches for women at high-risk for OHSS perform in terms of efficacy and safety when pitted against each other.

Pre-trial registration: NCT02148393, registered May 16th, 2014.
Introduction and rationale

Before the first successful in-vitro fertilization (IVF) was performed in 1978 (Steptoe and Edwards, 1978), the treatment of infertility was very limited. Since then, IVF has expanded greatly and now accounts for up to 4% of all live births (Sunderam, et al., 2012). This multidirectional improvement led to the optimization of ovarian stimulation (OS) and to a better assessment of embryo quality, ultimately causing a steady increase of live birth rates (LBR) until the late 1990s (Diedrich, et al., 2007). Despite that, LBR have always remained relatively low and, since the year 2000, rather stagnant (Andersen, et al., 2005, Nyboe Andersen, et al., 2009).

Furthermore, IVF still does not come without risks, as it may complicate with ovarian hyperstimulation syndrome [OHSS (Papanikolaou, et al., 2006, Griesinger, et al., 2007, Devroey, et al., 2011)] and is associated with an increased incidence of ectopic pregnancy [EP (Ishihara, et al., 2011, Polyzos and Devroey, 2012, Shapiro, et al., 2012, Decleer, et al., 2014)] and multiple pregnancies (Gerris, et al., 1999, Gerris, et al., 2003). Of these, OHSS is of particular importance, since it is exclusively an OS-related iatrogenic complication which occurs in women who are frequently otherwise healthy (Papanikolaou, et al., 2006). While its incidence is approximately 2% to 3% per cycle, OHSS can occur in up to a third of all cases of high-risk patients undergoing OS (Engmann, et al., 2008). As a primary preventive measure, the now widely adopted use of a gonadotropin-releasing hormone (GnRH) antagonist protocol has significantly decreased the incidence of OHSS (Kolibianakis, et al., 2006, Al-Inany, et al., 2011). Besides, with the use of GnRH antagonists, the option to trigger final oocyte maturation with a GnRH agonist is possible, effectively further reducing (and practically excluding) the incidence of severe OHSS (Engmann, et al., 2008, Humaidan, 2009). However, the use of GnRH agonists for triggering hinders IVF pregnancy rates (Al-Inany, et al., 2011) and still does not completely avoid the occurrence of severe OHSS in current clinical practice (Iliodromiti, et al., 2013). In light of the reduced clinical pregnancy rates associated with GnRH agonist triggering, several possible strategies have been proposed. The first strategy is intensified luteal phase support with either an intensified supplementation (Engmann, et al., 2008) of progesterone (P) and estradiol (E2) or a low-dose administration of human chorionic gonadotropin (hCG) immediately after oocyte retrieval (Humaidan, 2012). These approaches seem to significantly increase pregnancy rates after GnRH triggering without limiting its OHSS
prevention capacity (Iliodromiti, et al., 2013, Iliodromiti, et al., 2013, Seyhan, et al., 2013). Meanwhile, a second strategy is to electively cryopreserve all embryos and then replace them in a subsequent artificially-supported unstimulated cycle (Griesinger, et al., 2007, Devroey, et al., 2011), which, until now, has been shown to be the best method to effectively reduce the occurrence of severe OHSS (Bodri, et al., 2010). Although both approaches seem reasonable, there is no consensus on which is the most adequate since no comparative trials have ever been performed.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women aged ≥18 and &lt;40 years</td>
<td>Other known reasons for impaired implantation (i.e. hydrosalpinx, fibroid distorting the endometrial cavity, Asherman's syndrome, thrombophilia or endometrial tuberculosis)</td>
</tr>
<tr>
<td>First or second ART cycle</td>
<td>Oocyte/embryos donation acceptors</td>
</tr>
<tr>
<td>High response to OS (defined as presence of ≥18 follicles of ≥11 mm on the day of GnRH triggering)</td>
<td>Embryos planned to undergo embryo biopsy</td>
</tr>
<tr>
<td>GnRH antagonist down-regulation</td>
<td>Body mass index ≥35 or ≤18</td>
</tr>
<tr>
<td>Signed informed consent</td>
<td>Women who have previously enrolled in the trial</td>
</tr>
<tr>
<td>Patients can be included only once in the trial</td>
<td>Those unable to comprehend the investigational nature of the proposed study</td>
</tr>
<tr>
<td>Planned replacement of 1 or 2 blastocysts</td>
<td></td>
</tr>
</tbody>
</table>

Table 15 – Study inclusion and exclusion criteria of the ICE trial

ART, assisted reproductive technologies; OS, ovarian stimulation; GnRH, gonadotropin-releasing hormone.

For a live-birth to occur after IVF, a viable embryo must successfully implant in a receptive endometrium (Fatemi and Popovic-Todorovic, 2013). This intricate dialogue between the embryo and the endometrium is generally only possible during a four-day timeframe in the luteal phase – the window of implantation (Diedrich, et al., 2007, Fatemi and Popovic-Todorovic, 2013). In a natural cycle, the window of implantation begins approximately 7 days after ovulation, at a time when the embryo is a blastocyst and at an optimal stage for implantation (Papanikolaou, et al., 2006). Although endometrial receptivity has been extensively studied, the mechanisms entailed remain largely a mystery. Research, thus far, has implicated endometrial non-receptivity as the cause for implantation failure in up to two-thirds of all sub-fertile couples with morphologically normal embryos (Simon, et al., 1998, Ledee-Bataille, et al., 2001, Achache and Revel, 2006, Fatemi and Popovic-Todorovic, 2013). Furthermore, it is now clear that the disruption of the natural endocrine function caused by OS hinders normal endometrial function and receptivity (Kolibianakis, et al., 2002, Bourgain and Devroey, 2003, Fatemi, et al., 2007, Fatemi and Popovic-Todorovic, 2013). The supra-physiologic concentrations of sexual steroids during OS appear to be at the root of an endometrial advancement and may also be the main cause for the luteal

The knowledge that OS hinders IVF outcomes has led to multiple efforts to adequately assess the optimum timing for ET. These research groups have come from many medical fields, including immunology, histology, immune-histochemistry, endocrinology, proteonomics and genomics (Diedrich, et al., 2007, Lessey, 2011, Fatemi and Popovic-Todorovic, 2013), developing, amongst other tools, optimal late-follicular progesterone cut-offs (Bosch, et al., 2010, Xu, et al., 2012) and endometrial receptivity arrays (Ruiz-Alonso, et al., 2013). Despite their great potential, these methods have limited value since they do not ensure pregnancy, do not avoid OHSS or EP, are expensive and, most importantly, merely diagnosis reduced endometrial receptivity.

In an attempt to circumvent these – so far “unresolvable” – issues of fresh ET following OS (the lower pregnancy rates and the risk of severe OHSS), some authors have advocated the “freeze-all” strategy in which embryos are routinely cryopreserved and transferred in a subsequent unstimulated cycle (Griesinger, et al., 2007, Cobo, et al., 2008, Devroey, et al., 2011, Shapiro, et al., 2011, Shapiro, et al., 2011, Kato, et al., 2012, Bodri, 2013, Fatemi and Popovic-Todorovic, 2013). Indirect evidence of a benefit of this approach stems from shared oocyte cycles, in which oocyte recipients have better pregnancy rates when compared to the donor who underwent stimulation (Simon, et al., 1995, Shapiro, et al., 2011, Roque, et al., 2013). Until now, retrospective and prospective studies have shown that elective all-embryo vitrification may increase pregnancy rates (Shapiro, et al., 2011), while reducing the risk for OHSS (Griesinger, et al., 2007), EP (Shapiro, et al., 2012) or even IVF-related adverse fetal/neonatal outcomes (Kansal Kalra, et al., 2011, Maheshwari, et al., 2012). However, most of the randomized controlled trials (RCT) and meta-analyses published until now to address this hypothesis have delivered controversial results. The most recent systematic review and meta-analysis (Roque, et al., 2013) included the only available 3 RCTs and found that elective frozen ET was associated with better ongoing pregnancy rates (47.3% versus 35.7%, RR 1.32, IC 95% 1.10-1.59) and similar miscarriage rates (4.7% for frozen ET versus 5.7% for fresh ET, RR 0.83, IC 0.43-1.60). However, 46.0% of its data was pooled from a publication that has recently been retracted based on the results of an investigation which found serious methodological flaws in the study (Aflatoonian, et al., 2010, JARG Editor-In-Chief, 2013). The remaining 2 RCTs derived from the same research
group and varied methodologically in terms of their study population, with one being performed in normo-responders (Shapiro, et al., 2011) while the other was implemented in a group of high-responders (Shapiro, et al., 2011). In the high responders’ trial, the ongoing
pregnancy rates did not vary significantly (77.6% for frozen ET versus 65.4% for fresh ET, p=0.1943). However, this study was limited by the fact that groups varied significantly in terms of day-5 ET rates (36.7% for frozen ET versus 84.6% for fresh ET, p<0.0001) and the number of supernumerary embryos (2.9 for frozen ET versus 1.5 for fresh ET, p<0.0002).

The lack of sufficient confirmation of the positive results found in the before-mentioned studies demanded a well-designed RCT comparing the pregnancy outcomes between fresh ET and an elective freeze-all strategy (Evers, 2013).

**Study objectives**

*Primary efficacy outcome*

The primary objective of this study is to compare the clinical pregnancy rates (CPR) between fresh ET and elective all-embryo vitrification with warming and ET in a subsequent cycle. Clinical pregnancy (CP) was defined in accordance with the recommendations by the International Committee for Monitoring Assisted Reproductive Technology (ICMART), specifically as the visualization of a gestational sac during transvaginal ultrasound (Zegers-Hochschild, et al., 2009).

*Secondary efficacy and safety outcomes*

The secondary efficacy endpoint was the live-birth delivery rates after 24 weeks. Furthermore, we intend to evaluate the following secondary safety outcomes among the two study groups:

- Incidence of severe OHSS, according to the criteria proposed by Golan and Weissman (2009);
- EP rates, diagnosed as such whenever an implantation occurs outside the uterine cavity [including heterotopic pregnancies, (Zegers-Hochschild, et al., 2009)];
- Early (<20 weeks) and late (≥20 weeks) spontaneous miscarriage rates, as defined by ICMART (Zegers-Hochschild, et al., 2009).
Other efficacy and safety outcomes

Furthermore, we intend to assess:

- The number of cycles in which a fresh ET could not be performed due to complications directly related to ovarian hyperstimulation;
- The endocrine profile during OS (namely, the circulating levels of E₂ and P);
- The occurrence of adverse events (AE) and serious adverse events (SAE) until the first pregnancy test, in accordance with the United States Food and Drug Administration’s (FDA) guidance.

Material and methods

Study design

We propose a two-arm randomized, single-centre, open-label clinical trial. Summarily, women undergoing exogenous gonadotropin ovarian stimulation for ART in a GnRH antagonist down-regulated cycle and at high risk to develop OHSS will be included in either the “fresh ET” or “vitrified-warmed ET” arm (Figure 18, p. 141). Women in the fresh ET group will undergo GnRH agonist triggering followed by intensified luteal phase support while the vitrified-warmed ET group will electively cryopreserve all viable embryos after GnRH agonist triggering and perform the ET in a subsequent unstimulated artificial vitrified-warmed embryo cycle.

Study population and eligibility criteria

Women undergoing exogenous gonadotropin OS for ART in a GnRH antagonist down-regulated cycle in the Centrum voor Reproductieve Geneeskunde (CRG), Universitair Ziekenhuis Brussel (UZ Brussel) will be screened and invited to participate in the trial. High response to OS will be defined exclusively by the presence of ≥18 follicles of ≥11 mm on the day of GnRH triggering (Papanikolaou, et al., 2006). All other criteria for inclusion and exclusion are listed in Table 15 (p. 139).
Patient recruitment and randomization

All patients commencing IVF are presented at the daily monitoring meeting. Patients eligible for recruitment will be screened and then contacted telephonically on the day of ovulation triggering to be provided with information on the trial. Patients who accept to participate in the trial will be requested to sign a consent on the day of oocyte retrieval.

Participants will be randomized using a computer-generated randomization list (SPSS Version 20) which was developed by a trusted partner from the study nurse department of our centre. Each entry of the list was sealed in a sequentially-numbered opaque envelope and will be allocated in that order to patients. Participating physicians do not have access to the randomization list.

Study restrictions

During the study, patients are restricted to use continuously non-steroids anti-inflammatory drugs or any other type of medication that may interfere with OS, embryology or early pregnancy.

Patient withdrawal and protocol violations

Patients may withdraw from the study at any time. Eligible patients who gave adequate consent to the study and later withdrew following randomization will not be replaced.

Whenever a patient does not follow the planned protocol, this will be considered as a protocol violation. This includes, for example, cancelled fresh ETs due to complications related to ovarian hyperstimulation, patient denial to perform the prescribed medication or failure to comply with the study restrictions. The nature of the protocol violation will be recorded in the electronic case report form (eCRF) and will be accounted for in the per protocol (PP) analysis.

Financial incentives

Under the current Belgian healthcare system, eligible subjects to this study are entitled to the reimbursement of up to six ART treatment cycles. Beyond this provision already
available to all eligible Belgian citizens, no other financial incentive will be provided to participating subjects.

_Treatment common to both study arms_

OS, ultrasound and hormonal monitoring, ovulation induction, oocyte retrieval, embryology procedures and IVF will be performed according to how they are normally performed in our centre in both treatment arms (Figure 18, p. 141). All women included will undergo OS using a GnRH antagonist down-regulation with daily injections of either ganirelix (Orgalutran®) or cetrorelix (Cetrotide®).

Treating physicians will opt on which exogenous gonadotropins should be used according to the patient’s profile and preference which may include either recombinant follicle-stimulating hormone (FSH) or highly purified urinary human menopausal gonadotropins (HMG). OS will commence after it is confirmed that the patient is not pregnant and has basal levels of E2, P, FSH and luteinizing hormone (LH). The stimulation will be monitored by pelvic ultrasound and hormonal analyses (E2, P, FSH and LH), starting on day 6 of stimulation and then every 1 to 3 days, according to the individual endocrine profile and follicular development.

Final oocyte maturation will be triggered with 0.2 mg triptorelin (Decapeptyl®, Gonapeptyl®) as soon as at least three follicles larger or equal to 17 mm are observed. A GnRH agonist will be the preferred triggering agent for both groups to reduce the before-mentioned risk of severe OHSS associated with hCG triggering in high-responders.

Oocyte retrieval will be performed approximately 36 hours after the GnRH agonist administration under either local anaesthesia with analgesic premedication or general anaesthesia, according to patient preference.

Conventional IVF or intracytoplasmic sperm injection (ICSI) will be performed, using the specimen of sperm made available by the male progenitor on the day of oocyte retrieval.

In all instances following the confirmation of an ongoing CP, women will be sent a questionnaire approximately one year after the on-study period to assess the occurrence of the secondary efficacy and safety outcomes.
**Difference between the fresh ET or vitrified-warmed ET study arms**

In the fresh ET group, following oocyte retrieval, intensified luteal phase support will be provided with a single administration of 1500 IU of hCG (Pregnyl®) approximately 1 h later followed by daily vaginally-administered P 600 mg/day (Utrogestan®) and oral E₂ valerate 4 mg/day (Progynova®). ET in the uterine cavity will be performed on the 5th day of embryo development at blastocyst stage under ultrasound guidance.

In the vitrified-warmed ET group, no luteal phase support will be provided immediately after oocyte retrieval. Instead, all viable embryos will be vitrified. The vitrification process will be performed under the same conditions of all other vitrification procedures usually performed in our centre, as described in detail in previous publications (Van Landuyt, et al., 2011). In summary, vitrification will be accomplished using closed CBS-VIT High Security straws (CryoBioSystem®) in combination with dimethylsulfoxide, ethylene glycol and sucrose as cryoprotectants (Irvine Scientific® Freeze Kit®).

In the meantime, patients will wait for a subsequent menstrual cycle before starting exogenous hormone therapy for endometrial preparation. This therapy will commence after it is confirmed that the patient has basal levels of E₂, P, FSH and LH and it will consist of daily oral E₂ valerate (Progynova®), 4 mg/day during 7 days followed by 6 mg/day for 6 days. After this pre-treatment, a pelvic ultrasound will be performed to assess endometrial development. If the endometrium is ≥ 7 mm, the embryo warming and ET will be scheduled. Artificial luteal support will be changed to include daily vaginal P 600 mg/day (Utrogestan®) from then onwards, while the E₂ valerate will be reduced to 4 mg/day. Otherwise, E₂ valerate will be continued at 6 mg/day until this endometrial threshold is met. If the endometrium does not reach 7 mm, the cycle will be cancelled.

On the day of embryo transfer, blastocyst(s) will be warmed one by one until one or two blastocysts are suitable for transfer. The choice to transfer one or two embryos will be decided by the clinician at consultation mainly depending on the patient’s age and the number of embryos replaced in the previous treatment cycles, according to Belgian law. ET to the uterine cavity will be performed under ultrasound guidance whenever possible.
Statistical analysis

Sample size calculation and feasibility

We performed a sample size calculation using PASS® version 11.0. We based our calculation on the results of a previous RCT which found a significantly higher pregnancy rate in women who underwent elective embryo cryopreservation after hCG triggering (Shapiro, et al., 2011). The relative difference between groups in this study in terms of CPR was 55% (84% for the cryopreservation group versus 55% for the fresh group). In another recent analysis using GnRH agonist triggering in patients with >14 follicles, the CPR was 35% (Humaidan, et al., 2013). We proceeded with the calculation on the adequate sample size needed to detect an increase of approximately 55% (from 35% in the control group to 54% in the intervention group) using a two-side pooled z-test with a significance level (alpha) of 0.05, planning beforehand two safety-check interim analyses (at one-third and two-thirds of recruitment). To achieve an 80% power using a 1:1 randomization ratio, each group would require 106 patients, adding up to a total of 212 patients required for the trial.

The centre undertaking this clinical trial performs over 5000 oocyte retrievals per year, with approximately 10-15 patients per month being eligible for the study. Assuming a participation rate of 40% to 50%, we estimate that we will require between 28 and 53 months to conclude this trial.

Statistical analysis

Continuous baseline patient and cycle characteristics will be detailed using descriptive measures of centre and spread. Specifically, normally-distributed data will be presented using means and standard deviations, while non-normal continuous data will be summarized using medians and interquartile ranges. Categorical data will be presented using absolute and relative within group frequencies. The reporting of data will be done in accordance to CONSORT guidance (Schulz, et al., 2010).

All primary and secondary dichotomous outcomes will be compared among the treatment groups using the $\chi^2$ test. All continuous outcomes will be compared with either the t-test or Mann-Whitney test according to the normality of their distribution. The primary analysis will be according to the intention-to-treat (ITT) principle. However, modified ITT (i.e. evaluating the effect of the study-arm crossover of women who ultimately required
the freeze-all strategy owing to early-onset OHSS complaints) and PP analyses will also be considered. All tests will be two-sided with a p-value being considered significant whenever below 0.05. For the statistical analysis, we will use Stata Software® version 13.1.

Missing outcome data

Missing efficacy and safety outcomes will be considered as negative, regardless of the cause which may justify the lack of said information (e.g. cycle cancelled due to the development of no embryo or loss to follow-up). These patients will be included in the analysis following the ITT principle.

Data management, monitoring and dissemination of results

Data will be collected in a secure and encrypted eCRF created specifically for the trial using Filemaker Pro® version 13 hosted on a dedicated server at the CRG-UZ Brussel. The database has inbuilt validation procedures to ensure the correct introduction of data and avoid cases of missing information where such is not applicable. The doctors, study nurses and research assistants collaborating in the trial will be responsible for the data collection. Data will be stored for at least 20 years.

Owing to a) the pragmatic nature of this clinical trial, testing head-to-head two widely established treatment modalities against each other, b) the short duration of the on-study period, c) the single-centre design and d) pre-specified efficacy and safety interim analyses, no data safety monitoring board will be established.

The results of the study will be publicly disseminated following submission in peer-reviewed scientific journals.

Ethics and quality assurance

This study has received ethical approval by the Ethical Committee of the UZ Brussel. The centre performing this clinical trial is fully accredited by the Association for the Accreditation of Human Research Protection Program (AAHRPP).
Discussion

A recent trial with over 1500 patients demonstrated that the freeze-all approach performs better than a fresh hCG-triggered ET (Chen, et al., 2016). While these results were interesting to observe, this study also adds noise to the current field of evidence, given that it applied a progressively unpopular approach for high responders (i.e. the use of hCG triggering). Furthermore, the freeze-all strategy is a very controversial topic in current medical practice, with many raising concerns towards the lack of available data to fully support it (Maheshwari and Bhattacharya, 2013, Ata and Seli, 2017). For these reasons, we consider that the clinical trial we propose is of utmost importance in order to adequately evaluate how the two currently most-used approaches for predicted high responders perform when pitted against each other.

Trial status

Ongoing trial (recruitment initiated in May 2014).
7.3.2 Study 3-B

**Compare MZT rates between fresh ET and FET**


*Full manuscript published*
Do ARTs affect the incidence of monozygotic twinning?

I. Mateizel1,⁎, S. Santos-Ribeiro1,2,⁎, E. Done3, L. Van Landuyt1, H. Van de Velde1, H. Tournaye1, and G. Verheyen1

1Centre for Reproductive Medicine, UZ Brussels, Laarbeeklaan 101, 1090 Brussels, Belgium 2Department of Obstetrics, Gynecology and Reproductive Medicine, Hospital Universitário de Santa Maria, Avenida Professor Egas Moniz, Lisbon 1649-035, Portugal 3Department of Obstetrics and Gynecology, UZ Brussels, Laarbeeklaan 101, 1090 Brussels, Belgium

⁎Correspondence address. E-mail: leenx.mateizel@uzbrussels.be

Submitted on April 5, 2016; resubmitted on June 28, 2016; accepted on July 27, 2016

STUDY QUESTION: Does the manipulation of gametes or embryos during ARTs increase the risk for monozygotic twinning (MZT)?

SUMMARY ANSWER: Frozen embryo transfer (ET) is associated with a lower MZT rate, while blastocyst culture is associated with an increased risk of monozygotic pregnancy.

WHAT IS KNOWN ALREADY: Monozygotic twins have a higher risk for perinatal complications. Although an increased incidence of monozygotic pregnancies after ART has been previously reported, data regarding the possible impact of different laboratory procedures are conflicting.

STUDY DESIGN, SIZE, DURATION: All clinical pregnancies after single ET carried out in our centre between 2004 and 2013 (n = 6096) were retrospectively analysed for the incidence of MZT. The effect of different laboratory procedures on the incidence of MZT was evaluated.

PARTICIPANTS/MATERIALS, SETTING, METHODS: The following ART risk factors were assessed: maternal age, type of ET (fresh versus frozen), zona pellucida (ZP) manipulation (specifically, ICSI, embryo biopsy and assisted hatching), use of donor oocytes, embryo stage at time of ET (cleavage, compaction, early or advanced blastocyst) and culture media.

MAIN RESULTS AND THE ROLE OF CHANCE: The overall MZT rate was 3.2% (136/6096). Frozen ET was associated with a significant reduction in MZT incidence (adjusted odds ratio (aOR) 0.48, 95% CI 0.29–0.80), while blastocyst transfer (early or advanced blastocyst) was associated with a significant increase in MZT risk (aOR 2.70, 95% CI 1.36–5.34; aOR 2.05, 95% CI 1.29–3.26, respectively). No significant differences were found between the MZT and singleton (non-MZT) groups regarding maternal age, the use of different ZP manipulation techniques, or type of culture media used.

LIMITATION, REASONS FOR CAUTION: This study is limited by its retrospective nature and the fact that monozygotic was not confirmed by genetic testing. Furthermore, since monozygotic pregnancies is a rare event, other ART parameters that may influence its incidence could not be assessed during our analysis.

WIDER IMPLICATION OF THE FINDINGS: Our findings warrant future studies designed to investigate the association between specific ART procedures and MZT, namely the potential risk of blastocyst transfer to increase MZT.

STUDY FUNDING/COMPETING INTEREST(S): No external funding was used for this study. There are no conflicts of interest.

Key words: monozygotic twins / ART / single embryo transfer / monochorionic.

Introduction

Although the exact origin of monozygotic twinning (MZT) remains unknown,Corner (1955) postulated that the abnormal embryo splitting at different time points may lead to one of the following three types of monozygotic placental arrangements: dichorionic-diamniotic (DC-DA), monochorionic-diamniotic (MC-DA) and monochorionic-monoamniotic (MC-MA).

1The authors consider that the first two authors should be regarded as joint first authors.

© The Author 2016. Published by Oxford University Press on behalf of the European Society of Human Reproduction and Embryology. All rights reserved.
For permissions, please e-mail: journals.permissions@oup.com
Several publications have raised concerns regarding the potential for an increased rate of MZT following ART (Aikani et al., 2003; Aston et al., 2008; Kawachiya et al., 2011; Knopman et al., 2014; Dawson et al., 2015; Parazzini et al., 2016; Sarai et al., 2016; Saravelos et al., 2016). These reports warrant further exploration into this matter, since monozygotic pregnancies are associated with an increased risk of maternal and foetal complications (Rao et al., 2004) such as foetal growth restriction, preterm delivery and perinatal mortality (Aston et al., 2008).

Although the risk factors and mechanisms responsible for embryo splitting are still unclear, it has been argued that specific patient characteristics, and ART procedures might play a role in the development of MZT. Previous publications have pointed out that oocyte age (Knopman et al., 2014), assisted hatching (AIH) (Taskatzis et al., 2002; Skladas et al., 2008), prolonged embryo culture (Kawachiya et al., 2011; Knopman et al., 2014; Nakasa et al., 2014) and embryo cohort quality (Franzisk et al., 2015) may affect the subsequent risk for developing MZT pregnancies, while other reports (Papanikolaou et al., 2010; Shara and Abdo, 2010; Wu et al., 2014) contradict these findings.

Therefore, we performed a retrospective analysis of the embryological and clinical data of a large cohort of single embryo transfers (SETs) in order to investigate the association between ART and MZT pregnancies.

**Materials and methods**

**Study population**

This study retrospectively reviewed all consecutive fresh or frozen embryo transfer (ET) cycles with autologous or donated oocytes performed in our centre between January 2004 and December 2013. Approval to retrieve and analyse the data was obtained from the Ethical Committee of UZ Brussels.

Only SET cycles resulting in a clinical pregnancy (CP) documented by ultrasonography were considered for the analysis.

**Ovarian stimulation, oocyte retrieval and embryo culture**

Ovarian stimulation, oocyte retrieval and embryo culture were carried out as previously described (Material et al., 2013). Different sequential media (25 μl individual droplets) were used during the period of this study: Irvine HTF medium (Irvine Scientific, Dublin, Ireland), Cook IVF Cleavage and Blastocyst Media (Cook Belgium, Strombeek-Bever, Belgium), Vetrolife Sequential media (Vetrolife, Goteborg, Sweden), Medicut, Embryo-Point and Blastocyst Media (Medicut, De Pinte, Belgium) and Sage Quin’s Advantage Protein Plus Cleavage and Blastocyst Media (Cooper Surgical, USA). On Day 3, the embryos were evaluated for the number and symmetry of blastomeres, percentage of fragmentation, vacuolization, granulation and multinucleation. On Day 5, the blastocysts were scored according to the grading system by Gardner and Schoolcraft (1999).

**ET and cryopreservation**

Patients were planned to have fresh ET followed by the freezing of good quality supernumerary embryos. The fresh ETs were performed on Day 3 or 5, as decided by the treating physician based on the baseline characteristics of each patient.

In rare instances (n = 86), no fresh ET was attempted and all the embryos were frozen and transferred in a subsequent frozen cycle: 60 due to excessive ovarian response during stimulation, 12 on patient request, 6 due to the suspicion of intrauterine pathology detected during the treatment cycle, 5 following the detection of elevated late-follicular serum progesterone levels and 3 after other medical complications that occurred during the cycle (two pelvic infections following oocyte retrieval and one suspicion of deep vein thrombosis).

On Day 3, fresh embryos were considered eligible for transfer if at least five blastomeres were present with a maximum of 50% fragmentation. On Day 5, fresh embryos were considered eligible for transfer if they reached the stage of early compaction, early (BB1 or BB2), full (BB3), expanded (BB4), hatching (BB5) or hatched (BB6) blastocyst. Advanced blastocyst stages fulfilled the criteria for transfer if they had at least an inner cell mass (ICM) Type C and trophoderm (TE) Type B.

The following embryos were considered eligible for cryopreservation: Day 3 embryos with >6 blastomeres and ≤20% fragmentation, or >20% and ≤50% fragmentation if at least an eight-cell stage; Day 5 early, full, expanded or hatching blastocysts with a Type A/B ICM and TE, and Day 6 full, expanded or hatching blastocysts with a Type A/B ICM and Type A/B TE. After thawing, Day 3 embryos were cultured overnight in blastocyst medium until they were transferred the following day (Day 4). Frozen blastocysts (Days 5 and 6) ETs were performed on the day of thawing. All frozen ETs were performed in either a natural or an artificial cycle, as previously described (Van der Vloet et al., 2014).

For the purpose of this study, all embryos were categorized according to their stage at the time of ET into one of the following groups: cleavage, compaction, early blastocyst (BB1 or BB2) or advanced blastocyst (BB3, BB4, BB5, and BB6 and collapsed).

**Clinical outcome**

CP was defined by the presence of an intrauterine gestational sac (GS) as visualized by transvaginal ultrasound examination (Zegers-Hochschild et al., 2009) and later subdivided in the two following groups: MZT and singleton (non-MZT). Monozygotic pregnancies were identified when more than one GS was visualized or when the number of foetal poles exceeded the number of GSs. Chorionicity was established by counting the number of GSs or by taking into account the thickness of the amniotic inter-twin membrane. An MZT pregnancy was considered DC when two separate GSs were visualized during early first-trimester (7 weeks) ultrasound. The thickness of the inter-twin membranes or the presence of the lambda or T signs were also used to confirm chorionicity in the presence of two adjacent GSs or at a later period of the gestation (Fineberg, 1992; Sepulveda et al., 1996). Cases classified as DC-DA twin pregnancy but later recognized as chromosomally discordant were not included in the analysis since the simultaneous occurrence of a spontaneous intra pregnancy in the same cycle could not be excluded. Monozygotic pregnancies were considered MC when at least two foetal poles were visualized within one GS with one chorionic area. Amnioticity was established by the presence or absence of a thin inter-twin membrane and by counting the number of yolk sacs or amnions within the sac. Monochorionic twin pregnancies were characterized by the presence of a single amniotic cavity and yolk sac (Bromley and Benacerraf, 1995).

**Parameters studied**

The following parameters were considered during the analytic maternal age (or the age of the oocyte donor), the use of donor oocytes, ZP manipulation (CGL, embryo biopsy for PGD and AHR), embryo stage at the time of ET (cleavage, compaction, early or advanced blastocyst), type of ET (fresh or frozen) and culture media used prior to cryopreservation (Sage, Medicut, Vetrolife or other). During the period of time included in the study, no changes were performed in the standard laboratory culture conditions except for culture media and the introduction of the vitrification
method for embryo cryopreservation. The impact of the cryopreservation protocols was not assessed due to the low number of MZT clinical pregnancies obtained after frozen ET.

Statistical analysis
The baseline sample characteristics were compared among the MZT and singleton groups using either the Fisher's exact test for categorical variables or Student's t (for continuous variables) tests followed by crude exploratory pairwise comparisons whenever necessary.

MZT rates per ET were assessed using multivariable logistic regression to account for the above-mentioned study parameters. Since not all embryos were cultured in the same type of medium before cryopreservation and after thawing, the culture medium could not be included as a possible confounder in the general analysis. We also performed a subgroup analysis dividing the sample according to the type of ET.

A P-value was considered significant if <0.05. For the statistical analysis, we used Stata Software version 13.1 (StataCorp, College Station, Texas, USA). Given the low incidence of MZT (<1%), odds ratios (ORs) could be considered adequate estimates of risk ratios (Zheng and Yu, 1998).

Results
A total of 6103 CEs after SET performed in our centre during the study period were assessed (Supplementary data, Fig. S1). Of these, seven CEs resulted in twin deliveries with a discordant chromosomal constitution (five cases with gender discordancy and two cases with one of the twins having an extra Chromosome 21) and were excluded from the analysis. From the remaining 6096 CEs, 136 (2.2%) were MZT pregnancies. Out of the 5341 CP with a known pregnancy outcome, 4351 resulted in at least one live birth, of which 98 originated from an MZT pregnancy including 66 (1.5%) monozygotic twin deliveries.

During the study period, the trend of MZT incidence did not vary significantly over time (P = 0.661, Fig. 1).

Table I shows the distribution of the baseline cycle characteristics within the MZT and singleton groups (while Supplementary data, Table S1 presents the incidence of MZT according to the same cycle characteristics). None of the cycles with AH resulted in an MZT pregnancy. The stage of the embryo and the types of ET were the only parameters unevenly distributed among the two groups. Specifically, when performing pairwise comparisons among the various embryo stages, early and advanced blastocysts were observed more frequently than cleavage and compacted embryos in the MZT group. Regarding the type of transfer, fresh ET occurred significantly more frequently than frozen ET in the MZT group (P = 0.008).

To understand the relationship between the type and day of ET, and the incidence of MZT, we performed a pairwise comparison stratified by these two variables. In the MZT group (n = 136), fresh Day 3 ET (n = 89) occurred more frequently than fresh Day 3 (n = 28), frozen

<table>
<thead>
<tr>
<th>Age</th>
<th>Singleton (n = 5960)</th>
<th>MZT (n = 136)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;25</td>
<td>313 ± 4.1</td>
<td>314 ± 3.8</td>
<td>0.644</td>
</tr>
<tr>
<td>25–29</td>
<td>235 (5.0%)</td>
<td>5 (3.7%)</td>
<td></td>
</tr>
<tr>
<td>30–34</td>
<td>1739 (29.2%)</td>
<td>32 (23.5%)</td>
<td></td>
</tr>
<tr>
<td>35–39</td>
<td>2688 (44.8%)</td>
<td>69 (50.7%)</td>
<td></td>
</tr>
<tr>
<td>≥40</td>
<td>1100 (18.5%)</td>
<td>28 (20.6%)</td>
<td></td>
</tr>
</tbody>
</table>

Recipients of donated oocytes

<table>
<thead>
<tr>
<th>Age</th>
<th>Singleton (n = 5960)</th>
<th>MZT (n = 136)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>5766 (96.7%)</td>
<td>133 (97.8%)</td>
<td>0.803</td>
</tr>
<tr>
<td>Yes</td>
<td>194 (3.3%)</td>
<td>3 (2.2%)</td>
<td></td>
</tr>
</tbody>
</table>

Insemination type

<table>
<thead>
<tr>
<th>Age</th>
<th>Singleton (n = 5960)</th>
<th>MZT (n = 136)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional IVF</td>
<td>777 (13.0%)</td>
<td>22 (16.2%)</td>
<td>0.308</td>
</tr>
<tr>
<td>ICSI</td>
<td>5183 (87.0%)</td>
<td>114 (83.8%)</td>
<td></td>
</tr>
</tbody>
</table>

Embryo biopsy

<table>
<thead>
<tr>
<th>Age</th>
<th>Singleton (n = 5960)</th>
<th>MZT (n = 136)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>5022 (84.3%)</td>
<td>109 (80.2%)</td>
<td>0.192</td>
</tr>
<tr>
<td>Yes</td>
<td>938 (15.7%)</td>
<td>27 (19.9%)</td>
<td></td>
</tr>
</tbody>
</table>

Assisted hatching

<table>
<thead>
<tr>
<th>Age</th>
<th>Singleton (n = 5960)</th>
<th>MZT (n = 136)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>591 (99.2%)</td>
<td>136 (100.0%)</td>
<td>0.627</td>
</tr>
<tr>
<td>Yes</td>
<td>46 (0.8%)</td>
<td>0 (0.0%)</td>
<td></td>
</tr>
</tbody>
</table>

Embryo stage

<table>
<thead>
<tr>
<th>Age</th>
<th>Singleton (n = 5960)</th>
<th>MZT (n = 136)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleavage</td>
<td>1786 (30.0%)</td>
<td>28 (20.6%)</td>
<td>0.304</td>
</tr>
<tr>
<td>Compaction</td>
<td>502 (8.4%)</td>
<td>5 (3.7%)</td>
<td></td>
</tr>
<tr>
<td>Early blastocyst</td>
<td>492 (8.3%)</td>
<td>17 (12.5%)</td>
<td></td>
</tr>
<tr>
<td>Advanced blastocyst</td>
<td>3180 (53.4%)</td>
<td>86 (63.2%)</td>
<td></td>
</tr>
</tbody>
</table>

Type of ET

<table>
<thead>
<tr>
<th>Age</th>
<th>Singleton (n = 5960)</th>
<th>MZT (n = 136)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fresh ET</td>
<td>4555 (76.4%)</td>
<td>117 (86.0%)</td>
<td>0.008</td>
</tr>
<tr>
<td>Frozen ET</td>
<td>1405 (23.6%)</td>
<td>19 (14.0%)</td>
<td></td>
</tr>
</tbody>
</table>

Statistically significant values are highlighted in bold.

MZT, monozygotic twinning; ET, embryo transfer.

1Student's t test.
2Fisher's exact test.
3Pairwise comparisons statistically significant between cleavage embryos and both early and advanced blastocysts (P = 0.001 and 0.003, respectively), while cleavage and compaction embryos (P = 0.333) and early and advanced blastocysts (P = 0.364) did not vary significantly.

Figure 1. Incidence of monozygotic twinning (MZT) and monochorionic (MC) MZT clinical pregnancies during the study period (2004–2013). The incidence of MZT (Poisson regression coefficient 0.02, 95% CI = 0.04 to 0.01; P = 0.660) and MC MZT (Poisson regression coefficient 0.05, 95% CI = 0.03 to 0.13; P = 0.192) did not vary significantly over time. For each year included in the study, the types of culture media used are represented on the vertical axis.
Day 4 (n = 6) and frozen Day 5 ETs (n = 13) (pairwise comparison P < 0.001, 0.015 and 0.006, respectively) (Supplementary data, Table S1). All study parameters with the exception of AH (since no MZT pregnancy occurred following this procedure) were included in a multivariable logistic regression model to evaluate their independent effect on the incidence of MZT in the full sample set (6096 cycles; Table II). The transfer of frozen embryos and blastocyst-stage embryos were the only two parameters independently associated with a significant impact on the incidence of MZT in both the univariable and multivariable analyses. Specifically, the transfer of a frozen embryo approximately halved the likelihood of having an MZT pregnancy when compared with a fresh ET (adjusted odds ratio (aOR) 0.48, 95% CI 0.29–0.80). Meanwhile, the transfer of a blastocyst was associated with a significant increase in MZT risk when compared to cleavage-stage embryos (aOR 2.70, 95% CI 1.36–5.34 for early blastocysts and aOR 2.05, 95% CI 1.29–3.26 for advanced blastocysts).

Table III presents a subgroup analysis according to the type of ET. In the fresh ET subgroup, the transfer of a blastocyst was associated with an increased likelihood of MZT pregnancy compared to cleavage-stage ET. In the frozen ET subgroup, none of these relationships maintained their statistical significance in the univariable analysis. Since only 19 MZTs occurred in this subgroup, we were unable to perform multivariable logistic regression.

Regarding the type of placentation, 131 out of 136 MZT pregnancies had a clear ultrasound interpretation (Supplementary data, Fig. S9). Specifically, 80 (61.1%) were MC and 51 (38.9%) were DC/TC MZT. Both types of placentation had similar baseline characteristics (Table IV), and no association between placentation type and embryo stage or type of ET was observed. Due to the low number of MZT cases, it was impossible to assess further the potential risk factors associated with a particular placentation subtype. The proportion of MC MZT over time did not vary significantly (P = 0.92) during the study period (Fig. 1). In the MC MZT group, 5 cases (6.2%) were MC-MA, 73 cases (91.3%) were MC-DA and 2 cases (2.5%) were MC-TC. No laboratory record indicating a splitting of the ICM was found. The five MC-MA CPs were a result of a blastocyst transfer; four CPs after fresh ET (one hatching, two expanded and one collapsed blastocyst) and one after frozen ET at Day 4 (early blastocyst). The two cases of MC-DA CPs were a result of a fresh (early stage) and of a frozen (expanded stage) blastocyst transfer. In the DC/TC MZT group, two cases (3.9%) were TCTA and were a result of fresh blastocyst-stage (one hatching and one expanded) embryo.

Moreover, with regards to the delivery outcome of the 131 cases with known placentation (Table IV), the DC/TC showed a borderline significantly lower live birth delivery rate compared to MC group (P = 0.046).

**Discussion**

In our retrospective study examining a large cohort of SETs, we report a 2.2% MZT pregnancy rate following ART, comparable to previous studies (Aston et al., 2008; Knopman et al., 2014). In addition, we present a 1.5% monzygotic twin live birth rate, higher than what is reported after spontaneous conception (0.40–0.45%; Aston et al., 2008).

In view of the widely accepted explanation for the origin of monozygotic twins (Comer, 1995), we considered investigating different ART parameters for their possible involvement in embryo splitting and, thus, for their potential association with MZT.

The transfer of a frozen embryo and the transfer of a blastocyst seemed to have an impact on the incidence of MZT pregnancy. However, ovocyte origin (own or donated), female age, manipulation of the ZP and the use of a specific culture medium had no influence on the MZT rate. Furthermore, the incidence of monozygotic pregnancy did not vary significantly over time and, as compared to previous studies (Moayeri et al., 2007; Knopman et al., 2014), showed no decrease in the last years. Conversely, our results cannot confirm nor refute the hypothesis of MZT clustering previously proposed by Vaughan et al. (2016), as this sort of temporal relationship with MZT incidence cannot be assessed by linear regression.

According to our data, frozen ET was associated with a significantly lower incidence of MZT when compared to fresh ET, which contradicts previous reports (Allikani et al., 2003; Kawachiya et al., 2011; Knopman et al., 2014; Nakasuji et al., 2014). A clear reason for this discrepancy is difficult to provide. Different policies for freezing may be a possible justification: while in our centre cleavage-stage embryos are frozen on Day 3 and transferred on Day 4, information regarding the transfer of frozen embryos is lacking in other reports (Allikani et al., 2003; Knopman et al., 2014; Nakasuji et al., 2014). Furthermore, differences in the endometrial environment in ETs occurring immediately after ovarian stimulation may also explain the increased incidence of MZT following a fresh ET, and this hypothesis should be investigated further.

Regarding the effect of the embryo stage at ET, our data confirm most of the previous reports that showed a higher MZT rate after blastocyst transfer (Kawachiya et al., 2011; Knopman et al., 2014). However, they also contradict previously reported data from our centre (Papanikolaou et al., 2010), which showed no increase in MZT rate after single fresh blastocyst transfer. This divergence may be attributed to the use of a different definition for CP. Similarly to what was proposed by Knopman et al. (2014), this study also includes earlier monozygotic pregnancies, some of which may have eventually miscarried or spontaneously reduced to a singleton pregnancy. In other words, we defined

### Table II. The association between ART parameters and monozygosity.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic cycle parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (mean ± SD)</td>
<td>1.01 (0.97–1.05)</td>
<td>1.02 (0.98–1.06)</td>
</tr>
<tr>
<td>Donor oocyte</td>
<td>1.69 (0.47–7.47)</td>
<td>1.04 (0.32–3.34)</td>
</tr>
<tr>
<td>Frozen ET</td>
<td>0.53 (0.32–0.86)</td>
<td>0.48 (0.29–0.80)</td>
</tr>
<tr>
<td>Zona pellucida manipulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICSI</td>
<td>0.77 (0.49–1.23)</td>
<td>0.77 (0.48–1.24)</td>
</tr>
<tr>
<td>Embryo biopsy</td>
<td>1.33 (0.87–2.03)</td>
<td>0.97 (0.61–1.56)</td>
</tr>
<tr>
<td>Embryo stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleavage</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Compaction</td>
<td>0.63 (0.24–1.65)</td>
<td>0.91 (0.34–2.38)</td>
</tr>
<tr>
<td>Early blastocyst</td>
<td>2.20 (1.20–4.06)</td>
<td>2.70 (1.36–5.34)</td>
</tr>
<tr>
<td>Advanced blastocyst</td>
<td>1.73 (1.12–2.65)</td>
<td>2.05 (1.29–3.26)</td>
</tr>
</tbody>
</table>

*OR, univariable logistic regression odds ratio; aOR, adjusted multivariable logistic regression odds ratio.
Table III  Subgroup analysis of the impact of different ART parameters on MZT incidence according to the type of ET.

<table>
<thead>
<tr>
<th></th>
<th>Fresh (n = 4672)</th>
<th>Frozen (n = 1424)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Singleton (n = 4555)</td>
<td>MZT (n = 117)</td>
</tr>
<tr>
<td>Basic cycle parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (mean ± SD)</td>
<td>31.2 ± 4.1</td>
<td>31.4 ± 3.7</td>
</tr>
<tr>
<td>Donorized oocytes</td>
<td>98 (2.2%)</td>
<td>2 (1.7%)</td>
</tr>
<tr>
<td>Zona pellucida manipulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICSI</td>
<td>3926 (86.2%)</td>
<td>95 (81.2%)</td>
</tr>
<tr>
<td>Embryo biopsy</td>
<td>789 (17.3%)</td>
<td>25 (21.4%)</td>
</tr>
<tr>
<td>Embryo stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleavage</td>
<td>1606 (37.2%)</td>
<td>27 (23.1%)</td>
</tr>
<tr>
<td>Compaction</td>
<td>205 (4.5%)</td>
<td>2 (1.7%)</td>
</tr>
<tr>
<td>Early blastocyst</td>
<td>330 (7.2%)</td>
<td>14 (12.0%)</td>
</tr>
<tr>
<td>Advanced blastocyst</td>
<td>2224 (51.0%)</td>
<td>74 (63.2%)</td>
</tr>
<tr>
<td>Fresh culture media</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sage</td>
<td>3119 (52.3%)</td>
<td>70 (51.5%)</td>
</tr>
<tr>
<td>Medicult</td>
<td>1677 (28.1%)</td>
<td>31 (22.8%)</td>
</tr>
<tr>
<td>Vitrolife</td>
<td>986 (16.5%)</td>
<td>31 (22.8%)</td>
</tr>
<tr>
<td>Other</td>
<td>178 (3.0%)</td>
<td>4 (2.9%)</td>
</tr>
</tbody>
</table>

NE, non-estimable.
Table IV Baseline patient characteristics and known delivery outcomes in the MZT group according to the type of chorionicity.

<table>
<thead>
<tr>
<th></th>
<th>Monochorionic (n = 80)</th>
<th>Dichorionic/</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>trichorionic (n = 51)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>31.2 ± 3.5</td>
<td>31.6 ± 4.2</td>
<td>0.592</td>
</tr>
<tr>
<td>&lt;25</td>
<td>4 (5.0%)</td>
<td>1 (2.0%)</td>
<td></td>
</tr>
<tr>
<td>25–29</td>
<td>16 (20.0%)</td>
<td>15 (29.4%)</td>
<td></td>
</tr>
<tr>
<td>30–34</td>
<td>45 (56.3%)</td>
<td>22 (43.1%)</td>
<td></td>
</tr>
<tr>
<td>35–39</td>
<td>15 (18.8%)</td>
<td>11 (21.6%)</td>
<td></td>
</tr>
<tr>
<td>≥40</td>
<td>0 (0.0%)</td>
<td>2 (3.9%)</td>
<td></td>
</tr>
<tr>
<td>Recipients of donated oocytes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>79 (98.8%)</td>
<td>50 (98.0%)</td>
<td>&gt;0.999</td>
</tr>
<tr>
<td>Yes</td>
<td>1 (1.2%)</td>
<td>2 (4.0%)</td>
<td></td>
</tr>
<tr>
<td>Implantation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional IVF</td>
<td>11 (13.5%)</td>
<td>10 (19.6%)</td>
<td>0.465</td>
</tr>
<tr>
<td>ICSI</td>
<td>69 (86.5%)</td>
<td>41 (80.4%)</td>
<td></td>
</tr>
<tr>
<td>Embryo biopsy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>64 (80.0%)</td>
<td>41 (80.4%)</td>
<td>&gt;0.999</td>
</tr>
<tr>
<td>Yes</td>
<td>16 (20.0%)</td>
<td>10 (19.6%)</td>
<td></td>
</tr>
<tr>
<td>Embryo stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleaveage</td>
<td>14 (17.5%)</td>
<td>11 (21.6%)</td>
<td>0.528</td>
</tr>
<tr>
<td>Compaction</td>
<td>4 (5.0%)</td>
<td>1 (2.0%)</td>
<td></td>
</tr>
<tr>
<td>Early blastocyst</td>
<td>12 (15.0%)</td>
<td>4 (7.8%)</td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>50 (62.5%)</td>
<td>35 (68.6%)</td>
<td></td>
</tr>
<tr>
<td>Blastocyst</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of ET</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fresh ET</td>
<td>69 (86.3%)</td>
<td>43 (84.3%)</td>
<td>0.802</td>
</tr>
<tr>
<td>Frozen ET</td>
<td>11 (13.8%)</td>
<td>8 (15.7%)</td>
<td></td>
</tr>
<tr>
<td>Live birth outcome</td>
<td>71 (88.8%)</td>
<td>47 (92.2%)</td>
<td></td>
</tr>
<tr>
<td>CP with known delivery outcome</td>
<td>45 (2/43/-)</td>
<td>20 (-/9/1)</td>
<td></td>
</tr>
<tr>
<td>Twin (MA/DA/TA)</td>
<td>12 (+/12/-)</td>
<td>16 (-/16/-)</td>
<td></td>
</tr>
<tr>
<td>Singleton</td>
<td>45 (2/43/-)</td>
<td>20 (-/9/1)</td>
<td></td>
</tr>
<tr>
<td>No live birth (MA/DA/TA)</td>
<td>14 (+/12/-)</td>
<td>11 (-/10/1)</td>
<td></td>
</tr>
</tbody>
</table>

The data do not include the five CPs with an unknown implantation (two resulted in a twin live birth delivery and four resulted in a singletons live birth delivery).

*Student’s t-test.

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.
Authors’ roles


Funding

No external funding was sought or obtained for this study.

Conflict of interest

None declared.

References


Finnius JH, Dodick Y, Molinaro TA, Hong KH, Forman EJ, Werner MD, Upham KM, Scott RT. Blastocyst transfer is not associated with increased rates of monzygotic twins when controlling for embryo cohort quality. Fertil Steril 2015;103:95–100.


Matei I, Van Landuyt L, Tournaye H, Verheyen G. Deliveries of normal healthy babies from embryos originating from oooytes showing the presence of smooth endoplasmic reticulum aggregates. Hum Reprod 2014;29:311–3117.


7.3.3 Study 3-C

Compare EP rates between fresh ET and FET


Full manuscript published
Trends in ectopic pregnancy rates following assisted reproductive technologies in the UK: a 12-year nationwide analysis including 160 000 pregnancies

Samuel Santos-Ribeiro1,2, Herman Tournaye1, and NikoPos P. Polyzos1,3,*

1Centre for Reproductive Medicine, Universiteit Ziekenhuis Brussels, Vrije Universiteit Brussel, Laarbeeklaan 101, Brussels 1090, Belgium
2Department of Obstetrics, Gynaecology and Reproductive Medicine, Santa Maria University Hospital, Avruadisa Professor Egas Moniz, Lisbon 1649-035, Portugal
*Department of Clinical Medicine, Faculty of Health, University of Aarhus, Hvidovre Hospital, 8000 Aarhus N, Denmark

Correspondence address. Tel: +32-2-477-66-60; Fax: +32-2-477-66-49; E-mails: n.polyzos@gmail.com, nikolas.polyzos@vubrussels.be

Submitted on July 13, 2015; resubmitted on November 14, 2015; accepted on November 25, 2015

STUDY QUESTION: Have the advancement of assisted reproductive technologies (ART) and changes in the incidence of specific causes of infertility-affected ectopic pregnancy (EP) rates following ART over time in the UK?

SUMMARY ANSWER: EP rates in the UK following IVF/ICSI have progressively decreased, and this appears to be associated with a reduction in the incidence of tubal factor infertility and the increased use of both a lower number of embryos transferred and extended embryo culture.

WHAT IS KNOWN ALREADY: Historically, EP rates following ART are known to have increased over time. However, the impact of progression in ART procedures and changes in both policy and the incidence of specific causes of infertility on the overall EP rate in the UK has yet to be studied.

STUDY DESIGN, SIZE, DURATION: A population-based retrospective analysis was carried out on all pregnancies following ART cycles carried out in the UK between 2000 and 2012 included in the anonymized database of the Human Fertilisation and Embryology Authority.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Overall, 161 967 treatment cycles resulting in a pregnancy were included in the analysis. Among them, 8852 pregnancies occurred after intrauterine insemination (IUI) and 153 115 following IVF/ICSI.

MAIN RESULTS AND THE ROLE OF CHANCE: During this period of 12 years, ~1.1% (n = 2244) of all pregnancies following ART were an EP. Crude EP rates were significantly higher after IVF/ICSI when compared with following IUI (1.4 versus 1.1%, p = 0.043). The incidence of EP decreased significantly over time for IVF/ICSI cycles (incidence rate ratios (IRR) 0.96 per year, 95% confidence interval (CI) 0.94–0.97), but not after IUI (IRR 0.96 per year, 95% CI 0.91–1.03).

Among pregnancies resulting from IVF/ICSI, multivariable logistic regression analysis demonstrated that the major risk factor for EP was the presence of tubal infertility (adjusted odds ratio (aOR) 2.23, 95% CI 1.93–2.58), followed by the increased number of embryos transferred (aOR 1.29 for 2 versus 1 embryo transferred, 95% CI 1.11–1.50; aOR 1.69 for 3 or more versus 1 embryo transferred, 95% CI 1.35–2.11). The use of extended embryo culture to Days 3–4 or 5–7 significantly reduced the risk of EP, when compared with the transfer of early cleavage (Days 1–2) embryos (respectively, aOR 0.85, 95% CI 0.76–0.94; and aOR 0.73, 95% CI 0.63–0.84). Finally, frozen embryo transfer (ET) had no effect on the risk of EP following IVF/ICSI (aOR 0.92, 95% CI 0.76–1.14).

LIMITATIONS, REASONS FOR CAUTION: Owing to the use of this particular registry data, well-established risk factors of EP, such as smoking habits or uterine surgery, could not be assessed.

WIDER IMPLICATIONS OF THE FINDINGS: Our results provide the first evidence of a potential benefit—in terms of the reduction in EP rates—of the implementation of national programmes aiming to reduce the incidence of tubal infertility, such as the National Chlamydia Screening Programme. In addition, campaigns for the widespread introduction of single ET may not only reduce the incidence of multiple pregnancies but also the incidence of EP following IVF/ICSI.
Introduction

Ectopic pregnancy (EP), which occurs in 2% of all spontaneous conceptions (Farquhar, 2005), is the most common cause of mortality during the first trimester of pregnancy in the UK (Carnell et al., 2011) and is responsible for up to 7% of all pregnancy-related deaths in the USA (Grimes, 2006).

Anecdotally, the first successful human IVF treatment resulted in an EP (Staple and Edwards, 1976). Ever since then, assisted reproductive technologies (ART) have been consistently associated with increased EP rates, reaching as high as 8.6% (Clayton et al., 2006). Although the reasons behind this increased incidence of EP following ART are not completely understood (Shaw et al., 2010), several studies have demonstrated that this association may be related to the confounding effect of other known infertility risk factors (Brandell et al., 1999). Such risk factors, particularly tubal disease, smoking and advanced maternal age, are both more prevalent in infertile women (Brandell et al., 1999) and also described as risk factors for EP (Farquhar, 2005). Of these, tubal disease has been the focus of much debate (Roberts et al., 2007; Land et al., 2010; Aghaizu et al., 2004; Turner et al., 2011; 2014; Gillespie et al., 2010; Buit et al., 2012; Gottlieb et al., 2013; Huang et al., 2013; Filer and Ison, 2014; de Vl et al., 2015), particularly with regard to the cost effectiveness of organized screening programmes for Chlamydia trachomatis, a sexually transmitted infection (STI) associated with pelvic inflammatory disease (PID), tubal disease and EP. In 2012, a total of 1,422,976 cases of C. trachomatis (643.3 and 262.6 cases per 100,000 females and males, respectively) were reported to have occurred in the USA alone (Centers for Disease Control and Prevention, 2014), making it the most commonly reported STI in developed countries (Roberts et al., 2007).

Confounding, however, does not account for the overall increased risk of EP that is associated with ART, suggesting that other contributing factors may still exist. In an attempt to provide a reasonable explanation for this higher incidence, investigators have assessed the potential role of several other possible mechanisms. Specifically, the increased number of embryos transferred (Clayton et al., 2006), a change in the expression of the adhesion protein e-cadherin in embryos transferred in vitro (Reval et al., 2008) and a detrimental effect of the supraphysiological sex steroid levels during ovarian stimulation on both endometrial receptivity (Klikk, 2007) and tubal ciliary function (Paltiel et al., 2000) have all been proposed as plausible causes of an increased EP rate following ART. Nonetheless, most of these studies have been based on small cohorts with a very small number of events (EPs), given the relatively low prevalence of EP among all pregnancies derived from ART.

Over the last two decades, ART has changed significantly with a tendency towards fewer embryos being transferred (Thurin et al., 2004), extended embryo culture (Papanikolaou et al., 2006), milder ovarian stimulation (Kleijnen et al., 2007) and improved cryopreservation techniques (Cobo et al., 2012). In addition, nationwide campaigns have successfully managed to reduce the incidence of known risk factors for EP such as smoking (Richardson et al., 2014) or chlamydia-associated tubal disease (French et al., 2011).

A recent study attempted to evaluate whether these changes, implemented to both the ART procedure itself and the lifestyle of women of reproductive age, had any impact on the overall incidence of EP in the USA (Kerkes et al., 2011). These authors concluded that EP rates decreased between 2001 and 2011 (from 2.0 to 1.6%) and that factors such as the transfer of multiple embryos seemed to have been associated with an increased risk of EP during this period. Nonetheless, since treatment policies and pregnancy outcomes are known to vary significantly between the USA and Europe (Gleicher et al., 2007; Baxer et al., 2010), these results may not be directly generalizable to the European situation. For this reason, the aim of this study was to assess the incidence of EP over time with one of the largest European ART registries to date, maintained by the UK Human Fertilisation and Embryology Authority (HFEA), and to evaluate whether specific actions taken over this period had any impact on the incidence of EP following ART in the UK.

Materials and Methods

Data extraction

We obtained anonymized data for ART cycles carried out in the UK between 2000 and 2012 from the HFEA database. The HFEA is the statutory regulator of ART in the UK, and following a mandatory reporting policy that was enforced by law, this institution has kept records of the ART cycles performed in the UK since August 1991. This database has been regularly audited, cross checked and upgraded to ensure the consistency and relevance of the collected data (Human Fertilisation and Embryology Authority, 2007). Furthermore, since March 2002, a double-entry system complemented with a thorough verification process with ~800 validation rules was set in place to maximize the integrity of the data (Human Fertilisation and Embryology Authority, 2007).

Inclusion and exclusion criteria

In order to avoid biases related to the quality of the data inserted in the database, we only considered eligible ART cycles carried out in the UK between 2000 and 2012. The rationale for this approach was based on the fact that the collection of information regarding the types of infertility treated only began after the year 2000.

In this regard, given that the type of infertility diagnosis has been shown to affect the risk of EP (Brandell et al., 1999; Farquhar, 2005; Clayton et al., 2006), we restricted our analysis to cycles performed after 1999 in which a valid early positive pregnancy outcome was registered. To minimize bias and allow adequate adjustment for potential confounders, we excluded data on cycles which the treatment details (indications for ART, specific type of treatment, use of ovarian stimulation, type of embryo transfer (ET) and the number and/or developmental stage of the transferred embryos) were unclear or unknown. Furthermore, we limited our analysis to either intratubine insemination (IUI), or IVF/ICSI cycles, excluding other less frequent or not currently used techniques (i.e., gamete intrafallopian transfer, zygote intrafallopian transfer or intravaginal insemination). We also excluded cycles in which donor oocytes were used or if the indication for treatment was not infertility, such as IVF for PGD in patients otherwise expected to be fertile. Finally, given that few reports supported a reduced
incidence of EP following the transfer of frozen embryos (Ishibashi et al., 2011; Shapiro et al., 2012; Huang et al., 2014; Fang et al., 2015; London et al., 2015), we did not include cycles having a simultaneous transfer of both fresh and frozen embryos. The full data selection process of the cycles included in the analysis is presented in Supplementary data, Figure S1.

Definition of EP

EP was defined as the presence of at least one gestational sac outside of the uterine cavity. In our definition, we also considered heterotopic pregnancies (the simultaneous visualization of both at least one intrauterine and one extraterine gestational sac; n = 58) as being ectopic. The incidence of EP was calculated as the number of EPs as a proportion of positive pregnancy tests after ART.

Statistical analysis

The incidence of EP was compared between the two types of studied ART treatments (IUI and IVF/ICSI). Given the low incidence of EP (< 10%), we considered odds ratios (OR) as adequate estimates of risk ratios (Zhang and Yu, 1998).

The statistical significance of the trends in EP rates and relevant potential confounders observed over time were determined using the Poisson regression. All graphical depictions of trends were smoothed using moving averages, adjusting for the mean EP rates of the year before and after, in order to minimize spurious variations that could inflit the visual interpretation of the temporal trend.

We assessed the relationship between potential confounders and EP using multivariable logistic regression. The variables included in the model as potential confounders were year of treatment, female age, use of exogenous ovulation stimulation, infertility diagnosis and type of treatment (IUI or IVF/ICSI). For IVF/ICSI cycles, the type of ET (fresh-frozen or fresh), embryo development stage (day of in vitro embryo culture until transfer), and the number of embryos transferred were also accounted for via a separate logistic regression model. A P-value was considered significant when below 0.05. The statistical analysis was performed using the Stata software version 13.1 (StataCorp, College Station, TX, USA).

Results

Baseline characteristics and demographics

Of the 684,247 ART cycles that led to 2,128,877 pregnancies evaluated, 161,967 treatment cycles resulting in a pregnancy were included in the analysis following the restriction performed using the above-mentioned inclusion and exclusion criteria (Supplementary data, Figure S1). Among these, 88,522 pregnancies occurred after IUI, while 153,145 were obtained after either IVF or ICSI. Baseline patient demographics are presented in Table I, while further information regarding specifically the IVF/ICSI cycles according to the type of ET are detailed in Table II.

Rates of EP

During a period of 12 years (2000–2012), 1.4% (n = 2,244) of all pregnancies following ART were EPs. According to our unadjusted analysis, EP was higher after IVF/ICSI (1.4%) than following IUI (1.1%) (P = 0.043).

Trends in EP over time

The trend in EP following ART has been evaluated according to ART treatment (IUI versus IVF/ICSI, Fig. 1). By using the Poisson regression analysis, the incidence of EP significantly decreased over time after IVF/ICSI (incidence rate ratio (IRR) 0.96, 95% confidence interval (CI) 0.95–0.97; from 20.0 EP/1,000 cycles in 2000 to 11.5 EP/1,000 cycles in 2012), whereas no significant reduction was found following IUI (IRR 0.96, 95% CI 0.91–1.02). When assessing IVF/ICSI cycles specifically (Fig. 2), the EP rates decreased significantly over time for fresh ETs (IRR 0.96, 95% CI 0.94–0.97), but not for frozen ETs (IRR 0.97, 95% CI 0.91–1.03).

Risk factors for EP following ART procedures (IUI, IVF or ICSI)

In order to evaluate the contribution of individual risk factors to the incidence of EP, we performed logistic regression analysis (Table III). Although the univariate analysis several factors were associated with a significant variation in EP risk, after adjustment, the major risk factor for EP was the diagnosis of tubal infertility (adjusted odd ratio aOR 2.26 for tubal factor only, 95% CI 1.97–2.62). The presence of unexplained infertility was also a risk factor for EP, although the aOR was relatively weaker (aOR 1.21, 95% CI 1.06–1.38).

EP risk was not affected by the type of the ART procedure (IUI, IVF or ICSI). Specifically, the occurrence of EP did not differ significantly between IUI and IVF (aOR 1.16, 95% CI 0.91–1.51) or ICSI (aOR 0.86, 95% CI 0.73–1.16).

Finally, this analysis including all ART cycles demonstrated a significant decrease in EP rates over time even after adjusting for potential confounders (aOR 0.97 per year, 95% CI 0.95–0.98).

Risk factors for EP following IVF/ICSI

In order to evaluate the effect of different risk factors on the incidence of EP following IVF/ICSI, we performed logistic regression restricting our analysis to pregnancies resulting from these procedures. As shown in Table IV, the major risk factor for EP was the presence of tubal infertility (aOR 2.23 for tubal factor only, 95% CI 1.93–2.58), followed by the increase in the number of embryos transferred (aOR 1.29 for 2 versus 1 embryo transferred, 95% CI 1.11–1.54; aOR 1.69 for 3 or more versus 1 embryo transferred, 95% CI 1.35–2.15). The presence of unexplained infertility increased the risk of EP, even when restricting the analysis to IVF/ICSI cycles, although this effect size was again lower (aOR 1.18, 95% CI 1.03–1.36) than the one resulting from the presence of tubal factor infertility.

The use of extended embryo culture to Days 3–4 or 5–7 significantly reduced the risk of EP when compared with the transfer of early cleavage (Days 1–2) embryos (aOR 0.85 for Days 3–4 versus early cleavage embryos, 95% CI 0.77–0.94; aOR 0.73 for Days 5–7 versus early cleavage embryos, 95% CI 0.63–0.84).

Furthermore, when compared with the use of conventional IVF, performing ICSI was associated with a lower EP risk (aOR 0.78, 95% CI 0.70–0.87). To assess the use of ICSI was not confounded by the presence of male factor infertility, we added the interaction male factor infertility and ICSI to the model. This interaction was later removed from the final analysis since it was not statistically significant (aOR 0.92, 95% CI 0.72–1.16). Finally, transferring frozen-thawed embryos had no significant effect on the risk of EP (aOR 0.93, 95% CI 0.76–1.11).

Trend of factors affecting EP risk over time

As shown in Figure 3, while the proportion of patients with tubal factor significantly decreased over time (IRR OR 0.95, 95% CI 0.94–0.95), there was a significant increase in the number of patients with
Table 1: Distribution of demographic characteristics of the study population by type of treatment.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total (n = 161 967) (%)</th>
<th>IUI (n = 8 853) (%)</th>
<th>IVF/ICSI (n = 153 115) (%)</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under 35</td>
<td>53.1</td>
<td>58.5</td>
<td>52.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>35–37</td>
<td>25.3</td>
<td>23.0</td>
<td>25.3</td>
<td></td>
</tr>
<tr>
<td>38–39</td>
<td>12.8</td>
<td>11.0</td>
<td>12.9</td>
<td></td>
</tr>
<tr>
<td>40–42</td>
<td>7.8</td>
<td>6.6</td>
<td>7.8</td>
<td></td>
</tr>
<tr>
<td>43–44</td>
<td>0.9</td>
<td>0.8</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Over 44</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Infertility diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tubal factor</td>
<td>18.3</td>
<td>0.5</td>
<td>19.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ovarian factor</td>
<td>12.4</td>
<td>1.9</td>
<td>13.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>6.0</td>
<td>0.2</td>
<td>6.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male factor</td>
<td>44.6</td>
<td>70.0</td>
<td>43.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Otherwise unexplained infertility</td>
<td>25.8</td>
<td>29.4</td>
<td>25.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Use of exogenous ovarian stimulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3.8</td>
<td>63.9</td>
<td>0.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Yes</td>
<td>96.2</td>
<td>36.1</td>
<td>99.7</td>
<td></td>
</tr>
<tr>
<td>Year of treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000–2001</td>
<td>10.3</td>
<td>25.1</td>
<td>9.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2002–2003</td>
<td>12.0</td>
<td>23.9</td>
<td>11.3</td>
<td></td>
</tr>
<tr>
<td>2004–2005</td>
<td>13.4</td>
<td>20.2</td>
<td>13.0</td>
<td></td>
</tr>
<tr>
<td>2006–2007</td>
<td>16.2</td>
<td>12.9</td>
<td>16.4</td>
<td></td>
</tr>
<tr>
<td>2008–2009</td>
<td>18.2</td>
<td>6.6</td>
<td>18.9</td>
<td></td>
</tr>
<tr>
<td>2010–2011</td>
<td>19.8</td>
<td>7.2</td>
<td>20.5</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>10.1</td>
<td>4.1</td>
<td>10.4</td>
<td></td>
</tr>
<tr>
<td>Pregnancy outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intrauterine foetal pulsation</td>
<td>79.3</td>
<td>82.9</td>
<td>79.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Miscarriage</td>
<td>19.3</td>
<td>16.0</td>
<td>19.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>EP</td>
<td>1.4</td>
<td>1.1</td>
<td>1.4</td>
<td>0.043</td>
</tr>
</tbody>
</table>

Data are presented as percentage; *χ2 test. Excluding 45 cases of gestational trophoblastic disease (4 and 41 after IUI and IVF/ICSI, respectively); IUI, intrauterine insemination; ART, assisted reproductive technologies; EP, ectopic pregnancy.

unexplained infertility (IRR 1.03, 95% CI 1.03–1.03) and the number of cycles in which ICSI (IRR 1.02, 95% CI 1.01–1.02), single embryo (IRR 1.31 per year, 95% CI 1.30–1.31) or blastocyst (IRR 1.32, 95% CI 1.32–1.33) transfers were performed.

In Table V, we have summarized the predicted probabilities for EP using multivariable logistic regression. Although the same variables as presented in Table IV were included in this model, the embryo developmental stage (cleavage versus blastocyst stage) and the number of embryos transferred (single versus multiple) were dichotomized to facilitate the interpretation of the results. The range of predicted probabilities in our sample varied from 6.5 to 28.9 per 1000 pregnancies after IVF/ICSI.

Discussion

Our study has demonstrated for the first time that the crude EP rates following ART have decreased over time in the UK, practically halving during the studied period. This trend appeared to be associated with both changes in the incidence of specific causes of infertility and modifications to the ART procedure itself.

EP rates according to the cause of infertility

Our study confirms that infertile women may have an increased risk of EP, irrespective of their fertility treatment, owing to the cause of their infertility (Fanjeh, 2005). This is in line with previous studies demonstrating that up to 85% of all EPs occurring after IVF may be related to the presence of tubal disease or a previous myomectomy (Sarstedt et al., 1999). In addition, although we identified a small, albeit statistically significant, increased risk of EP in patients with no known cause of infertility, caution is needed prior to assuming an association between EP and unexplained infertility. Since previous studies have clearly demonstrated that several patients with unexplained infertility may have undiagnosed tubal damage (Dubuisson et al., 1991), one can postulate that tubal infertility might be the reason behind this increased incidence of EP in patients with unexplained infertility identified by our study. Furthermore, recent evidence has led to a reduction in the assessment of tubal patency...
Table II Gamete and embryo characteristics of IVF/ICSI cycles.

<table>
<thead>
<tr>
<th>Gamete and embryo details</th>
<th>Total (n = 153 115)</th>
<th>Fresh ET (n = 145 373)</th>
<th>Frozen ET (n = 7742)</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oocytes collected (mean ± SD)</td>
<td>11.2 ± 5.8</td>
<td>11.1 ± 4.9</td>
<td>11.3 ± 5.8</td>
<td></td>
</tr>
<tr>
<td>Embryos created (mean ± SD)</td>
<td>6.9 ± 4.1</td>
<td>6.9 ± 4.1</td>
<td>6.9 ± 4.1</td>
<td></td>
</tr>
<tr>
<td>Embryos cryo/recovered (mean ± SD)</td>
<td>1.5 ± 2.5</td>
<td>1.5 ± 2.5</td>
<td>1.5 ± 2.5</td>
<td></td>
</tr>
<tr>
<td>Embryos deriving from ICSI</td>
<td>52.7%</td>
<td>53.1%</td>
<td>44.8%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Embryo developmental stage on day of transfer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1–2</td>
<td>47.3%</td>
<td>45.1%</td>
<td>89.7%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Day 3–4</td>
<td>32.5%</td>
<td>33.8%</td>
<td>8.4%</td>
<td></td>
</tr>
<tr>
<td>Day 5–7</td>
<td>20.2%</td>
<td>21.1%</td>
<td>1.9%</td>
<td></td>
</tr>
<tr>
<td>Number of embryos transferred</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>15.6%</td>
<td>15.8%</td>
<td>12.0%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>77.6%</td>
<td>77.4%</td>
<td>80.9%</td>
<td></td>
</tr>
<tr>
<td>3 or more</td>
<td>6.8%</td>
<td>6.8%</td>
<td>7.1%</td>
<td></td>
</tr>
<tr>
<td>Pregnancy outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intraperitoneal fertilization</td>
<td>79.1%</td>
<td>79.6%</td>
<td>71.2%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Miscarriage</td>
<td>19.5%</td>
<td>19.0%</td>
<td>27.3%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ectopic</td>
<td>1.4%</td>
<td>1.4%</td>
<td>1.5%</td>
<td>0.448</td>
</tr>
</tbody>
</table>

Data are presented as a percentage, unless otherwise specified: *χ² test; †Excluding 41 cases of gestational trophoblastic disease (34 after fresh and 7 after frozen ETs, respectively).

Figure 1 EP rates per 1000 cycles of IVF/ICSI (n = 153 115) and IUI (n = 8862) in the UK between 2000 and 2012. The incidence of EP decreased significantly over time for IVF/ICSI (incidence rate ratio (IRR) 0.96, 95% CI 0.95–0.97), but not for IUI (IRR 0.96, 95% CI 0.91–1.02); plotting smoothed by moving averages.

Figure 2 EP rates per 1000 transfer cycles of fresh (n = 145 373) and frozen (n = 7742) embryos in the UK between 2000 and 2012. The incidence of EP decreased significantly over time after fresh (IRR 0.96, 95% CI 0.94–0.97), but not following frozen (IRR 0.96, 95% CI 0.91–1.03) ETs; plotting smoothed by moving averages.

(namely, with either hysteroscopy [angiography or diagnostic laparoscopy] prior to ART altogether (Bosteels et al., 2007), and such a trend may have played a role in the association of unexplained infertility with EP risk. While this approach seems to be more cost effective (Verhoeve et al., 2013), it may have left some cases of tubal infertility undiagnosed. This, coupled with the fact that other factors known to decrease ectopic endometrial receptivity, such as endometrial thickness on the day of embryo transfer or abnormal uterine contractility (Kiick, 2007; Rombaux et al., 2015), are not routinely assessed in clinical practice, renders unexplored infertility causes to be frequently labelled as ‘otherwise unexplained’.

The National Chlamydia Screening Programme (NCCP) for England, launched in 2003, has become increasingly widespread, and although recent studies have shown a reduction in the rates of definitive PID in the UK (French et al., 2011), they have failed to correlate this reduction in PID with a similar trend in the general rate of EP (Woodhall et al.,
Table III  Crude OR and aOR for EP after ART (n = 161,967).

<table>
<thead>
<tr>
<th>Female age (years)</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18–34</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>35–37</td>
<td>0.94 (0.84–1.04)</td>
<td>0.92 (0.83–1.02)</td>
</tr>
<tr>
<td>38–39</td>
<td>1.10 (0.97–1.24)</td>
<td>1.10 (0.97–1.25)</td>
</tr>
<tr>
<td>40–42</td>
<td>0.98 (0.84–1.15)</td>
<td>1.00 (0.85–1.18)</td>
</tr>
<tr>
<td>43–44</td>
<td>0.94 (0.60–1.47)</td>
<td>0.98 (0.62–1.52)</td>
</tr>
<tr>
<td>45–50</td>
<td>0.72 (0.18–2.90)</td>
<td>0.73 (0.18–2.96)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Infertility diagnosis</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male factor only</td>
<td>2.73 (2.43–3.06)</td>
<td>2.36 (1.97–2.62)</td>
</tr>
<tr>
<td>Tubal factor only</td>
<td>1.23 (1.01–1.48)</td>
<td>1.08 (0.88–1.32)</td>
</tr>
<tr>
<td>Endometriosis only</td>
<td>1.27 (0.99–1.63)</td>
<td>1.10 (0.85–1.44)</td>
</tr>
<tr>
<td>Multiple female (including tubal) factors</td>
<td>2.21 (1.71–2.85)</td>
<td>1.88 (1.43–2.46)</td>
</tr>
<tr>
<td>Multiple male (excluding tubal) factors</td>
<td>1.17 (0.92–1.47)</td>
<td>1.07 (0.84–1.37)</td>
</tr>
<tr>
<td>Male + female (including tubal) factors</td>
<td>1.83 (1.47–2.29)</td>
<td>1.69 (1.35–2.13)</td>
</tr>
<tr>
<td>Male + female (excluding tubal) factors</td>
<td>1.02 (0.88–1.28)</td>
<td>1.01 (0.80–1.27)</td>
</tr>
<tr>
<td>Otherwise unexplained infertility</td>
<td>1.36 (1.21–1.52)</td>
<td>1.21 (1.06–1.38)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Use of exogenous ovarian stimulation</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>1.13 (0.97–1.32)</td>
<td>1.13 (0.95–1.34)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of treatment</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IUI</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>IVF</td>
<td>1.26 (1.27–1.91)</td>
<td>1.16 (0.93–1.47)</td>
</tr>
<tr>
<td>ICSI</td>
<td>0.94 (0.76–1.15)</td>
<td>0.86 (0.73–1.16)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year of treatment</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per year</td>
<td>0.96 (0.95–0.97)</td>
<td>0.97 (0.95–0.98)</td>
</tr>
</tbody>
</table>

Crude ORs were calculated using univariable logistic regression, while aORs were calculated with multivariable logistic regression adjusting for female age, infertility diagnosis, the use of exogenous ovarian stimulation, type of treatment and year of treatment. CI confidence interval.

2013). Our study reveals, for the first time, indirect evidence supporting that EP rates following ART decreased over time in the UK, in line with a proportional reduction in the incidence of tubal factor infertility, a subset of the general population more likely to be affected by preventive policies such as the NCSP.

Modifications to the ART procedure and the risk of EP

Our study provides evidence demonstrating that progress made in the ART procedure had a substantial impact on the reduction of EP over time in the UK. Meanwhile, according to the Centre for Maternal and Child Enquiries report, the incidence of EP in the general UK population during 2000–2008 remained between 11.0 and 11.3 cases per 1000 pregnancies (Cartwell et al., 2011). Such a reduction is likely to be more pronounced with the reduction in tubal factor infertility, the various changes implemented in ART during recent years as an attempt to enhance euploic pregnancy rates and neonatal outcomes may have ultimately decreased EP rates as well.

Extended in vitro culture until blastocyst stage prior to ET has been increasingly used owing to a significant increase in pregnancy rates after IVF/ICSI (Papankolsou et al., 2006). Our data highlight an additional benefit, namely a lower EP risk. Previous literature regarding this topic, however, has been rather controversial. While initial studies using data from cycles performed between 1998 and 2003 failed to support this hypothesis (Milki and Jun, 2003; Zegers et al., 2007), most of the available evidence suggests that EP rates following blastocyst transfer did not vary significantly (aOR 1.07, 95% CI 0.98–1.16) from those occurring after transfer of cleavage stage embryos (Perkins et al., 2015). On the other hand, another recent US registry analysis concluded that the EP rates after blastocyst transfer were lower than those occurring after transfer of cleavage stage embryos (Perkins et al., 2015). This discrepancy raises one to question if recent differences in laboratory techniques/stage of development between the USA and UK during extended embryo culture may not be ultimately contributing to a differential risk of EP.

As far as the number of transferred embryos is concerned, the ideal outcome after assisted reproduction has progressively shifted from live birth delivery to “birth emphasizing a successful singleton at term” (BSETT) (Min, 2004). While an increasing number of countries, such as Sweden (Karlstrom and Bergh, 2007) and Belgium (Van Laer et al., 2006), have enforced single ET policies to limit the amount of embryos
Table IV Crude OR and aOR for EP after IVF/ICSI (n = 153115).

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18–34</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>35–37</td>
<td>0.94 (0.85–1.04)</td>
<td>0.91 (0.82–1.01)</td>
</tr>
<tr>
<td>38–39</td>
<td>1.10 (0.97–1.25)</td>
<td>1.07 (0.94–1.22)</td>
</tr>
<tr>
<td>40–42</td>
<td>0.97 (0.82–1.14)</td>
<td>0.87 (0.73–1.04)</td>
</tr>
<tr>
<td>43–44</td>
<td>0.93 (0.81–1.07)</td>
<td>0.83 (0.72–1.02)</td>
</tr>
<tr>
<td>45–50</td>
<td>0.74 (0.68–2.99)</td>
<td>0.66 (0.61–2.66)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Infertility diagnosis</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male factor only</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Tubal factor only</td>
<td>2.74 (2.41–3.10)</td>
<td>2.33 (1.93–2.80)</td>
</tr>
<tr>
<td>Ovulatory factor only</td>
<td>1.23 (1.02–1.40)</td>
<td>1.08 (0.87–1.32)</td>
</tr>
<tr>
<td>Endometriosis only</td>
<td>1.28 (0.99–1.64)</td>
<td>1.08 (0.83–1.41)</td>
</tr>
<tr>
<td>Multiple female (including tubal) factors</td>
<td>2.22 (1.71–2.87)</td>
<td>1.85 (1.41–2.43)</td>
</tr>
<tr>
<td>Multiple female (excluding tubal) factors</td>
<td>1.17 (0.91–1.48)</td>
<td>1.07 (0.84–1.36)</td>
</tr>
<tr>
<td>Male + female (including tubal) factors</td>
<td>1.85 (1.48–2.32)</td>
<td>1.69 (1.34–2.12)</td>
</tr>
<tr>
<td>Male + female (excluding tubal) factors</td>
<td>1.01 (0.80–1.28)</td>
<td>1.00 (0.79–1.26)</td>
</tr>
<tr>
<td>Otherwise unexplained infertility</td>
<td>1.36 (1.20–1.53)</td>
<td>1.18 (1.03–1.36)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of cryopreservation</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional IVF</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>ICSI</td>
<td>0.60 (0.55–0.66)</td>
<td>0.78 (0.70–0.87)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of ET</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cryopreserved/thawed</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Fresh</td>
<td>1.08 (0.95–1.29)</td>
<td>0.82 (0.72–0.94)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Embryo developmental stage on day of transfer</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days 1–2 (early cleavage)</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Days 3–4 (late cleavage)</td>
<td>0.82 (0.75–0.90)</td>
<td>0.85 (0.77–0.94)</td>
</tr>
<tr>
<td>Days 5–7 (blastocyst)</td>
<td>0.62 (0.53–0.73)</td>
<td>0.73 (0.66–0.80)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of embryos transferred</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>2</td>
<td>1.48 (1.29–1.70)</td>
<td>1.29 (1.11–1.50)</td>
</tr>
<tr>
<td>3 or more</td>
<td>1.55 (1.41–2.36)</td>
<td>1.69 (1.35–2.11)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year of treatment</th>
<th>Crude OR (95% CI)</th>
<th>aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per year</td>
<td>0.96 (0.94–0.97)</td>
<td>0.97 (0.98–1.01)</td>
</tr>
</tbody>
</table>

Crude ORs were calculated using univariate logistic regression, whilst aORs were calculated with multivariate logistic regression adjusting for age, infertility diagnosis, type of IVF, type of ET, embryo developmental stage, the number of embryos transferred and year of treatment.

In fact, the use of both conventional IVF and ICSI has become increasingly more liberal (Boulet et al., 2015), and the application of these procedures in couples with milder forms of infertility may partially explain both the increase of unexplained infertility and the decrease in the overall incidence of EP overtime. If such an assumption is true, then our analysis may have artificially associated a protective effect with the ICSI procedure simply due to unmeasured confounding.

Finally, we need to highlight that our research contrasts with previous studies indicating that frozen ET may reduce EP pregnancy rates. While some researchers have also failed to show a relationship between EP and the type of ET (Decuir et al., 2014), others have concluded that frozen-thawed ETs seem to be associated with a lower risk of EP when compared with fresh cycles (Bhikara et al., 2011; Shapiro et al., 2012;...


Huang et al., 2014), including a recent US population-based study (Londra et al., 2015). Our results did not show any association between the type of ET and EP rates in both the univariable and multivariable analyses. Furthermore, crude EP rates seemed to have declined over time solely for fresh ET cycles, a result that contrasts strongly with US registry data, which show a sudden, otherwise unexplained and sharp decrease of EP after frozen embryo since 2004 (Perkins et al., 2015).

Strengths, limitations, need for future research and main conclusions of the study

The major strength of the current study is that it is the first to provide an insight into the trends in the incidence of EP following ART over a period of 12 years in the UK and to investigate the evolution of the incidence of known EP risk factors over time and their effect on the nationwide incidence of ART-related EP.

Besides utilizing one of the largest and most complete ART datasets worldwide (the HFEA registry), we also performed a robust analysis with multiple adjustments for potential confounders in order to minimize the likelihood of confounding. In this regard, we believe that our study provides a superior insight regarding the trends in the incidence of EP in the UK and its evolution over time.

However, we must also highlight the significant limitations of our study and why it should be interpreted with caution.

First, although the HFEA registry contained data from all cycles in the UK up to 2009, from October 2009 onwards patients had the option to not allow the use of their anonymized data for research purposes. Thus, it is highly likely that a substantial proportion of cycles are not presented in the results after that period. Although this means that the sample used from 2010 onwards in under-representing the total UK population undergoing ART treatment, we have no specific reason to believe that the omission of these cycles might have significantly biased our results. Furthermore, since the data were anonymized, we could not account for the clustering of events by patient or treatment centre.

Second, owing to its retrospective design, several data in the dataset were either missing, censored broadly, conflicting or even potentially incorrect. To this extent, it is important to stress that all data included after March 2002 underwent the before-mentioned double-entry and multiple-rule validation process. Furthermore, the data collected prior to 2006 were covered by the Histric Audit Project which followed up

**Figure 3** Trend of factors affecting occurrence of EP after IVF/ICSI (n = 131115) in the UK between 2000 and 2012. Tubal factor significantly decreased over time (IRR 0.95, 95% CI 0.94–0.95), while unexplained infertility (IRR 1.03, 95% CI 1.03–1.03) and the number of cycles in which either ICSI (IRR 1.02, 95% CI 1.02–1.02), single embryo (IRR 1.31, 95% CI 1.30–1.33) or blastocyst (IRR 1.32, 95% CI 1.32–1.33) transfers were performed significantly increased; plotting smoothed by moving averages.

**Table V** Predicted cases of EP/1000 pregnancies after IVF/ICSI according to type of IVF, embryo developmental stage, the number of embryos transferred and indication for ART (n = 131115).

<table>
<thead>
<tr>
<th>Indication for IVF/ICSI</th>
<th>Predicted EP/1000 pregnancies</th>
<th>After conventional IVF</th>
<th>After ICSI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Single embryo transferred</td>
<td>Single embryo transferred</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cleavage</td>
<td>Blastocyst</td>
</tr>
<tr>
<td>Tubal factor infertility:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tubal factor only</td>
<td>28.9</td>
<td>23.1</td>
<td>22.8</td>
</tr>
<tr>
<td>Multiple female (including tubal) factors</td>
<td>24.2</td>
<td>19.3</td>
<td>19.1</td>
</tr>
<tr>
<td>Male + female (including tubal) factors</td>
<td>22.0</td>
<td>17.6</td>
<td>17.4</td>
</tr>
<tr>
<td>Without tubal factor infertility:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Otherwise unexplained infertility:</td>
<td>15.5</td>
<td>12.3</td>
<td>12.2</td>
</tr>
<tr>
<td>Endocrineosious only:</td>
<td>14.2</td>
<td>11.3</td>
<td>11.2</td>
</tr>
<tr>
<td>Ovulotory factor only:</td>
<td>14.1</td>
<td>11.2</td>
<td>11.1</td>
</tr>
<tr>
<td>Multiple female (excluding tubal) factors:</td>
<td>14.0</td>
<td>11.2</td>
<td>11.1</td>
</tr>
<tr>
<td>Male + female (excluding tubal) factors:</td>
<td>13.1</td>
<td>10.5</td>
<td>10.4</td>
</tr>
<tr>
<td>Male factor only:</td>
<td>13.1</td>
<td>10.4</td>
<td>10.3</td>
</tr>
</tbody>
</table>

Predicted cases/1000 pregnancies calculated after logistic regression adjusting for female age, infertility diagnosis, type of IVF, type of ET, year of treatment, embryo developmental stage (cleavage versus blastocyst) and the number of embryos transferred (single versus multiple).
outcomes of these treatment cycles, offering additional assurance to the
group of women with a history of tubal infertility. This is a significant
study that provides important insights into the treatment outcomes for
women with a history of tubal infertility.

Conflict of interest
None declared.

References
the cost of pelvic inflammatory disease during pregnancy? An analysis of the
Barker VL, Jones CE, Cornette B, Hoeller F, Salle B, Urbancsik J, Soules MR.
Factors affecting success rates in two concurrent clinical IVF trials: an analysis of
potential explanations for the difference in pregnancy rates between the United
Boyes J, Van Herendael B, Weyers S, D’Hooghe T. The position of diagnostic
of use of reproductive outcomes associated with intracytoplasmic sperm
Carterwell RL, Chuwun-Brock T, Cooper G, Dawson A, Drife J, Garrett D,
Harper A, Hubbert D, Lucas S, McClure J et al. Saving mothers’ lives:
reviewing maternal deaths to make motherhood safer, 2006–2008. The
eighth report of the confidential enquiries into maternal deaths in the
Centers for Disease Control and Prevention. Sexually Transmitted Disease,
Department of Health and Human Services, 2014.
Champion HB, Schieve LA, Peterson HB, Jamieson DJ, Reynolds MA, Wright VC.
Ectopic pregnancy risk with assisted reproductive technology procedures.
Cobo A, de la Serna M, Castell D, Gamin P, Campos P, Remohi J. Outcomes of
Deckler W, Osmanullah K, Moganik G, Grove P. Slightly lower incidence of
de Vos GC, Visser EA, Schriek DR, van Bergen JE, van den Broek PV, van der
Sande MA, Wite R, Op de Coul EL, Kretzschmar ME. Chlamydia screening is
not cost-effective at low participation rates: evidence from a
Dobson WJ, Aults T, Mattes M, Foulad L, Mendeleto L, de Jolijenh:JB.
Risk factors for ectopic pregnancy in 556 pregnancies after in vitro
fertilization: implications for preventive management. Fertil Steril 1991;
56:690–699.
Feng C, Huang R, Wei LN, Jia L. Frozen-thawed day 5 blastocyst transfer is
associated with a lower risk of ectopic pregnancy than day 4 transfer and
Fletcher I, Howson CA. Nuclear acid amplification tests for the diagnosis of
Neisseria gonorrhoeae in low-prevalence settings: a review of the evidence. Sex
Transm Infect 2014;90:577–579.
French CE, Hughes G, Nicholison A, Yung M, Ross JD, Williams T, Solkan K.
Estimation of the rate of pelvic inflammatory disease diagnoses: trends in
Gillespie D, O’Neill C, Adams T, Turner K, O’Donovan D, Bhargav R,
Vaughn D, O’Connell E, Cormican M, Balle M et al. The cost and

Supplementary data
Supplementary data are available at http://humrep.oxfordjournals.org/

Authors’ roles
N.P.P. developed the study concept and designed the study. N.P.P. and
S.S.R. acquired the data for analysis and wrote the first draft of the manu-
script. S.S.R. performed the statistical analysis. All authors contributed
in the interpretation of the data and provided critical revisions for important
intellectual content. approved the final version to be published and agree
and are accountable for all aspects of the work in ensuring that questions
related to the accuracy or integrity of any part of the article are appropri-
ately investigated and resolved.

Funding
None.


Min JK. What is the most relevant standard of success in assisted reproduction? The singleton, term gestation, live birth rate per cycle initiated: the BESTS endpoint for assisted reproduction. Hum Reprod 2006;21:137–137.


7.4 Specific aim 4

Determine the temporal extent of the hindering effect of COS on ER

7.4.1 Study 4-A

Following a failed fresh ET


Full manuscript published
To delay or not to delay a frozen embryo transfer after a failed fresh embryo transfer attempt?

Samuel Santos-Ribeiro, M.D.,¹,² Johannie Siffain, M.D.,³ Nikolaos P. Polyzos, M.D., Ph.D.,¹,²,³ Arne van de Vijver, M.D.,⁴ Lisbet van Landuyt, M.Sc.,⁵ Dominic Stoop, M.D., Ph.D.,⁶ and Christophe Blockeel, M.D., Ph.D.⁷

¹ Center for Reproductive Medicine, Universiteit Ziekenhuis Brussel, Vrije Universiteit Brussel, Brussels, Belgium; ² Department of Obstetrics, Gynecology, and Reproductive Medicine, Santa Maria University Hospital, Lisbon, Portugal; and ³ Department of Clinical Medicine, Faculty of Health, University of Aarhus, Aarhus, Denmark

Objective: To evaluate if increasing the interval between a failed fresh embryo transfer and a subsequent frozen embryo transfer (FET) cycle has any effect on clinical pregnancy rates (CPRs).

Design: Retrospective cohort study.

Setting: University-based tertiary referral center.

Patients: Women who underwent at least one FET after ovarian stimulation for in vitro fertilization (IVF) and a failed fresh embryo transfer attempt from January 2010 to November 2014. We divided our sample according to the “timing” of the first FET (TF-FET), defined by the interval between oocyte retrieval and the FET cycle start date. The start of the FET was classified as either immediate (≤22 days after oocyte retrieval) or delayed (>22 days after oocyte retrieval).

Intervention(s): None.

Main Outcome Measure(s): CPR after the first FET.

Results: A total of 1,183 FET cycles (performed in 1,087 women) were included in our study. No significant differences were found between the immediate and delayed FET groups regarding age, number of oocytes retrieved, number of good-quality embryos produced, embryo developmental stage at FET, and number of frozen embryos transferred. Most importantly, the CPRs of the first FET did not differ significantly according to the TF-FET (22.9% after immediate FET vs. 31.7% after delayed FET), even after adjusting for potential confounding with the use of multivariable logistic regression.

Conclusion(s): FETs performed immediately after fresh IVF cycles had CPRs similar to those postponed to a later time. Therefore, deferring FETs may unnecessarily prolong time to pregnancy. (Fertil Steril © 2016 by American Society for Reproductive Medicine.)

Key Words: Frozen embryo transfer, endometrial receptivity, time to pregnancy, assisted reproduction, embryo cryopreservation

Discuss: You can discuss this article with its authors and with other ASRM members at http://fertstertforum.com/santosribeiros-delay-fet-increase-cpr/
to minimize any conceivable residual effect that ovarian stimulation may have on endometrial receptivity [17]. However, the literature on this matter is rather scarce [16, 19]. For this reason, although this empirical decision may be based on the best of intentions, the elective deferral of FETs may unnecessarily frustrate couples who wish to become pregnant as soon as possible.

The objective of the present study was to evaluate if increasing the interval between a failed fresh embryo transfer and a subsequent FET cycle has any effect on clinical pregnancy rates (CPRs).

MATERIALS AND METHODS
Study Population and Design

We performed a retrospective cohort study including all women who underwent at least one FET after ovarian stimulation for in vitro fertilization (IVF) from January 2010 to November 2014 at our center. Approval to retrieve and analyze the data was provided by the Ethics Committee of Brussels University Hospital (Dutch-Speaking Free University of Brussels).

Only the outcomes of the first FET cycles performed after ovarian stimulation and a failed fresh embryo transfer attempt were assessed. To minimize bias, we included only FETs that followed fresh cycles in which a GnRH antagonist and hCG alone were administered for down-regulation and ovulation triggering, respectively.

Women who were acceptors of donated oocytes or performed either in vitro maturation or blastocyst biopsy for preimplantation genetic diagnosis were excluded from the study. Furthermore, if during the preceding ovarian stimulation cycle ovulation was triggered with a drug other than hCG (e.g., a GnRH agonist, either alone [20] or in combination with hCG [21]) or hCG was administered for reasons other than ovulation triggering (e.g., for late follicular ovarian stimulation [22] or luteal phase support [23]), those cycles were also disregarded. Finally, FET cycles performed under GnRH agonist down-regulation or with concomitant exogenous ovarian stimulation also were excluded from the sample.

Ovarian Stimulation Performed during the Preceding Failed Fresh Embryo Transfer Cycles

Ovarian stimulation was initiated on day-2 of the menstrual cycle with either recombinant FSH (rFSH; Gonal-F [Merck Serono Pharmaceuticals], Puregon [Merck Sharp and Dohme], or Eleva [Merck Sharp and Dohme]) or highly purified hMG (hp-hMG; Menopur [Ferring Pharmaceuticals]). Pituitary down-regulation was performed by means of daily administrations of either cetorelix (Cetrotide; Merck Serono Pharmaceuticals) or ganirelix (Orgalutran; Merck Sharp and Dohme) starting from day 7 of the menstrual cycle. Cycles were monitored with the use of serial vaginal ultrasound scans and serum determination of E2, P, LH, and FSH. Whenever necessary, dose adjustments of rFSH/hp-hMG were performed according to ovarian response.

As soon as three follicles with mean diameters ≥ 17 mm were observed, final oocyte maturation and ovulation were triggered with the use of hCG (5,000–10,000 IU highly purified urinary hCG [Pregnyl; Merck Sharp and Dohme] or 250 IU recombinant hCG [Ovitrelle; Merck Serono Pharmaceuticals]).

Oocyte Retrieval, Insemination, Embryo Quality Assessment, and Cryopreservation

Cumulus-oocyte complexes were collected by means of transvaginal aspiration ~36 hours after triggering. The insemination of the collected oocytes was performed with the use of either conventional IVF or intracytoplasmic sperm injection (ICSI). Fertilization was assessed ~18 hours after insemination, and from then onward embryo development was graded daily until embryo transfer or cryopreservation according to the following parameters: number and size of blastomeres, rate of fragmentation, multiseedation of the blastomeres, and early compaction. Blastocyst quality on day 5/6 was assessed according to the criteria proposed by Schoolcraft et al. [24].

Good-quality embryos that were not used for the failed fresh embryo transfer attempt were cryopreserved by means of vitrification with the use of a closed vitrification system with high-security straws (CBS-VIT-HS; Cryobio-systems) in combination with dimethylsulfoxide and ethylene glycol bis (succi-inidyl) succinate as cryoprotectants (Irvine Scientific Freeze Kit; Irvine Scientific), as described by van Landuyt et al. [25]. Embryos were vitrified as cleavage-stage embryos on day 3 or full-to-expanded blastocysts on day 5 or 6 of embryo culture. Day 3 embryos were warmed the day before FET and transferred as day 4 embryos in day 4 endometrium. Day 5/6 blastocysts were warmed in the morning of the day of transfer and transferred in day 5 endometrium.

Endometrial Preparation for the FET

The FETs took place in either a natural or an artificially supplemented cycle monitored by both pelvic ultrasound and blood sampling of E2, P, LH, and FSH. In a natural cycle, ovulation occurred either spontaneously (detected by means of serial plasma LH assessments until a LH peak was noted) or artificially triggered (with the use of 5,000 IU hCG, as soon as one follicle ≥ 17 mm and endometrial thickness ≥ 7 mm were observed). In artificially supplemented cycles, preparation of the endometrium consisted of sequential administration of E2 valerate and micronized vaginal P as previously described [26]. In brief, we administered 2 mg E2 valerate twice per day (Progynova; Bayer-Schering Pharma) for 7 days, followed by 6 days 2 mg E2 valerate three times per day. On day 13, endometrial thickness was measured by means of ultrasound scan. If the endometrial thickness was ≥7 mm, supplementation with 200 mg micronized vaginal P (Utrogestan; Besins) three times per day was initiated. If the endometrial thickness was <7 mm, patients continued to take 2 mg E2 valerate orally three times per day until the endometrium thickness was ≥7 mm, at which point P supplementation was started.
The “Timing” of the First Frozen Embryo Transfer

The “timing” of the first FET (TF-FET) was defined as the interval between oocyte retrieval and the start of the first FET cycle. We divided our sample in cycles with either an immediate (≤22 days after oocyte retrieval) or delayed (>22 days after oocyte retrieval) start of FET cycle (Fig. 1). This cutoff was devised by adding the interval between oocyte retrieval and the first pregnancy test (15 days) to an extra interval of up to 7 days necessary for the patients to have their withdrawal bleeding and begin their first FET cycle. By using these intervals, we essentially divided our sample into: 1) women who had an immediate FET; and 2) women who waited at least one menstrual cycle before having their transfer.

Embryos were transferred under ultrasound guidance with the use of a 9.8-French catheter (Cook). The choice to transfer one or two embryos was decided by the clinician depending on patient age and according to Belgian law (27).

Main Outcome Measure and Statistical Analysis

Basic demographic characteristics were compared between the women who underwent immediate and delayed FET, with the use of the Student’s t-test for continuous and of the chi-square (χ²) (for categoric variables) tests.

Clinical pregnancy, defined by the International Committee for Monitoring Assisted Reproductive Technology as the visualization of a gestational sac during transvaginal ultrasound at 7 weeks of gestational age (28), was the main outcome of our study. Our secondary outcome was live birth after 24 weeks, with unknown outcomes (patients lost to follow-up) being considered as negative.

CPR and live birth rates per FET were assessed both crudely and with the use of multivariable logistic regression accounting for the following known potential confounders for FET cycle outcome: the woman’s age, number of good-quality embryos produced, type of FET cycle, stage and number of embryos transferred, and quality of the best embryo transferred. Crude and adjusted odds ratios (ORs) were estimated, adjusting the standard errors to eventually allow for more than one fresh cycle performed in the same women to be included in the analysis.

A P value was considered to be significant at <.05. For the statistical analysis, we used Stata software version 13.1 (StataCorp).

RESULTS

A total of the 1,183 first FET cycles (performed in 1,087 women) were included in the analysis. The indications for IVF included: male-factor infertility (n = 589; 49.8%), tubal-factor infertility (n = 164; 13.9%), ovulatory disorders (n = 109; 9.2%), endometriosis (n = 61; 5.2%), and otherwise unexplained infertility (n = 346; 29.3%). The majority of FET cycles (n = 986; 83.4%) were initiated after a waiting period of >22 days after oocyte retrieval, regardless of the year of treatment (Fig. 2).

Patient Demographics and General Characteristics of the Treatment Protocol

The baseline characteristics of the preceding ovarian stimulation and IVF cycles according to TF-FET are presented in Table 1. No significant differences were found between the groups regarding age, total dose of exogenous FSH administered, number of oocytes retrieved, and number of good-quality embryos either produced or used during the failed fresh embryo transfer attempt.

FIGURE 1

Study groups according to timing of first frozen embryo transfer (FET). The FET cycles were divided in either (A) immediate (≤22 days after oocyte retrieval) or (B) delayed (>22 days after oocyte retrieval).

**Relationship between TF-FET and FET Pregnancy Outcomes**

Further details regarding the FET cycles are presented in Table 2, including the pregnancy outcome. The embryo developmental stage at transfer, type of FET cycle, and number of embryos transferred did not vary between the immediate and delayed FET groups. Regarding CPR per FET, these did not differ significantly according to TF-FET (12.9% after immediate FET versus 11.7% after delayed FET, $P = .830$), even after adjusting for age, number of good-quality embryos produced, type of FET cycle, stage and number of embryos transferred, and quality of the best embryo transferred with the use of multivariable logistic regression (predicted probabilities of 32.6% for immediate FET versus 31.7% for delayed FET; $P = .803$; crude and adjusted ORs are presented in Supplemental Table 1 [available online at www.fertstert.org]).

Of the 377 clinical pregnancies in our sample, 280 had a live delivery after 24 weeks, 86 had no live birth, and 5 were lost to follow-up. Live birth rates did not vary significantly between groups (24.4% after immediate FET versus 24.1% after delayed FET; $P = .946$), even after accounting for the potential confounders (predicted probabilities of 24.5% for immediate FET versus 24.1% for delayed FET; $P = .895$).

**DISCUSSION**

This is, to our knowledge, the first study to comprehensively assess the trends and effects of FET scheduling during ART, revealing that the intentional postponement of FET cycles occurred frequently and did not enhance pregnancy outcomes.

Because delaying FETs may be a potential source for ART-related patient stress and treatment discontinuation (29, 30), we considered that a broader understanding of the motives behind such a frequent decision deserved further scrutiny. Our failure to show any clinical expression of a residual effect of ovarian stimulation on the endometrial receptivity of a subsequent cycle may reassure physicians who might otherwise hesitate to schedule FETs without delay.

On the other hand, patients may opt to purposely delay an FET cycle owing to a number of clinically unrelated reasons. To this extent, a scientific presentation including 271 FET cycles in 2008 was the first to propose that delaying FET cycles might actually even reduce the chances of achieving pregnancy, because delaying embryo transfer was associated with a significant absolute difference of 14.2% in CPR (35.2% for immediate FET versus 21.0% for delayed FET; $P < .01$) (17). However, we consider that our larger sample

**TABLE 1**

**Baseline characteristics of the ovarian stimulation and IVF cycles according to timing of first frozen embryo transfer (FET).**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Immediate FET (n = 197)</th>
<th>Delayed FET (n = 986)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woman's age (y)</td>
<td>32.4 ± 4.4</td>
<td>32.5 ± 4.3</td>
<td>.697</td>
</tr>
<tr>
<td>Total dose of exogenous FSH (IU)</td>
<td>1,562 ± 493.7</td>
<td>1,605 ± 553.2</td>
<td>.333</td>
</tr>
<tr>
<td>Oocytes retrieved</td>
<td>11.1 ± 6.1</td>
<td>10.4 ± 5.5</td>
<td>.135</td>
</tr>
<tr>
<td>Good-quality embryos produced</td>
<td>4.5 ± 2.4</td>
<td>4.4 ± 2.5</td>
<td>.829</td>
</tr>
<tr>
<td>Embryos transferred in the failed fresh cycle</td>
<td>1.3 ± 0.5</td>
<td>1.2 ± 0.4</td>
<td>.338</td>
</tr>
</tbody>
</table>

Note: Immediate FET occurred ≤22 days after oocyte retrieval, and delayed FET occurred >22 days after oocyte retrieval.

---

TABLE 2

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Immediate FET (n = 197)</th>
<th>Delayed FET (n = 986)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embryo developmental stage at transfer, n (%)</td>
<td>95 (48.2)</td>
<td>472 (49.7)</td>
<td>.928</td>
</tr>
<tr>
<td>Cleavage stage</td>
<td>102 (51.8)</td>
<td>514 (50.3)</td>
<td>.565</td>
</tr>
<tr>
<td>Type of FET cycle, n (%)</td>
<td>35 (17.8)</td>
<td>181 (18.4)</td>
<td>.496</td>
</tr>
<tr>
<td>Artificially supplemented cycle</td>
<td>89 (44.1)</td>
<td>396 (40.1)</td>
<td></td>
</tr>
<tr>
<td>Natural cycle (spontaneous LH peak)</td>
<td>75 (38.1)</td>
<td>409 (41.5)</td>
<td>.854</td>
</tr>
<tr>
<td>Embryo quality of best embryo transferred, n (%)</td>
<td>122 (61.9)</td>
<td>603 (61.2)</td>
<td>.665</td>
</tr>
<tr>
<td>1</td>
<td>54 (27.4)</td>
<td>264 (26.8)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>21 (10.7)</td>
<td>119 (12.1)</td>
<td></td>
</tr>
<tr>
<td>Pregnancy outcomes, n (%)</td>
<td>1.4 ± 0.5</td>
<td>1.4 ± 0.5</td>
<td>.895</td>
</tr>
<tr>
<td>Crude clinical pregnancy rate per FET</td>
<td>64 (32.5)</td>
<td>313 (31.7)</td>
<td>.838</td>
</tr>
<tr>
<td>Adjusted clinical pregnancy rate per FET, %a</td>
<td>32.6</td>
<td>31.7</td>
<td>.802</td>
</tr>
<tr>
<td>Crude live birth delivery rate per FET</td>
<td>48 (24.4)</td>
<td>238 (24.1)</td>
<td>.946</td>
</tr>
<tr>
<td>Adjusted live birth delivery rate per FET, %a</td>
<td>24.5</td>
<td>24.1</td>
<td>.854</td>
</tr>
</tbody>
</table>

Note: *immediate FET occurred within 12 days after oocyte retrieval, and delayed FET occurred >22 days after oocyte retrieval. Data are presented as n (%) unless otherwise specified.

a Predicted probability using multivariable logistic regression adjusting for women’s age, number of good-quality embryos produced, quality of the best embryo transferred, type of FET cycle, and stage of frozen embryo transfer (unavailable and multivariable odds ratios are presented in Supplemental Table 1).

b Cycles without a live birth outcome (n = 58) were considered to be normal births.


and confounder-adjusted analysis offer a more accurate inference that may serve as a better basis to counsel women seeking to temporarily postpone their next FET cycle. Furthermore, our results agree and add to the already existing body of evidence showing, so far, a lack of an effect of the duration of the embryo cryopreservation (31) and uterine ageing (32) on FET pregnancy outcomes.

Although the present study included a large sample size and adjusted for potentially confounding differences between the groups, it is limited by its retrospective nature and by the possibility of unmeasured confounding. We considered that a retrospective design was the most appropriate for this research hypothesis because we had ethical reservations about performing a prospective study offering a treatment modality which, at first glance, had a biological plausibility of being detrimental and seemed likely an inferior alternative. Furthermore, one can also assume that such a clinical trial would be difficult to conduct, because a noninferiority trial capable of detecting even the smallest difference we found (1.2%) between immediate and delayed FET (32.5% vs. 31.7%) would require >40,000 cycles per group to achieve a reasonable 80% power with a significance level of 5%. Regarding unmeasured confounding, we were unable to account for all possible confounders, such as smoking and body mass index (BMI). However, it is unlikely that patients with heavy smoking habits or extreme BMIs would be proposed different FET scheduling schemes based on these characteristics alone.

Finally, we should also reiterate that this study evaluated only the effect of FET-FET on CPRs and live birth rates in GnRH antagonist down-regulated cycles and that our results should not be assumed as valid surrogates for the potential carryover effect of ovarian stimulation on other pregnancy outcomes (such as preterm birth, birth weight, and fetal development) nor following GnRH agonist-suppressed ovarian stimulation cycles. To this extent, a previous study providing translational evidence to support that endometrium exposed to ovarian stimulation with the use of GnRH antagonist treatment mimics natural endometrium better than one exposed to GnRH agonist down-regulation (33) can not be ignored when attempting to extrapolate our results.

CONCLUSION

This study provides the first potential answer to a very frequent question posed by couples seeking parenthood with the use of IVF: “Will waiting before performing my FET cycle increase my chances to become pregnant?” Ovarian stimulation did not seem to have a carryover effect on CPR per FET, allowing patients to opt to perform their FET cycle either without delay or at their own convenience, potentially reducing the frustration associated with the various waiting periods of IVF treatment.

REFERENCES

Results (Study 4-A)


7.4.2 Study 4-B

Following a “freeze-all protocol”


Full manuscript published
The effect of an immediate frozen embryo transfer following a freeze-all protocol: a retrospective analysis from two centres

Samuel Santos-Ribeiro1,2,*, Nikolaos P. Polyzos1,3, Vuong Thi Ngoc Lan4,5, Johannie Siffain1, Shari Mackens1, Lisbet Van Landuyt1, Herman Tournaye1, and Christophe Blockeel1,6

1Centre for Reproductive Medicine, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel, Brussels 1090, Belgium. 2Department of Obstetrics, Gynaecology and Reproductive Medicine, Santa Maria University Hospital, Avenida Professor Egas Moniz, Lisbon 1649-035, Portugal. 3Department of Clinical Medicine, Faculty of Health University of Aarhus, Incubus/Skoleby, 562, Polle Juul Jensens Boulevard 82, 8200 Aarhus N, Aarhus, Denmark. 4Department of Obstetrics & Gynaecology, University of Medicine and Pharmacy at Ho Chi Minh City, 217/19 Hang Bang, District 5, Ho Chi Minh City, Vietnam. 5VPMD, My Duc Hospital, 4 Nui Thanh, District Tan Binh, Ho Chi Minh City, Vietnam. 6Department of Obstetrics & Gynaecology, School of Medicine, University of Zagreb, Zagreb, Croatia.

*Correspondence address. Centre for Reproductive Medicine, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel, Laarbeeklaan 101-1090 Brussels, Belgium. Tel: +322-477-66-60, Fax: +322-477-66-49; E-mail: samuel.santosribeiro@gmail.com

Submitted on February 29, 2016; resubmitted on May 24, 2016; accepted on July 13, 2016

STUDY QUESTION: Does the timing of the first frozen embryo transfer (FET) after gonadotropin-releasing hormone (GnRH) agonist triggering with the elective cryopreservation of all embryos affect pregnancy outcome?

SUMMARY ANSWER: FETs performed immediately after a freeze-all cycle did not vary significantly from delayed FETs in terms of pregnancy rates.

WHAT IS KNOWN ALREADY: As interest in, and use of, the freeze-all strategy expands in the field of reproductive medicine, the optimal timing to perform the subsequent FET has become increasingly important. Thus far, all clinical trials evaluating the efficacy of the segmentation strategy have opted to electively defer the first FET for at least one menstrual cycle. However, this more empirical approach may cause unnecessary distress to infertile patients eager to conceive as soon as possible.

STUDY DESIGN, SIZE AND DURATION: This retrospective cohort study included the first FET cycle of all women who underwent a freeze-all protocol between October 2010 and October 2015 in two reproductive medicine centres (in Belgium and Vietnam, respectively).

PARTICIPANTS/MATERIALS, SETTING AND METHODS: A total of 333 FET cycles were included in the analysis. Following the freeze-all cycle, the preparation of the endometrium consisted of the sequential administration of estradiol valerate and micronized vaginal progesterone. The start of the FET was classified as either immediate (following the GnRH agonist withdrawal bleeding) or delayed (by at least one menstrual cycle). Clinical pregnancy rate (CPR) was the main outcome of our study.

MAIN RESULTS AND THE ROLE OF CHANCE: Women in the immediate FET group were slightly younger on average (30.9 ± 4.1 versus 31.8 ± 4.3, P = 0.045) on the date of oocyte retrieval. Moreover, women in the immediate FET group received a blastocyst transfer more frequently (53.6% versus 41.6%, P = 0.038) and had fewer embryos transferred on average compared to the delayed FET group (1.8 ± 0.8 versus 2.0 ± 0.8, P = 0.013). CPR/FET was marginally significantly higher in the immediate FET group in our crude analysis (52.9% after immediate FET versus 41.6% after delayed FET, P = 0.046). In order to assess if CPR/FET remained unaltered after adjusting for measured confounding, we performed mixed-effects multivariable regression analysis. Timing of the FET no longer affected significantly the CPR of the first FET in the adjusted analysis (adjusted odds ratio (aOR): 0.62, 95% CI 0.38–1.00; predicted CPR of 52.9% for immediate FET versus 41.8% for delayed FET).
LIMITATIONS, REASONS FOR CAUTION: The results are limited by the retrospective design and the potential for unmeasured confounding. Furthermore, we only evaluated the effect of the FET timing of the first FET on CPRs in artificially supplemented cycles and, thus, the results should not be extrapolated to live birth rates or natural-cycle FETs.

WIDER IMPLICATIONS OF THE FINDINGS: This study offers a simple but potentially relevant measure to increase patient satisfaction and adherence in couples who seek to become pregnant both safely and as soon as possible.

STUDY FUNDING/COMPETING INTEREST(S): No funding was received for this study. The authors have no conflicts of interest to declare.

Key words: freeze-all strategy / segmentation concept / frozen embryo transfer / embryo transfer / ovarian hyperstimulation syndrome

Introduction

More than 30 years have passed since the first studies showing that a gonadotropin-releasing hormone (GnRH) agonist could be effectively used as an alternative to human chorionic gonadotropin (hCG) for final oocyte maturation during in vitro fertilization (IVF) were published (Bennick et al., 1990; Goren et al., 1990). Since then, this approach has been increasingly acknowledged as a worthy strategy to minimize the risk of ovarian hyperstimulation syndrome (OHSS) (Devroey et al., 2011) and is currently applied more broadly, namely in cycles with abnormal late follicular progesterone levels (Roque et al., 2015) or in oocyte donation programmes (Stoop et al., 2012).

However, while seeming equally efficient in terms of oocyte competence (Herrero et al., 2011), the generalized use of GnRH agonist triggering has remained thus far limited by the fact that this approach seems to significantly reduce IVF pregnancy outcomes (Griesinger et al., 2006). While enhanced luteal-phase support methods such as low-dose hCG (Humaidan et al., 2010) and aggressive hormonal supplementation (Engmann et al., 2008) have proven to somewhat mend this issue, these studies have been conflicting and came potentially at the cost of increasing the risk of OHSS (Iliodromiti et al., 2013; Seyhan et al., 2013; Youssel et al., 2014). This has led many authors to propose the elective cryopreservation of all embryos (i.e. the freeze-all protocol) as an alternative to attempting to rescue the luteal phase for a fresh embryo transfer (Devroey et al., 2011; Griesinger et al., 2011; Wong et al., 2014; Roque, 2015; Bloock et al., 2016).

The most recent meta-analysis collecting information from a few clinical trials designed to date to compare the pregnancy rates between fresh and electively frozen embryo transfers (FETs) to date demonstrated that the freeze-all approach did seem to perform better (Roque et al., 2013). However, in all three clinical trials included, the patients in the elective cryopreservation arm had to wait for a considerable amount of time to perform their first embryo transfer, since they either underwent pituitary down-regulation with a GnRH agonist first (Shapiro et al., 2011a,b) or elected to wait for two months (Aftanazian et al., 2010). With this waiting period, physicians were able to circumvent any potential complications that the abrupt luteolysis (which occurs after GnRH agonist triggering without luteal support) could have on a subsequent FET cycle. While this empirical decision to distance the subsequent FET from the artificially shortened luteal phase (Griesinger et al., 2007; Humaidan et al., 2010) seems to be thoughtful even without the clear scientific evidence to support it, it may cause unnecessary distress to infertile patients who are eager to conceive as soon as possible. For this reason, we decided to assess whether increasing the interval between the end of ovarian stimulation and the subsequent FET had any effect on the pregnancy outcomes following a freeze-all cycle.

Material and Methods

Study population and design

We performed a retrospective cohort study including the first FET cycle of all women who underwent their first freeze-all protocol between October 2010 and October 2015 at one of the two following hospitals: Dutch-speaking free Brussels University Hospital (Centre 1, Belgium) and VIMV, My Duc Hospital, Ho Chi Minh City (Centre 2, Vietnam). We included only artificially supplemented FETs. In order to minimize the potential for confounding, women who used donated oocytes or whose cycles included either in vitro maturation or preimplantation genetic diagnosis were excluded from the analysis. Finally, two delayed FET cycles which were preceded by a cancelled attempt of immediate FET due to a persistently thin endometrium were also excluded from the analysis to avoid misleadingly skewing the pregnancy outcome in the delayed FET group.

The freeze-all cycle

Ovarian stimulation was initiated on Days 2 and 3 of the menstrual cycle with either recombinant follicle-stimulating hormone (rFSH) or highly purified human menopausal gonadotropin (hMG). The starting dose was selected according to the woman's baseline characteristics, including age, body mass index (BMI) and antral follicle count (AFC). Cycles were monitored by means of serial vaginal ultrasound scans and determination of serum estradiol (E2) and progesterone (P) and, whenever necessary, dose adjustments of rFSH/hMG were performed according to ovarian response.

Pituitary down-regulation was performed with daily administration of either cetrotrex or goserelin starting from Day 5 to Day 7 of the menstrual cycle. As soon as three follicles with a mean diameter of ≥7 mm were observed, final oocyte maturation and oovaculture were triggered using 0.2 mg of recombinant, which was administered at least 8 h after the last dose of GnRH antagonist.

Cumulus-oocyte complexes were collected by transvaginal aspiration 36 h after triggering. The insemination of the collected oocytes was performed via either conventional IVF or intracytoplasmic sperm injection (ICSI). Fertilization was assessed approximately 18 h after insemination and, from then onwards, embryo development was graded daily until embryo cryopreservation. The embryo grading system varied according to the participating centres, both of which have been previously described in detail [by Papamichail et al. (2006) for Centre 1 and according to the Alpha Scientists Istanbul Consensus for Centre 2 (Alpha Scientists in Reproductive and Embryology, 2011)].
Results

Endometrial preparation for the FET

The artificial endometrial preparation consisted of a sequential administration of E2 valerate and micronized vaginal P as previously described for both Centre 1 (van de Vier et al., 2014) and Centre 2 (Lin et al., 2008). Summarily, 2 mg of E2 valerate was administered at least twice daily for 10–14 days, and the dose was later adjusted according to the endometrial thickness measured by ultrasonic scan. If the endometrial thickness was 7 mm or more, vaginal P supplementation was initiated. If the endometrial thickness was below 7 mm, patients continued to take oral E2 until the endometrium reached the necessary threshold, at which point P supplementation was started. In the delayed FETs, we registered all cases in which progesterone was administered during the preceding menstrual cycle.

The ‘timing’ of the first frozen embryo transfer

The ‘timing’ of the first frozen embryo transfer (TF-FET) was defined as the interval between oocyte retrieval and the start of the first FET cycle. We divided our sample into women who entered with an immediate start (following the GnRH antagonist withdrawal bleeding) and those with a delayed (by at least one menstrual cycle) start of FET cycle (Fig. 1).

Embryos were transferred under ultrasonic guidance using a soft embryo transfer catheter. The choice to transfer one or more embryos was decided by the clinician depending on the patient age, embryo quality, and cycle rank.

Main outcome measure and statistical analysis

Basic demographic characteristics were compared amongst the women who performed immediate and delayed FET, using the t-test (for continuous variables) or χ² test (for categorical variables). Clinical pregnancy rate (CPR), defined by the International Committee for Monitoring Assisted Reproductive Technology (ICMART) as the visualization of a gestational sac during transvaginal ultrasound at 7 weeks of gestational age (Zegers-Hochschild et al., 2009), was the main outcome of our study. CPR/FET were assessed both crudely and using mixed-effects multivariable regression analysis, accounting for the potential for correlated outcomes between cycles performed in the same centre. In this regression analysis, we also accounted for variables which were either unevenly distributed amongst the study groups or presumed to be potential confounders, namely female age (as a continuous variable), indication for freeze-all (high-risk of OHSS, low follicular P < 1.5 ng/ml, low follicular endometrium < 7 mm, and patient preference or other reasons), number of good quality embryos produced (as a continuous variable) and number transferred (single versus multiple), developmental stage (cleavage versus blastocyst stage) and quality of the embryos transferred (quality of the best embryo transferred, classified as either 1, 2 or 3).

Although our sample size would likely be limited to perform an adequate confounder-adjusted sensitivity analysis, for preliminary exploratory purposes, we crudely assessed whether the main outcome parameter presented obvious disparities according to potential confounders distributed unequally amongst the study groups, the centre where the procedure was performed and the indication for freeze-all, by evaluating the CPR/FET according to these subgroups.

A P-value was considered to be statistically significant whenever <.05. For the statistical analysis, we used Stata software version 13.1 (StataCorp, College Station, TX, USA).

Results

A total of 333 FET cycles were included in the analysis. The indications for IVF included male factor (n = 140, 42.0%), tubal factor (n = 51, 11.9%), and advanced maternal age (n = 17, 5.1%). The remaining patients (n = 295, 87.0%) underwent IVF for other indications.

Figure 1 Study groups according to TF-FET. The FET cycles were divided into either (A) immediate (following the GnRH agonist withdrawal bleeding) or (B) delayed (by at least one menstrual cycle). GnRH, gonadotropin-releasing hormone; FET, frozen embryo transfer; IVF, in vitro fertilization; TF-FET, timing of first frozen embryo transfer.
**Table I** Baseline characteristics of the freeze-all cycle according to TF-FET.

|                      | Immediate FET (n = 208) | Delayed FET (n = 125) | P-value  
|----------------------|-------------------------|-----------------------|-------
| Female age and exogenous ovarian stimulation (mean ± SD) | 30.9 ± 4.1 | 31.8 ± 4.2 | 0.045 |
| Female age (years)  | 180.5 ± 77.3 | 188.9 ± 84.4 | 0.434 |
| Total dose of exogenous FSH (IU) | 22.1 ± 10.6 | 22.2 ± 11.1 | 0.914 |
| Oocytes retrieved   | 22.1 ± 10.6 | 22.2 ± 11.1 | 0.914 |
| Good quality embryos produced | 6.7 ± 4.8 | 7.0 ± 4.2 | 0.551 |
| Reason for freeze-all (%) | 109 (52.4%) | 62 (49.6%) | 0.051 |
| High-risk of OHSS  | 39 (15.8%) | 14 (11.2%) | 0.30 |
| Late-follicular P > 1.5 mg/ml | 24 (1.5%) | 13 (10.4%) | 0.01 |
| Late-follicular endometrium < 7 mm | 36 (17.3%) | 36 (28.8%) | 0.05 |
| Patient preference and other reasons | 0 | 0 | 0.3 |
| Preformed effective freeze-off | 30 | 26 | 0.3 |
| Belief in Feng Shui | 3 | 2 | 0.3 |
| Poor personal commitments | 3 | 3 | 0.3 |
| Hydrocele or stage 1 intracavitary adhesions suspected during ovarian stimulation | 0 | 0 | 0.3 |

**Table II** Baseline characteristics of the FET cycle according to TF-FET.

|                      | Immediate FET (n = 208) | Delayed FET (n = 125) | P-value  
|----------------------|-------------------------|-----------------------|-------
| Embryo developmental stage at transfer (%) | 97 (46.3%) | 71 (58.4%) | 0.039 |
| Cleavage stage | 111 (53.4%) | 52 (41.6%) | 0.046 |
| Blastocyst stage | 64 (30.8%) | 30 (24.0%) | 0.046 |
| Embryo quality 1 | 134 (64.4%) | 89 (71.2%) | 0.046 |
| Embryo quality 2 | 64 (30.8%) | 30 (24.0%) | 0.046 |
| Embryo quality 3 | 10 (4.8%) | 6 (4.8%) | 0.046 |
| Frozen embryo transfer and pregnancy outcome per FET | 1.8 ± 0.8 | 2.0 ± 0.8 | 0.013 |
| Number of frozen embryos transferred (mean ± SD) | 123 (59.1%) | 59 (47.2%) | 0.034 |
| Clinical pregnancy rate (%) | 110 (52.9%) | 52 (41.6%) | 0.046 |

**Sensitivity analysis**

Figure 2 presents the CPR/FET according to the treatment centre and indication for freeze-all. Although the point estimates for CPR/FET following an immediate FET were frequently better than those after a delayed FET, these differences were only significant following a freeze-all protocol due to patient preference or other reasons (47.2% for immediate FET versus 22.2% for delayed FET, P = 0.026), a result that deserves a cautious interpretation due to the heterogeneity of this subgroup. Furthermore, the high-risk of OHSS subgroup, which represents half of the sample, the pregnancy rates seemed very similar (56.0% for immediate FET versus 56.0% for delayed FET, P = 0.94).

We also performed subgroup analyses to crudely assess the effect of potential confounding which varied significantly amongst the TF-FET groups. The point estimates once again argued any potential detrimental effect of immediate FET on CPR (Supplementary Table I).

**Discussion**

Although the freeze-all strategy itself has drawn much attention in recent literature (Blockeel et al., 2016), the optimal timing to perform the subsequent FET has been either overlooked or relayed as best postponed owing to putative concerns of an unpredictable outcome (Ozgur et al., 2015). Consequently, the current study provides...
relevant evidence suggesting that FETs done immediately after a freeze-all protocol appear to result in pregnancy rates comparable to FETs deferred to a later time.

Two previous studies have examined the effect of FET timing after a failed fresh ET on pregnancy rates. Specifically, a small study presented in 2008 proposed that immediate FETs resulted in higher pregnancy rates (Mass et al., 2008), a more recent publication including more than 1000 cycles clearly showed that FET timing had no significant effect on pregnancy outcomes in these patients (Santos-Ribeiro et al., 2016). However, the question of whether a delay of ET may benefit pregnancy outcome is more crucial after a freeze-all cycle, taking into consideration that the luteal phase is substantially reduced after GnRH agonist triggering (Beckers et al., 2003; Acedova et al., 2006). To this extent, our study seems to dispel the empirical notion that an artificially shortened luteal phase would cause immediate FETs to perform sub-optimally.

Given the fact that both our unadjusted and adjusted analyses of the effect of FET timing on CPR/FET show a borderline significance in favour of immediate FET, it would be tempting to encourage the widespread preference of immediate FET over a potentially worse delayed FET. And, indeed, some of our results support previous evidence that lead one to posit that differences in the luteal phase milieu of the preceding freeze-all cycle, namely in terms of P, may potentially be advantageous to women performing immediate FET. Specifically, having in mind that there is a sharp decrease in P levels immediately following GnRH agonist triggering (Beckers et al., 2003), it is interesting to associate such knowledge to previous studies pointing to a potential detrimental effect of both elevated endogenous and exogenous luteal-phase P levels in a subsequent treatment cycle, namely prior to ovulation induction (Diamond et al., 2012) and IVF (Greisinger et al., 2010). The authors of these studies postulated that an elevated pre-treatment P exposure may hinder endometrial receptivity either directly (by modifying the secretion of sex hormones or proteins related to endometrial receptivity, namely the endometrial androgen receptor, which may have an important role in defining the timing of the window of implantation (Mertens et al., 2000; Appart et al., 2002; Kawaski et al., 2006) or indirectly (by altering the hypothalamic-pituitary-ovarian axis secretion pattern). Hence, by applying a short-lived exogenous ovulation trigger (such as a GnRH agonist), one could be minimizing the detrimental effect of luteal-phase P on the subsequent cycle. Nonetheless, we would refrain from performing such a recommendation, due not only to the fact that our sensitivity analysis showed that there was significant heterogeneity among our sample, but also because we were unable to completely exclude selection bias in our study owing to its retrospective design. For that reason, instead of interpreting our P-values unequivocally, we opted rather to evaluate more exhaustively what our sample set could and could not potentially allow for in terms of extrapolation. To that extent, while limited by its sample size, the results of the sensitivity analysis also warrant more need for caution. Specifically, as this small exploratory analysis could not adequately assess potentially significant differences amongst the multiple subgroups, it left us questioning whether the apparent general benefit of immediate FET may not be a chance finding due to the conflation between many smaller subsets of freeze-all indications and the larger subgroup which included women at high-risk for OHSS. While we consider that the study robustly contradicts the conventional wisdom of a dire fate bestowed upon immediate FETs, we would avoid unwittingly depriving couples of their just freedom to choose a TF-FET that better fits their wishes until future studies on this matter may allow physicians to adequately distinguish which patients could indeed benefit from an immediate FET.

A recent retrospective analysis with 411 FET cycles sought to compare the pregnancy outcomes of FETs after a freeze-all protocol scheduled with either oral contraceptive pill (OCP) or GnRH agonist pretreatment (Ozgur et al., 2015). In this study, while women in the former group waited for 10 days to start the OCP and took the same for at least 21 days before commencing the FET (after a 3-day wash-out period), the patients in the GnRH agonist group began the endometrial preparation for the FET much sooner, specifically 17 days after the oocyte retrieval (10 days after initiating the daily GnRH agonist down-regulation). Although the latter strategy seemed to shorten in half the time necessary to start the FET, evidence supporting a benefit of such a strategy is lacking (Gluevsky et al., 2010) and, most importantly, almost all women in the immediate FET group of our study started their cycle even earlier and without the need for any pituitary down-regulation.

The main strengths of our study are derived from the fact that we included data from two centres and accounted for multiple confounding factors. However, our results are limited by the retrospective nature of
the study and the potential for unmeasured confounding. To that extent, we considered that a retrospective study was warranted before advancing to future prospective studies so that these could be performed without any ethical concerns of potentially delivering a treatment alternative to patients who had plausible reasons to perform worse. Finally, we would also caution that this study only evaluated the effect of TF-FET on CPR following artificially supplemented cycles and, thus, the results should not be extrapolated to other pregnancy outcomes (such as live birth) or natural-cycle FETs. That said, future studies should also assess whether known potential benefits (e.g., lower risk of prematurity and low birthweight) and risks (e.g., increased risk for macrosomia) of the freeze-all protocol (Maheshwari et al., 2012; Wennemo et al., 2013) may be altered by performing an immediate FET.

In conclusion, by showing that immediate FETs consistently do not perform worse than delayed FETs after a freeze-all protocol regardless of the indication for treatment, this study offers a simple but potentially clinically relevant measure to increase the patient satisfaction and adherence to modern-day ART in couples who seek to become pregnant both safely and as soon as possible.

Supplementary data
Supplementary data are available at http://humrep.oxfordjournals.org/.

Authors’ roles
S.S.R. and C.B. developed the study concept and designed the study. S.S.R., V.T.N.L., and J.S. acquired the data for analysis. S.S.R. performed the statistical analysis and wrote the first draft of manuscript. C.B., L.V.L., N.P., V.T.N.L., S.M. and H.T. contributed to the interpretation of the data and provided critical revision for important intellectual content. All authors approved the final version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.

Funding
No funding was received for this study.

Conflict of interest
The authors have no conflicts of interest to declare.

References
Results (Study 4)
Results (Study 4 – B)


SUMMARY OF THE MAIN RESULTS

Specific aim 1

Better understand the relationship between ART outcome and circulating P during COS

1. P produced during COS is associated with a hindered IVF on multiple levels, which includes both embryo quality and ER;
2. The detrimental effect of elevated P on ongoing pregnancy rates was corroborated using treatment outcomes which are more relevant to both patients and clinicians – specifically, LBR and CLBR;
3. The relationship between LFEP and ART pregnancy outcome seems to be non-linear and, contrary to previous beliefs, may be of varying importance according to the number of oocytes retrieved. These conclusions limit the ease of use of this biomarker in an everyday clinical setting;
4. The exposure to P during COS could affect LBR even in women who would be originally expected to have unaltered chances of conception if only late-follicular P levels were to be measured;

Specific aim 2

Predict and optimize in-cycle ER

5. EMT, as a prognostic tool for IVF, may still have the potential to predict both LBR and neonatal birthweight. This predictive capacity may be independent from the effect of the supraphysiologic endocrine milieu on ER, a conclusion that goes against the evidence thus far pointing to the limited value of ultrasonographic endometrial evaluation in predicting treatment outcomes beyond CP (Kasius, et al., 2014). Nonetheless, better tools to predict ER are still needed;
6. The clinical value of endometrial scratching and in-cycle transcriptomic expression analysis requires validation in an adequately-powered RCT;
Specific aim 3

Assess the benefits of deferring ETs away from COS

7. FETs are associated with lower MZT rates. However, whether this is due to an effect caused by the embryo cryopreservation method or by the transfer of an embryo in an unstimulated endometrium requires further investigation;
8. EP rates following ART have decreased overtime and may no longer be lower following FET;
9. The differences between the freeze-all strategy and fresh ET approach proposed by Humaidan, et al. (2010) requires further scrutiny in a RCT addressing specifically high-responder women;

Specific aim 4

Determine the temporal extent of the hindering effect of COS on ER

10. Intentionally deferring FETs beyond the menstrual cycle immediately following COS may be of no benefit for ART pregnancy outcomes, regardless if it is to be performed following a failed fresh ET attempt or a freeze-all protocol.
OVERALL DISCUSSION AND CONCLUSIONS

Despite the development of progressively more robust embryo morphology classification systems, up to two thirds of all morphologically good to top quality embryos eventually fail to implant (Fatemi and Popovic-Todorovic, 2013). The extent of frustration and burden that such failure causes to both couples and physicians highlights the need for a better understanding on the reasons behind this poor success rate of ART. Furthermore, it stresses the need for IVF to focus more attention on ER.

This thesis attempted to tackle the issue of endometrial non-receptivity by addressing two overarching hypotheses: a) which are the best tools a physician may use to predict and optimize ER during a fresh ET? (studied in specific aims 1 and 2) and b) what are the clinical implications of a pragmatic approach (i.e. the freeze-all strategy) towards hindered ER during a fresh ET? (evaluated in specific aims 3 and 4).

9.1 Clinical implications

Specific aim 1

Better understand the relationship between ART outcome and circulating P during COS

Our results shed new light on the evidence of the detrimental effect of P during COS on ART outcome. Mainly, they allowed us to get closer to answering the question that is increasingly concerning physicians regarding this matter: how can one (and should one) avoid the elevation of P and its detrimental effects during COS?

Most of the literature thus far has proposed three potential solutions for overcoming this issue: a) replace the type of exogenous stimulation with one having LH activity to stimulate the theca-cell conversion of P into androgens (Smitz, et al., 2007, Bosch, et al., 2010, Arce and Smitz, 2013), b) lower the total dose of gonadotropins (Bosch, et al., 2010, Venetis, et
Overall discussion and conclusions

188

al., 2016) or c) apply the freeze-all strategy (Bosch, et al., 2010, Roque, et al., 2015). However, with the collective knowledge derived from the 3 studies performed in specific aim 1, some of these “solutions” may soon cease to merit that title.

Firstly, there seems to be a very limited benefit of using drugs with LH activity to prevent LFEP, which appears to be a corroboration of the fundamental studies published just recently by Oktem, et al. (2017) pointing towards a FSH-derived production of P (Figure 7, p. 38). This, beyond potentially justifying our results, may also be one of the reasons why RCTs thus far have unanimously failed to provide any convincing evidence of benefit of one COS regimen over the other in terms of LBR (van Wely, et al., 2003, Andersen, et al., 2006, Devroey, et al., 2009).

Secondly, while we did confirm the findings of initial studies proposing that the decrease in the total dose of exogenous gonadotropins effectively reduces the production of P, this approach also reduces the number of oocytes retrieved. While one could make the argument that a small decrease in oocyte retrieval rates might have little direct effect on LBRs (Sunkara, et al., 2011), we are now confronted with data from concurrent research initiatives showing that oocyte retrieval rates affect CLBR independently (Drakopoulos, et al., 2016). So, while reducing the total dose of COS might mitigate the issues in the short term (i.e. LBR), it may very well affect long term ART outcome (i.e. CLBR). However, if one were to be still tempted to study this hypothesis, it is possible that a step-down approach would be a better alternative, given the fact that both P’s production and function seem to be concentrated on the late-follicular phase. This approach, originally proposed by Simon, et al. (1998) for elevated circulating E2 levels, may prove even more beneficial for circulating P. However, this hypothesis requires further confirmation.

Moreover, while the freeze-all approach is the most frequently preferred approach for LFEP, this thesis has shown that ER may not be the only factor causing the decrease in LBRs. Our results have subsequently been partially confirmed in multiple studies concurrent to this thesis (Bu, et al., 2014, Huang, et al., 2016, Healy, et al., 2017, Vanni, et al., 2017), in which both embryo quality and CLBR seemed to be affected by LFEP.

Finally, this thesis sheds light on the non-linear relationship between late-follicular P and LBR and on the limitations of evaluating P levels only once, i.e. the day of ovulation triggering.
Specific aim 2

Predict and optimize in-cycle ER

Much to our surprise, the results of this thesis give a refreshing place to the thought to be outdated measurement of EMT (Kasius, et al., 2014), namely in its capacity to predict ART outcome beyond CPR (e.g. for LBR and neonatal birthweight). However, while we do consider that EMT assessment may still have place in modern-day medicine for now, we consider it unable to predict LBR well enough to consider it a good marker of ER, underscoring the need for better in-cycle prediction tools for the future of ART.

Specific aim 3

Assess the benefits of deferring ETs away from COS

Four years later, the freeze-all strategy is still a very contentious topic, with many raising concerns towards the lack of available data to support it fully (Maheshwari and Bhattacharya, 2013, Ata and Seli, 2017).

A recent trial with over 1500 patients demonstrated that the freeze-all approach performs better than a fresh hCG-triggered ET (Chen, et al., 2016). While this result is interesting to observe, this study also adds more noise to the current field of evidence, given that it applied a progressively unpopular approach for high responders (i.e. the use of hCG triggering). For this reason, we await impatiently for the results of the ICE trial in order to evaluate how the two currently most-used approaches for predicted high responders (those who stand to gain the most from strategies curtailing excess hormone production) perform when pitted against each other.

While waiting for the results from the ICE trial, it was interesting to assess how EP rates no longer seems to be affected by COS, while MZT rates do. However, given the retrospective nature of these studies, we would recommend caution in establishing any clinical guidance until validated further.
Specific aim 4

Determine the temporal extent of the hindering effect of COS on ER

The confirmed lack of any carryover effect of COS plays an important role in clinical practice, as many centres intentionally defer FETs (Aflatoonian, et al., 2010, Shapiro, et al., 2011, Shapiro, et al., 2012, Shi, et al., 2014), leading to unnecessary patient distress. These results were recently confirmed in women performing freeze-all using LBR as the studied pregnancy outcome in a subsequent large retrospective study (Lattes, et al., 2017). While still requiring confirmation in prospective trials, the impact this information may have on the wellbeing of couples performing ART every day warrants at least a strong reconsideration on the counselling around the “policy” of purposely delaying FETs for reasons that no longer seem clinically sound.

9.2 Strengths and limitations

We consider that the main strength of this thesis was the fact that it attempted to address the question of ER from multiple fronts using either prospective or large retrospective studies to do so. With the ongoing ICE and REFRESH trials, we aimed simultaneously to pragmatically assess a) how good is the theoretical “best case scenario” one can offer to patients, by transferring their best embryo (the first embryo transferred) in a subsequently unstimulated endometrium, and b) if we could reverse or predict hindered ER following COS using endometrial scratching followed by in-cycle ER assessment. In the meantime, we also re-evaluated the modern-day value of the in-cycle ER prediction (late-follicular P and EMT evaluation) and “treatment” (intentionally deferring FETs) strategies most currently used in everyday clinical practice. Understanding how one could potentially predict ER during a fresh ET today, while we wait the years that are needed for better tools to be developed and distributed worldwide, was an outcome that could not be evaluated in the prospective trials that we commenced.

Another major strength of this thesis, namely in most of the retrospective studies, is the large sample size included in each analysis. The sheer number of cycles included in these studies allowed for the use of robust statistical analyses to account for multiple factors, namely potential confounding, the clustering of cycles performed by the same women and the eventual non-linearity between the associations being evaluated.
The effect of COS-derived excessive circulating P has been subject to much debate, with some authors advocating aggressive clinical action (Roque, et al., 2015) and others a complete halt in its measurement (Martinez, et al., 2016). One of the potential reasons for this controversy may be the fact that the majority of the previous literature failed to use (C)LBR as a main outcome. To this extent, we consider that the studies included in specific aim 1 provide data which is more relevant for patient counseling. Furthermore, by using CLBR in study 1-B, we could determine a potentially relevant relationship between circulating P and oocyte retrieval rates which had been missed before (Venetis, et al., 2015). Moreover, regarding specifically study 1-C, we consider that we have tackled the question of P exposure in a manner which is much more robust. While indeed a single evaluation of P (e.g. on the day of ovulation triggering, as used in the overwhelming majority of the previous literature) is much simpler to apply in clinic, this approach assumes several suppositions that required further scrutiny. Firstly, by performing a single evaluation, there is an assumption that P reaches its maximal level on the day of ovulation triggering. Secondly, by not measuring P more often, one also supposes that P exposure prior to this measurement has no effect on ER and embryo quality, an assumption that is in contradiction with the results of Huang, et al. (2012). Conversely, while the study performed by these researchers improved our understanding on the effect of circulating P exposure on IVF outcomes, it still lacked robustness as it failed to account for each individual’s basal P level and did not use LBR as the studied outcome. In study 1-C, we adequately accounted for these nuances, determining better which factors actually may be associated with P variation throughout COS and, ultimately, what was the association between cumulative P exposure and LBR.

Despite the before-mentioned strengths, this thesis also has a number of limitations which require caution and accounting for when attempting to extrapolate its results. First and foremost, although most retrospective studies included a large sample size and adjusted for many potentially confounding factors, these studies are limited by their inability to establish causation and by the possibility of unmeasured confounding (Grimes and Schulz, 2002). We considered that a retrospective design was the most appropriate design for a number of our research hypothesis because we either a) had ethical reservations in performing a prospective study offering a treatment modality which, at first glance, had enough biologic plausibility and previous clinical evidence to be an inferior alternative (specific aims 1 and 4), or b) considered that the outcome in study was too sporadic (studies 3-B and 3-C) or c) the difference being compared was too small (specific aim 4), to make a
feasible RCT. Moreover, some of the evaluated outcomes (such as the neonatal outcomes assessed in study 2-A) would be impossible to evaluate in a clinical trial, given that the subject unit (i.e. the liveborn) cannot be randomized at the start of the trial. In essence, while the ideal scenario would be to perform a RCT “always”, there are a number of situations in which such is impossible or unlikely to be completed, rendering a robust retrospective study the better option. In fact, retrospective studies are often used to establish equipoise (i.e. the balance between the potential benefit of an intervention and the uncertainty of its use), which is an ethical requirement when conceiving a future RCT (Braakhekke, et al., 2017). Retrospective studies are, in such situations, frequently used and have become the mainstay with regard to evaluating the effectiveness of healthcare (Black, 1996, Rochon, et al., 2005). Furthermore, adequately performed observational studies frequently find similar results to RCTs in a manner that is more cost-effective and comprehensive (Benson and Hartz, 2000, Concato, et al., 2000), given the possibility for multiple post-hoc sensitivity analyses.

As the use of retrospective studies is unlikely to ever stop in medical research, it is important that the readers are aware of the potential limitations of such analyses. Firstly, retrospective studies are more prone to selection, recall, (missing) information and confounding bias (Ioannidis, et al., 2001, Grimes and Schulz, 2002, Mann, 2003). Secondly, owing to the frequent need for data modelling via regression analysis, there is a risk of model over-fitting, which may overestimate the true effect of the variable being studied (Ioannidis, et al., 2001, Mann, 2003). Collectively, these concerns require that the readers take caution when extrapolating results into their daily clinical practice.

Finally, as other researchers have pointed out in studies published within the same timeframe of this thesis (de Ziegler, et al., 2017, Healy, et al., 2017), most of the currently used P assays outside of a research laboratory setting may not have the required accuracy to determine subtle P changes occurring during the follicular phase, given that they were developed to be used in the luteal phase. These less-precise P measurements may justify simultaneously the worse than hoped predictive value of late-follicular P and why multiple P assessments (such as the ones utilized in study 1-C) may potentially deliver more clinically-relevant predictions. Such a hypothesis should soon be confirmed in future studies using more sensitive assays.
9.3 Future prospects

Four years later, despite an intense amount of effort, this thesis leaves its main proponent with the frustration that fresh transfer ER is still widely unpredictable in current clinical practice. Notwithstanding, with two ongoing trials and a renewed batch of clinical hypotheses, our research group has planned multiple endeavours to build on the outcomes that resulted from this thesis.

Firstly, we – together with the general ART community (Evers, 2013, Simon and Bellver, 2014, van Wely, 2014, Blockeel, et al., 2016) – vehemently wait for definitive results on the use of the “real” clinical use of freeze-all protocol (Wei, et al., 2017) and endometrial scratching (Lensen, et al., 2016, van Hoogenhuijze, et al., 2017). Our belief is that the ICE and REFRESH trials will provide interesting results on both matters. Among the possible answers from these studies, we are particularly interested in one of our secondary outcomes of the REFRESH trial, which involves the comparative transcriptomic expression analysis of the late proliferative phase endometrial biopsies between women with good-quality embryos who became pregnancy versus those who did not. An increasing amount of recent evidence is pointing to the existence of a specific expression profile in women with recurrent implantation failure potentially associated with factors involved in the proliferative phase of endometrial development (Koot, et al., 2016, Macklon, 2017). Our ongoing research takes this information a step further, by attempting to evaluate whether an in-cycle expression test (that allows sufficient time for endometrial repair and same-cycle transfer) can be translationally useful for optimizing and tailoring fresh ETs. Even if the ICE trial shows that the freeze-all approach is superior to a fresh ET, this information is still relevant for patients, given that any delay in the timeframe between oocyte retrieval and the much-desired pregnancy test necessarily entails an extension of the most stressful period of ART (Gameiro, et al., 2012). Moreover, we are extremely interested in evaluating further the effect of P exposure on both ER and embryo quality. While we may lack the ethical equipoise to perform a RCT in which one would compare the pregnancy outcome between women who did and did not measure circulating P during COS (given the overwhelming evidence of the hindering effect of P), we are confident that there are many avenues still open for one to test prospectively the scientific hypotheses we generated retrospectively. Specifically, we consider that it would be essential to evaluate the effect of extending the late-follicular phase on both embryo quality and ER. One way to do so would be to perform a cross-over study in which oocyte
Donors would perform two stimulation cycles with different ovulation triggering points (e.g. when at least three follicles of 17 mm versus 20 mm are present). This would effectively prolong the period of COS associated with the largest extent of P production and allow for an efficient study design which could evaluate extensively both ER (by performing endometrial biopsies in both scenarios) and embryo quality (by comparing the utilization rate and, potentially, pregnancy outcomes of the oocyte recipients). This study may assist us to finally establish the “sweet-spot” between the highest possible number of oocytes retrieved at the lowest possible level of circulating P.

Another interesting study would be to concomitantly assess the predictive power of two different P measurement tools – such as an electrochemiluminescence immunoassay (as the one used throughout this thesis) versus liquid chromatography coupled to tandem mass spectrometry (LC-MS), the widely accepted gold standard (Taylor, et al., 2015) – on fresh ET pregnancy outcomes.

Finally, recent advances have allowed physicians to better evaluate the genital microbiota (Garcia-Velasco, et al., 2017, Moreno and Franasiak, 2017). This recent technological development adds new – and largely uncharted – territory which may allow doctors one day to accurately determine why so many apparently good quality embryos fail to progress beyond their transfer into the uterus. For this reason, our research group has developed a prospective observational trial that attempts to a) better characterize the female urogenital microbiota in infertility, b) determine whether the supraphysiologic sexual hormone milieu caused by IVF affects the female urogenital microbiota, c) evaluate if in-vitro embryo development is affected by the microbiota present in the embryo culture or in the uterus at the time of embryo transfer, and d) study whether specific microbiotic profiles can predict either IVF pregnancy or neonatal outcomes (NCT03105453). Our final aim with this project is to contribute to the development of a translational clinical test which could predict both a high-risk for female infertility and an unfavourable ART pregnancy outcome, opening potential solutions for the prevention of infertility and treatment of infertile women.
ACKNOWLEDGEMENTS

Along the way that led to this thesis I had the opportunity to be accompanied by a group of tremendous individuals who made this daunting endeavour a pleasure to perform.

Firstly, I would like to thank dearly the members of the jury for their effort to evaluate critically this thesis. Your insight and remarks have not only contributed to the improvement of this thesis, but also to the work that will follow.

To Prof Christophe Blockeel, the most humane of all people, who has his office door (and even house) always open and an emotional intelligence which is unparalleled. Thank you for always being there.

To Prof Calhaz Jorge, whom without this thesis would have never seen daylight. I express my deepest gratitude for your encouragement and support which was given always without expecting anything in return.

To Prof Herman Tournaye, who saw me leave, come back, leave for a while and come back. I appreciate your everlasting efforts to try to make things work and help everyone.

To the research collaborators and all co-authors, thank you for your hard work and for being so insightful and present. I am especially grateful to Nikolaos Polyzos, Neelke De Munck, Shari Mackens, Dominic Stoop, Ileana Mateizel, Vania Costa Ribeiro, Biljana Popovic, Hilde Van de Velde, Annalisa Racca and Panagiotis Drakopoulos, who were always so willing to help and give input. I hope I can one day help you all as much as you help me!

The CRG is an exceptional place to be! This uniqueness comes from an excellent team of doctors, embryologists, psychologists, nurses, midwives, secretaries, lab technicians and managers. I have never seen such an extraordinary place to work at and I believe I never will! Thank you to you all.

The CRG is not only a workplace, but it is also a family. During this time, I have had the privilege to be carefully taken care of by many of the members of the CRG’s extended
family, a gesture that was a kind surprise. I would like to especially thank Michel Camus, Martine Nelis, Valerie Schutyser, Steven Coulembier and, once again, Christophe Blockeel and Elisabeth De Waele. Your kind-heartedness made me feel at home.

Bruno, Ines, Aline and Pedro, you are the dearest of friends! Thank you!

I also would like to thank Flavia Carvalho for her artistic input during the production of this thesis. Even in the middle of chaos you found the time.

A PhD thesis does not stand by itself. Rather, it is the result of a journey with a fair amount of serendipity. There were some special people who were present during this long expedition that deserve my dearest thank you.

To Dr. Ana Paula Soares, for being the best of mentors and coaches.

To Catia Rodrigues, for being a super woman.

To Walter Meul and Yves Thorrez, for being quick and efficient as lightening!

To Romy Souffreau, for being just soooo Romy Souffreau.

To all my friends, who never forgot me despite the distance.

A special appreciation must go to my mother and father, who have lived their lives around making everything possible for their children. I hope one day to be half the man you both are so casually every day. Thank you.

Finally, Marta. With you, my life began (and I do not miss my last one for even a second!). Your calm and zeal are effortless, genuine and unshaken. You transform any hurdle into insignificance. Since I met you, I cannot stop smiling of joy. I love you.
REFERENCES


Abdalmageed OS, Keyhan S, Acharya KS, Acharya CR, Hurd WW, Muasher SJ. Conventional in vitro fertilization (IVF) versus intracytoplasmic sperm injection (ICSI) for couples with polycystic ovary syndrome (PCOS) and normozoospermic semen: an analysis of 4679 cycles from the SART registry. Fertility and sterility 2015;104: e300-e301.


Adamson GD, Tabangin M, Macaluso M, de Mouzon J. The number of babies born globally after treatment with the assisted reproductive technologies (ART). Fertility and sterility 2013;100: S42.


Alper MM, Fauser BC. Ovarian stimulation protocols for IVF: is more better than less? Reproductive biomedicine online 2017;34:345-353.


Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ (Clinical research ed) 1996;312: 1215-1218.


References


Bridges PJ, Komar CM, Fortune JE. Gonadotropin-induced expression of messenger ribonucleic acid for cyclooxygenase-2 and production of prostaglandins E and F2alpha in bovine preovulatory follicles are regulated by the progesterone receptor. Endocrinology 2006;147: 4713-4722.


Casper RF. It's time to pay attention to the endometrium. Fertility and sterility 2011;96: 519-521.


Check JH, Wilson C, Choe JK, Amui J, Basile D. Evidence that high serum progesterone (P) levels on day of human chorionic gonadotropin (hCG) injection have no adverse effect on the embryo itself as determined by pregnancy outcome following embryo transfer using donated eggs. *Clinical and experimental obstetrics & gynecology* 2010;37: 179-180.

Chen CH, Zhang X, Barnes R, Confino E, Milad M, Puscheck E, Kazer RR. Relationship between peak serum estradiol levels and treatment outcome in in vitro fertilization cycles after embryo transfer on day 3 or day 5. *Fertility and sterility* 2003;80: 75-79.


De Geyter C, De Geyter M, Huber PR, Nieschlag E, Holzgreve W. Progesterone serum levels during the follicular phase of the menstrual cycle originate from the crosstalk between the ovaries and the adrenal cortex. *Human reproduction* 2002;17: 933-939.


de Jong D, Macklon NS, Eijkemans MJ, Mannaerts BM, Coelingh Bennink HJ, Fauser BC. Dynamics of the development of multiple follicles during ovarian stimulation for in vitro fertilization using recombinant follicle-stimulating hormone (Puregon) and various doses of the gonadotropin-releasing hormone antagonist garelix (Orgalutran/Antagon). *Fertility and sterility* 2001;75: 688-693.


DeUgarte DA, DeUgarte CM, Sahakian V. Surrogate obesity negatively impacts pregnancy rates in third-party reproduction. *Fertility and sterility* 2010;93: 1008-1010.


References

202


Doldi N. Elevated serum progesterone on the day of HCG administration in IVF is associated with a higher pregnancy rate in polycystic ovary syndrome. Human reproduction 1999;14: 601-605.


Fang C, Huang R, Wei LN, Jia L. Frozen-thawed day 5 blastocyst transfer is associated with a lower risk of ectopic pregnancy than day 3 transfer and fresh transfer. Fertility and sterility 2015;103:655-661 e653.


Finn CA, Martin L. Endocrine control of the timing of endometrial sensitivity to a decidua l stimulus. Biology of reproduction 1972;7:82-86.


Fleming R, Jenkins J. The source and implications of progesterone rise during the follicular phase of assisted reproduction cycles. Reproductive biomedicine online 2010;21:446-449.


Friedler S, Zimerman A, Schachter M, Raziel A, Strassburger D, Ron El R. The midluteal decline in serum estradiol levels is drastic but not deleterious for implantation after in vitro fertilization and embryo transfer in patients with normal or high responses. *Fertility and sterility* 2005;83: 54-60.


Humaidan P, Polyzos NP, Alsbjerg B, Erb K, Mikkelsen AL, Elbaek HO, Papanikolaou EG, Andersen CY. GnRH trigger and individualized luteal phase hCG support according to ovarian response to stimulation: two


Imudia AN, Awonuga AO, Doyle JO, Kaimal AJ, Wright DL, Toth TL, Styer AK. Peak serum estradiol level during controlled ovarian hyperstimulation is associated with increased risk of small for gestational age and preeclampsia in singleton pregnancies after in vitro fertilization. Fertility and sterility 2012;97: 1374-1379.


JARG Editor-In-Chief. Retraction note to: Can fresh embryo transfers be replaced by cryopreserved-thawed embryo transfers in assisted reproductive cycles? A randomized controlled trial. Journal of assisted reproduction and genetics 2013;30: 1245.


Klinkert ER, Broekmans FJ, Looman CW, Habbema JD, te Velde ER. Expected poor responders on the basis of an antral follicle count do not benefit from a higher starting dose of gonadotrophins in IVF treatment: a randomized controlled trial. *Human reproduction* 2005;20: 611-615.


Lattes K, Checa MA, Vassena R, Brassesco M, Vernaeve V. There is no evidence that the time from egg retrieval to embryo transfer affects live birth rates in a freeze-all strategy. *Human reproduction* 2017;32: 368-374.


Letur-Konirsch H, Delanian S. Successful pregnancies after combined pentoxifylline-tocopherol treatment in women with premature ovarian failure who are resistant to hormone replacement therapy. *Fertility and sterility* 2003;79: 439-441.


Ma WG, Song H, Das SK, Paria BC, Dey SK. Estrogen is a critical determinant that specifies the duration of the window of uterine receptivity for implantation. Proceedings of the National Academy of Sciences of the United States of America 2003;100: 2963-2968.


Macklon N. Recurrent implantation failure is a pathology with a specific transcriptomic signature. Fertility and sterility 2017;108: 9-14.


Massin N. New stimulation regimens: endogenous and exogenous progesterone use to block the LH surge during ovarian stimulation for IVF. Human reproduction update 2017;23: 211-220.


Min JK. What is the most relevant standard of success in assisted reproduction? The singleton, term gestation, live birth rate per cycle initiated: the BESST endpoint for assisted reproduction. Human reproduction 2004;19: 3-7.


Moffitt DV, Queenan JT, Jr., Shaw R, Muasher SJ. Progesterone levels on the day of human chorionic gonadotropin do not predict pregnancy outcome from the transfer of fresh or cryopreserved embryos from the same cohort. Fertility and sterility 1997;67: 296-301.

Moffitt DV, Queenan JT, Jr., Shaw R, Muasher SJ. Progesterone levels on the day of human chorionic gonadotropin do not predict pregnancy outcome from the transfer of fresh or cryopreserved embryos from the same cohort. Fertility and sterility 1997;67: 296-301.


Ozcakir HT, Levi R, Tavmergen E, Goker EN. Premature luteinization defined as progesterone estradiol ratio ≥1 on hCG administration day seems to adversely affect clinical outcome in long gonadotropin-releasing hormone agonist cycles. *The journal of obstetrics and gynaecology* 2004;30: 100-104.

References

213


References


Santos-Ribeiro S, Polyzos NP, Haentjens P, Smits J, Camus M, Tournaye H, Blockeel C. Live birth rates after IVF are reduced by both low and high progesterone levels on the day of human chorionic gonadotrophin administration. *Human reproduction* 2014;29: 1698-1705.


Schieve LA, Reynolds MA. What is the most relevant standard of success in assisted reproduction?: challenges in measuring and reporting success rates for assisted reproductive technology treatments: what is optimal? *Human reproduction* 2004;19: 778-782.


References


Simon C, Belliver J. Scratching beneath The Scratching Case: systematic reviews and meta-analyses, the back door for evidence-based medicine. *Human reproduction* 2014;29: 1618-1621.


Stanger JD, Yovich JL. Reduced in-vitro fertilization of human oocytes from patients with raised basal luteinizing hormone levels during the follicular phase. Br J Obstet Gynaecol 1985;92:385-393.


Tarlatzis BC, Qublan HS, Sanopoulou T, Zepiridis L, Grimbisiz G, Bontis J. Increase in the monozygotic twinning rate after intracytoplasmic sperm injection and blastocyst stage embryo transfer. Fertility and sterility 2002;77:196-198.


References


Valbuena D, Martin J, de Pablo JL, Remohi J, Pellicer A, Simon C. Increasing levels of estradiol are deleterious to embryonic implantation because they directly affect the embryo. Fertility and sterility 2001;76: 962-968.


van der Westerlaken L, Helmerhorst F, Dieben S, Naaktgeboren N. Intracytoplasmic sperm injection as a treatment for unexplained total fertilization failure or low fertilization after conventional in vitro fertilization. Fertility and sterility 2005;83: 612-617.

van Hoogenhuijze NE, Torrance HL, Mol F, Laven JSE, Scheenjes E, Traas MAF, Janssen C, Cohlen B, Teklenburg G, de Bruin JP et al. Endometrial scratching in women with implantation failure after a first IVF/ICSI cycle; does it lead to a higher live birth rate? The SCRaTCH study: a randomized controlled trial (NTR 5342). BMC women's health 2017;17: 47.


Wu CH, Kuo TC, Wu HH, Yeh GP, Tsai HD. High serum estradiol levels are not detrimental to in vitro fertilization outcome. *Taiwan J Obstet Gynecol* 2007;46: 54-59.


Younis JS. Elevated P level on the day of hCG administration is related to FSH dose: is it the whole truth? *Human reproduction* 2011;26: 498-499.

References

Younis JS, Matilsky M, Radin O, Ben-Ami M. Increased progesterone/estradiol ratio in the late follicular phase could be related to low ovarian reserve in in vitro fertilization-embryo transfer cycles with a long gonadotropin-releasing hormone agonist. *Fertility and sterility* 2001;76: 294-299.


Zavy MT, Craig LB, Wild RA, Kahn SN, O'Leary D, Hansen KR. In high responding patients undergoing an initial IVF cycle, elevated estradiol on the day of hCG has no effect on live birth rate. *Reproductive biology and endocrinology : RB&E* 2014;12: 119.


