Ovarian Reserve and Assisted Reproduction

THOMAS BRODIN
Abstract


Treatment success in IVF-ICSI is mainly limited by female age, but differences in ovarian reserve (OR; the remaining pool of oocytes and their quality) between individuals modify treatment prerequisites among women of similar age. OR may be assessed by OR tests (ORTs). The main aims of this work were to study menstrual cycle length (MCL), basal levels of circulating gonadotrophins, antral follicle count (AFC) and serum Anti-Müllerian hormone (AMH) levels and their associations with and prognostic capacities regarding IVF-ICSI outcome in large cohorts of unselected women.

Age-adjusted MCL was positively and linearly associated with pregnancy rates (PRs), live-birth rates (LBRs) and ovarian response to controlled ovarian hyperstimulation. An MCL of >34 days almost doubled the LBR compared with an MCL of <26 days.

The grouped variable ‘combined FSH and LH levels’ was superior to both individual gonadotrophin levels and the LH:FSH ratio. The highest mean PR was seen in connection with a combination of FSH <6.7 U/l with LH >4.9 U/l; PRs were lowest when FSH-LH levels were opposite to this (high-low) and intermediate when FSH-LH levels were low-low or high-high. Associations with LBR and ovarian response were similar as those for PR.

AFCs and serum AMH levels were positively and log-linearly associated with PR, LBR and ovarian response. Success rates levelled out above AFC 30 or AMH 5 ng/ml. Treatment outcome was superior among women with polycystic ovaries.

Among the studied ORTs, logAFC and logAMH concentration correlated most strongly. After multivariate testing, entering all studied ORTs, AMH and female age remained independently associated with LBR. AMH+AFC+age predicted both poor and excessive ovarian responses with high accuracy.

In conclusion, measures of OR are strongly associated with PR, LBR and ovarian response in a log-linear fashion, and partly reflect oocyte quality. The OR spectrum is continuous, from small ‘oligofollicular’ ovaries (the low extreme) to polycystic ovaries (the high extreme). Among the studied ORTs, AMH together with age provide the most powerful basal estimate for IVF/ICSI outcome.

Keywords: AFC, AMH, anti-Müllerian hormone, antral follicle count, follicle-stimulating hormone, FSH, ICSI, infertility, intracytoplasmic sperm injection, in vitro fertilization, IVF, LH, live birth, luteinizing hormone, menstrual cycle, menstrual cycle length, ovarian reserve, pregnancy, reproductive endocrinology, reproductive technology

Thomas Brodin, Uppsala University, Department of Women’s and Children’s Health, Obstetrics and Gynaecology, Akademiska sjukhuset, SE-751 85 Uppsala, Sweden.

© Thomas Brodin 2013

ISSN 1651-6206
ISBN 978-91-554-8592-4
urn:nbn:se:uu:diva-192998 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-192998)
Marie
Johanna
Max

...memento vivere
This work is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the respective publishers.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>11</td>
</tr>
<tr>
<td>IVF</td>
<td>11</td>
</tr>
<tr>
<td>The ovaries</td>
<td>12</td>
</tr>
<tr>
<td>Oocyte development, follicular maturation and regulation of the menstrual cycle</td>
<td>12</td>
</tr>
<tr>
<td>The uterus and the endometrium</td>
<td>15</td>
</tr>
<tr>
<td>Ovarian reserve</td>
<td>16</td>
</tr>
<tr>
<td>Means of assessing ovarian reserve</td>
<td>16</td>
</tr>
<tr>
<td>Ovarian reserve at the time of assisted reproduction</td>
<td>19</td>
</tr>
<tr>
<td>Controlled ovarian hyperstimulation</td>
<td>21</td>
</tr>
<tr>
<td>Ovarian reserve tests and IVF-ICSI treatment outcome</td>
<td>22</td>
</tr>
<tr>
<td>Aims</td>
<td>24</td>
</tr>
<tr>
<td>Material and methods</td>
<td>25</td>
</tr>
<tr>
<td>Swedish legislation regulating ART</td>
<td>25</td>
</tr>
<tr>
<td>Effects of Swedish legislation and use of a prediction model</td>
<td>25</td>
</tr>
<tr>
<td>Ethical considerations</td>
<td>26</td>
</tr>
<tr>
<td>Patients and treatments</td>
<td>26</td>
</tr>
<tr>
<td>Design and methods</td>
<td>27</td>
</tr>
<tr>
<td>Ovarian sensitivity index</td>
<td>27</td>
</tr>
<tr>
<td>Paper I</td>
<td>28</td>
</tr>
<tr>
<td>Paper II</td>
<td>28</td>
</tr>
<tr>
<td>Paper III</td>
<td>29</td>
</tr>
<tr>
<td>Paper IV</td>
<td>30</td>
</tr>
<tr>
<td>Paper V</td>
<td>31</td>
</tr>
<tr>
<td>Statistics</td>
<td>32</td>
</tr>
<tr>
<td>Results</td>
<td>35</td>
</tr>
<tr>
<td>Paper I</td>
<td>35</td>
</tr>
<tr>
<td>Paper II</td>
<td>39</td>
</tr>
<tr>
<td>Paper III</td>
<td>42</td>
</tr>
<tr>
<td>Paper IV</td>
<td>45</td>
</tr>
<tr>
<td>Paper V</td>
<td>48</td>
</tr>
<tr>
<td>Oocyte quality?</td>
<td>51</td>
</tr>
</tbody>
</table>
Discussion .................................................................................................................. 52
Methodological considerations and comments ......................................................... 52
Ovarian reserve tests and treatment outcome ............................................................ 55
  Do ovarian reserve tests predict oocyte quality? .................................................. 60
  Do ovarian reserve tests predict pregnancy and live birth? ................................. 61
Polycystic and oligofollicular ovaries ..................................................................... 63
Limitations .............................................................................................................. 65
General conclusions ............................................................................................... 69
Sammanfattning på svenska .................................................................................... 71
Acknowledgements ............................................................................................... 75
References ............................................................................................................. 78
Abbreviations

AFC(s)  Antral follicle count(s)
AMH  Anti-Müllerian hormone
ART  Assisted reproduction technology
AUROC  Area under the receiver operating characteristic curve
CI; [CI]  95% confidence interval
COH  Controlled ovarian hyperstimulation
DET  Double embryo transfer
eSET  Elective single embryo transfer
ET  Embryo transfer
FSH  Follicle stimulating hormone
GEE  Generalized estimating equation
GnRH  Gonadotrophin releasing hormone
hCG  human chorionic gonadotrophin
hMG  human menopausal gonadotrophin
ICSI  Intracytoplasmic sperm injection
IMC  Integrated morphology cleavage embryo score
IVF  In vitro fertilization
LBR  Live birth rate
LH  Luteinising hormone
MCL  Menstrual cycle length
OHSS  Ovarian hyperstimulation syndrome
OPU  Ovum pick-up
OR  Ovarian reserve
ORT(s)  Ovarian reserve test(s)
OSI  Ovarian sensitivity index
PCO  Polycystic ovaries
PCOS  Polycystic ovary syndrome
PR  Pregnancy rate
rFSH  recombinant FSH
SD  Standard deviation
SEM  Standard error of the mean
SET  Single embryo transfer
2D  Two dimensional
Introduction

Worldwide, since the first successful IVF treatment carried out by Edwards and Steptoe in 1977 – 1978 (1), more than four million infants have been born after in vitro fertilization (IVF). IVF has become established as a safe and efficient treatment for infertility and subfertility, affecting about 10% of all couples trying to conceive.

IVF

The principle of IVF is the retrieval of mature oocytes from a woman’s ovaries, fertilization with sperm, cultivation in the laboratory and transfer of the embryo to the uterine cavity.

The first successful treatments in humans were carried out in natural cycles, i.e. relying on a woman’s natural ovulation. The chance of a resulting pregnancy with that approach is small. After the development of methods of controlled ovarian hyperstimulation (COH, see below), treatment cycles are at present generally stimulated for retrieval of multiple eggs and thus increased chances of successful treatment due to the possibility of selecting a high quality embryo(s) for transfer.

Ovum pick-up (OPU) is carried out with an ultrasonographically guided needle piercing the vaginal wall to aspirate the follicular fluid together with ova from ovarian follicles. After fertilization with sperm, the embryo is cultivated in an incubator for two (cleavage stage) to five days (blastocyst stage). Male infertility may be overcome by intracytoplasmic sperm injection, ICSI.

The best embryos available are selected for embryo transfer (ET), which is carried out with a thin plastic catheter inserted through the cervix uteri to the uterine cavity. Embryo quality is predominantly assessed on morphological grounds. There are different embryo-scoring systems. All embryos included in the studies on which this thesis is based were scored according to an embryo-scoring system developed at the Carl von Linné Clinic, Uppsala, Sweden – the integrated morphology cleavage (IMC) embryo score (2). The IMC score is the first evidence-based embryo-scoring system for embryos transferred on day two after OPU.
After ET, luteal phase support is given with progesterone, usually for two weeks as a vaginal preparation. A surplus of good-quality embryos may be cryopreserved for later transfer.

The ovaries

Oocyte development, follicular maturation and regulation of the menstrual cycle

If not otherwise indicated, this section refers to: Fritz, M.A. and Speroff L., Clinical Gynecologic Endocrinology and Infertility. 8th ed. 2011, Philadelphia, Lippincott Williams & Wilkins. Chapters 3, 4 and 6.

Oocytes, follicles and follicle recruitment

Primordial germ cells migrate from the yolk sac to the gonadal ridge and the forming ovaries of the female foetus. As proliferation of the germ cells continues, they give rise to oogonia, reaching a maximum number of six to seven million oogonia in gestational weeks 16–20 when formation of the primordial follicles begins. The oogonia develop to primary oocytes as they enter meiosis I. At birth, meiosis of the oocytes is arrested in the diplotene stage of meiosis I (primary oocytes with 46 chromosomes).

The oocytes are enclosed in primordial follicles surrounded by stroma in the inner part of the ovarian cortex. A primordial follicle consists of an oocyte surrounded by a layer of pre-granulosa cells.

Oocyte (follicle) recruitment for maturation is activated during puberty. The transition of primordial follicles first to primary follicles and thereafter to pre-antral (secondary) follicles is a complex process that is not fully understood. Gap junctions develop between the oocyte and the surrounding granulosa cells and are necessary for nutritional and signal interchange. Several peptide growth factors are involved, exerting paracrine and autocrine signalling. Many of the stimulating growth factors belong to the ‘transforming growth factor-β (TGF-β) superfamily’, such as Bone Morphogenic Proteins (BMPs), growth and differentiation factors (GDFs) and activin. Other factors, such as kit ligand (KL), leukaemia inhibitory factor (LIF) and neurotrophins, also promote follicle maturation and oocyte growth at this early stage (3). Anti-Müllerian Hormone (AMH), also a member of the TGF-β superfamily, is secreted from granulosa cells of already growing pre-antral and small antral follicles (4). AMH inhibits the transition of primordial follicles to primary and pre-antral follicles together with inhibins and it may also be involved in dominant follicle development by reducing the FSH responsiveness of growing antral follicles (3, 5). These first steps of follicle recruitment and maturation, i.e. from primordial follicle to the pre-antral
stage, last for at least 70 days and are not gonadotrophin-dependent (Figure 1).

Figure 1. Schematic illustration of follicular development. Anti-Müllerian Hormone (AMH) is produced in growing follicles up to the small antral stage. Within the ovary, AMH inhibits recruitment from the primordial follicle pool and reduces the FSH responsiveness of growing antral follicles, thereby contributing to dominant follicle development.

At the pre-antral stage, the enlarging oocyte is surrounded by the zona pellucida. From here on, follicle maturation is under endocrine control and is dependent upon gonadotrophins. The theca layer forms and the granulosa cells proliferate and are now capable of synthesising oestrogens under stimulation by follicle-stimulating hormone, FSH, secreted from the anterior pituitary. The secretion of FSH is in turn regulated by pulsatile secretion of gonadotrophin-releasing hormone, GnRH, from the hypothalamus together with feedback mechanisms from the target organs, as described below. FSH and oestrogens act synergistically, enhancing granulosa cell proliferation and they also increase the amount of FSH receptors in the follicle. As follicle growth continues, oestrogen and FSH stimulate the production of follicular fluid, forming an antrum, and granulosa cells near the oocyte form the cumulus oophorus that surrounds the oocyte.

The dominant follicle
To develop into a dominant follicle, conversion from an androgen-dominated microenvironment to an oestrogen-dominated one is crucial; small follicles are poor in aromatising androgens to oestrogens and are therefore more prone to atresia. In larger follicles and under the influence of luteinising hormone, LH (from the anterior pituitary), the theca cells synthesise androgens, which are in turn aromatised to oestrogens in the proliferat-
ing granulosa cells under the influence of FSH. The dominant follicle becomes more sensitive to FSH as a result of increases in both the amount of granulosa cells and FSH receptor density. Together with inhibins, oestrogens exert negative feedback on the secretion of FSH in the pituitary, which reduces the stimulation of less mature follicles even more, leaving them androgen-dominated and destined for atresia.

Ovulation
Together with a rise in estradiol, FSH induces the expression of LH receptors in the granulosa cells. When the dominant follicle is 15–20 mm in diameter, the rising estradiol levels increasingly stimulate pituitary release of LH, leading to the ovulatory LH surge. The rise in LH also starts to luteinise the granulosa cells, leading to an increase in progesterone synthesis. Together, LH, progesterone and FSH stimulate proteolytic enzymes, further enhanced by prostaglandins, causing rupture of the follicle, and ovulation. Ovulation is only possible if the follicle is mature; an adequate level of estradiol from the follicle about to ovulate coordinates the LH surge.

Just before ovulation and at fertilization, the oocytes undergo their final meiotic divisions, expelling the first (secondary oocyte) and second polar body, resulting in a mature haploid oocyte with 23 chromosomes.

The corpus luteum
The ruptured follicle transforms to a corpus luteum as LH luteinises the granulosa cells. Also, theca lutein cells become part of the corpus luteum. The luteal phase lasts for 14 (±3) days. Unless rescued by human chorionic gonadotrophin (hCG) in the case of pregnancy, the corpus luteum declines 9–11 days after ovulation. The mechanisms of corpus luteum demise, luteolysis, are not fully understood.

The endometrium and regulation of the menstrual cycle
The endometrium (uterine lining) proliferates during the follicular (proliferative) phase under the stimulation of oestrogen. After ovulation, under the influence of progesterone and oestrogen, proliferation stops and secretory glands develop. Through the action of oestrogens, progesterone, growth factors and prostaglandins, the stroma becomes increasingly oedematous, which is the main feature of the endometrium at the time of implantation (cycle days 21–22). In the absence of pregnancy, oestrogen and progesterone levels fall as the corpus luteum regresses, causing the endometrium to be shed in menstrual bleeding.

Changes in the menstrual cycle with age
Regulation of the menstrual cycle is closely related to the regulation of follicle maturation and ovulation as described above.
Generally, cycles become ovulatory and regular during the years after menarche and exhibit the most frequent interval of 28 days \( (6, 7) \). A subtle gradual shortening of menstrual cycles occurs with increasing age. As the follicle pool deteriorates (see ‘Ovarian reserve’ below), levels of inhibins decrease and FSH levels increase \( (8-10) \) and FSH is enabled to rise earlier in the luteal phase. The result is an ‘overlap’ of cycles; the coming follicular phase starts earlier during the ending luteal phase, with a shortening of the interval between menstrual bleedings as a result \( (11) \). In the late 40s and before menopause, cycles become irregular due to anovulatory cycles as an effect of exhaustion of the follicle pool \( (10) \).

The uterus and the endometrium

After fertilization, it takes 5–8 days until the embryo implants. For implantation to occur, the endometrium and the embryo have to be synchronised. The endometrium is embryo-receptive for a limited period of time – the ‘implantation window’ \( (12, 13) \). Oestrogens, and particularly progesterone, together with growth factors, adhesion molecules, cytokines and prostaglandins regulate endometrial receptivity \( (13) \).

In assisted reproductive technology (ART) it is common that the woman is in her (late) 30s; thus at an age when a reduction in fertility with age plays a significant role. However, the endometrium is not considered to limit the chance of a healthy pregnancy in older women \( (14) \). The so-called ‘ovarian concept’ of reduced fertility with increasing (ovarian) age suggests that oocyte quality decreases as the amount of oocytes wanes (see ‘Ovarian reserve’ below). The ovarian concept is supported by good results in oocyte donor cycles where the donor is young and the recipient is a woman of greater age \( (15) \).

It is possible, however, that failed IVF treatments can also be attributed to uterine factors and/or ‘poor endometrial receptivity’. An intrauterine pathological condition such as polyps and adhesions may affect the chance of successful implantation and uterine anomalies and fibroids have also been reported to have negative influences \( (16-18) \). Moreover, COH may alter the development of the endometrium; in a recent report it was pointed out that the endometrium also plays an important role, along with embryo quality, in successful implantation \( (19) \). Although only a very abnormal endometrium seems to affect implantation \( (20) \), there are indications of impaired endometrial receptivity as a result of high estradiol levels during COH, with premature induction of progesterone receptors, leading to an advanced endometrial stage \( (19) \).

This work is focused on ovarian reserve and the ‘ovarian concept’ of fertility.
Ovarian reserve

The size of the remaining pool of oocytes and their quality at any given time is often referred to as the ovarian reserve (15). Ovarian reserve is probably largely determined by genetic factors, causing inter-individual differences in follicle density and the number of oocytes at birth. This is possibly caused by different endowment and mitotic rates of germ cells in the foetal ovary (21), and possibly also by different rates of follicle depletion (15).

The amount of female germ cells reaches a maximum at about half way through foetal life (week 16–20 of gestation), after which there is a fall in the number of immature follicles. At birth, there is about a quarter to half a million follicles containing one oocyte each in each ovary. The reduction in the amount of primordial follicles continues and at the time of menarche, about half of them remain. Only a small number of the oocytes left at menarche reach maturation and are ovulated over approximately 400 menstrual cycles; the majority (>99%) of follicular oocytes are destined for continuous atresia and apoptosis (22, 23).

The rate of follicle depletion declines increasingly, either ‘bi-exponentially’ with an increase in the depletion rate in the late 30’s (24), or gradually with increasing age (25). In parallel, there is a decrease in oocyte quality due to increasing aneuploidy (14, 15), leading to a lower probability of conception and an increased risk of spontaneous abortion (15). That is, as oocyte quality deteriorates with advancing age, oocytes with the ability to give rise to a euploid embryo after fertilization become increasingly scarce. The point of time at which there is more definite infertility, as a result of deteriorating oocyte quality and quantity, precedes menopause by about ten years (15).

When the ovarian reserve is depleted, menopause occurs. Age at natural menopause is normally distributed (15, 26). The close relationship between depletion of the ovarian reserve and menopause makes it highly probable that the distribution of ovarian reserve is also Gaussian.

Means of assessing ovarian reserve

For obvious reasons, counting all the follicles in the ovaries is not possible. Ovarian reserve assessment is carried out by using different proxy variables and several ovarian reserve tests (ORTs) have been described. The most commonly used include assays of circulating levels of inhibin-B, gonadotrophins and anti-Müllerian Hormone (AMH), and sonographic estimates of ovarian volume and the number of antral follicles – the antral follicle count (AFC) (27).

Some ORTs, like the ‘Clomiphene Citrate Challenge Test’ (CCCT) (28) and other forms of ‘stress test’ like the ‘Gonadotrophin-releasing hormone
agonist stimulation test’ (GAST) (29) or the ‘Exogenous FSH Ovarian Reserve Test’ (EFORT) (30) have generally now been abandoned.

The methods for assessing ovarian reserve used in the present work are as follows:

**Menstrual cycle length, MCL**
From the age of 20 to 40, as the follicle pool diminishes, menstrual cycle intervals shorten, by a mean of two days (6, 31). Previous studies have shown an association between menstrual cycle length and fecundity (32) and also studies from the 1980s indicated inferior ART outcome among women with menstrual intervals shorter than 27 days (33, 34). A shortened menstrual cycle interval may thus indicate advanced ovarian aging. Menstrual cycle length has previously not been described as a marker of ovarian reserve.

**Follicle-stimulating hormone, FSH**
FSH is secreted from the anterior pituitary gland and is the main regulator of follicle growth. The level of FSH in serum varies throughout the menstrual cycle and is subject to modulation by feedback mechanisms on the pituitary and the hypothalamus by oestrogen together with regulatory peptides (inhibin A and B, activin and follistatin) from the ovaries. Low oestrogen levels and low levels of inhibins cause an increase in the secretion of FSH, while high levels of oestrogen and inhibin do the opposite (35).

Basal (measured on menstrual cycle day 3±1) FSH (36, 37) serves as an indirect measure of follicle cohort size through the feedback mechanisms on the pituitary and the hypothalamus. A low basal FSH level would thus indicate a good ovarian reserve and vice versa. Accordingly, the reliability of basal FSH is dependent on a functioning hypothalamus–pituitary–ovary feedback mechanism. Basal FSH levels may fluctuate for several reasons (38) and they have also been shown to vary within the individual and from cycle to cycle (39), which is why assay of basal FSH alone is an unstable and not entirely reliable ORT.

The interpretation of serum FSH concentrations may also be confounded by receptor polymorphism, which may affect endogenous FSH levels (40) and the response to COH (41).

**Luteinising hormone, LH**
Assay of LH alone is usually not referred to as an ORT. The results of some studies suggest that low basal levels of LH may influence ART treatment outcome negatively, although LH mainly has been evaluated as part of the FSH:LH (or LH:FSH) ratio (42-47). An elevated LH level or a high LH:FSH ratio is also a common feature in women with polycystic ovary syndrome (PCOS) (48), a group with generally well preserved ovarian re-
serves at a relatively high age (49), also suggesting that LH levels give information on ovarian reserve.

**Antral follicle count, AFC**

Only sufficiently large maturing follicles (containing fluid, in the antrum) can be visualized by vaginal ultrasonography and usually follicles of 2–10 mm in size are counted (50). It has long been demonstrated that the amount of small antral follicles predicts oocyte yield and the cancellation rate in IVF-ICSI (51, 52). The AFC proportionally reflects the underlying number of primordial follicles (53) and therefore serves as a measure of ovarian reserve.

![Figure 2](image_url). Ovaries observed in vaginal ultrasonography. This example illustrates the difference in ovarian size and in the numbers of antral follicles (arrows) in a large polycystic ovary (left) and a normal ovary (right).

**Anti-Müllerian Hormone, AMH**

AMH is a dimeric glycoprotein belonging to the transforming growth factor-β (TGF-β) family. AMH is synthesised in growing pre-antral and small antral follicles (4) (Figure 1) and is considered to reflect the follicle pool in a stable manner in adult women. Serum AMH levels increase during puberty, peak at 24–25 years of age (54) and then decline with age, being undetectable before menopause (55-57). AMH levels are generally analysed by using enzyme-linked immunosorbent assays (ELISAs).

At the beginning of the 2000s, AMH levels were shown to be positively associated with the response to COH (58) and assay of AMH has since been proposed as an ORT with several advantages: it is (so far) the marker with the least inter-cycle and intra-cycle variability (59-61) and can thus be measured regardless of cycle day. The level of AMH is unaffected by the use of combined hormonal contraceptives (62), but this has also been contradicted (63, 64).

Assay of AMH is the ORT in which results change earliest with (ovarian) age (65) and it is also of help in differentiation between causes of anovulation and amenorrhoea (66-68).
Ovarian reserve at the time of assisted reproduction

Female age is the main single determinant of the chance of delivering a healthy child, in natural procreation as well as in assisted reproduction.

The decreasing success rate of IVF with increasing female age (Figure 3) is primarily due to a gradual loss of oocyte quality (27). There is evidence that causes of infertility, or medical indications for treatment, such as tubal occlusion, endometriosis or male factor infertility, have minor or no impact on treatment outcome (69) and that the main limitations of successful IVF treatment are advanced female age together with the ovaries’ ability to respond to ovarian hyperstimulation and produce euploid oocytes of high quality (22, 69). The chance of success of IVF may thus, at least partly, be reflected by the response to ovarian stimulation; a poor response to ovarian stimulation is associated with a reduced ovarian reserve and poor IVF results (22) and is also predictive of early menopause (70, 71), while the chance of a pregnancy and a live birth is higher among women who respond better to COH.

Although female age alone influences IVF treatment outcome and ovarian reserve co-varies with age, differences in follicle number (at birth) cause differences in fertility potential (i.e. ovarian reserve) between individuals of similar ages. This is coarsely illustrated in Figure 4. For example, previous findings have shown that young women with reduced ovarian reserves and poor responses to COH have poor chances of treatment success (72), while older women with an ORT result indicating a well preserved ovarian reserve have fair chances of a pregnancy (73). However, it is still unclear as to what extent differences in ovarian reserve at different ages matter compared with the influence of age itself. Moreover, there is a possibility that the decrease in ovarian reserve may proceed at a different pace than aging itself (15) and ovarian reserve may also be (negatively) influenced by adnexal surgery, chemotherapy, pelvic radiotherapy and lifestyle factors such as smoking (15, 74, 75).
Figure 3. Births per OPU in different age cohorts in Sweden in 1994–2010. (With kind permission from PO Karlström, Karolinska Hospital, Stockholm.)

Figure 4. Arbitrary illustration of how ovarian reserve may influence ART results (graphs) relative to age (x-axis). The chance of a pregnancy is age-dependent but varies according to differences in ovarian reserve/genetic prerequisite (y-axis; good, "normal" and poor). Decline may be accelerated (purple) due to, for example, smoking or surgery (black) such as oophorectomy or ovarian resection.
Controlled ovarian hyperstimulation

The introduction of controlled ovarian hyperstimulation (COH) in the 1980s markedly increased IVF treatment success rates. The principle of COH is to stimulate the ovaries to produce numerous mature oocytes by inducing follicle growth using supra-natural doses of exogenous FSH. It is only possible to recruit follicles that have already reached the FSH-sensitive stage of maturation by endogenous recruitment, as described above.

Typically, FSH is given as daily subcutaneous injections for 10±2 days using individual doses of recombinant FSH (rFSH) or a urinary-derived gonadotrophin such as human menopausal gonadotrophin (hMG) or uro-follitropin. The ovarian response is monitored by vaginal ultrasonography after 5–7 days of injections, usually in combination with assay of serum estradiol. Depending on the response, further monitoring and doses are individualized.

During COH, estradiol from the growing follicles reaches well above normal levels, which, due to the feedback mechanism on the anterior pituitary, may cause a premature LH surge. To prevent premature luteinisation and ovulation, a premature endogenous LH surge has to be blocked. There are two principal ways of doing this:

1. Long down-regulation of the pituitary gland by use of a GnRH agonist, usually started a week ahead of expected menstrual bleeding – after which injections are started – and typically continued during the injection phase until ovulation is induced with hCG. The GnRH agonist exhausts the pituitary, and shuts off gonadotrophin secretion after an initial flare-up.
2. Use of a GnRH antagonist. The patient starts FSH injections on one of the first 2–3 days of menstrual bleeding. The GnRH antagonist (daily injections) is added when the largest follicles reach the size of 10–12 mm, or more often (according to a fixed protocol) on day 5 of FSH stimulation and it is then continued during the remaining period of injections.

When the leading follicles reach 18–22 mm in diameter the stimulation ends with the induction of final oocyte maturation and ovulation, by simulating the LH surge by means of an injection of hCG. OPU is commenced about 36 hours later.
Ovarian reserve tests and IVF-ICSI treatment outcome

ORTs and ovarian response

An excessive response to COH may result in ovarian hyperstimulation syndrome (OHSS). OHSS is a complication characterized by increased capillary permeability, with a shift in fluid from the intravascular space to the extravascular space and the peritoneal and pleural cavities. In its most severe form, including marked ovarian enlargement, haemoconcentration, and an increased risk of thrombosis together with renal failure (35), OHSS may be life-threatening.

The opposite of OHSS is the so-called “poor response” (or in the worst cases, non-response), when only a few follicles grow as an effect of COH. There is no consensus definition of poor response (76), although (the arbitrary) definitions usually specify less than three to less than five follicles of sufficient size at the time for hCG (ovulation induction) at COH, or corresponding amounts of retrieved oocytes (4).

As ovarian reserve is reflected by ovarian responsiveness to COH (27), ovarian reserve is the main determinant of ovarian responsiveness. Individual ovarian responses to COH can thus be estimated by assessment of ovarian reserve. The aim of COH is to achieve a response as good as possible for the expected poor responder and to reduce the risk of OHSS among women likely to be high responders. For the individualisation of COH, it is therefore of great importance to assess ovarian reserve before IVF (77).

Apart from chronological age alone, it has long been suggested that more specific information on ovarian reserve may improve prediction of pregnancy chances where ART is concerned. Numerous studies have been aimed at describing the associations between different ORTs and treatment outcome in assisted reproduction.

Previously, studies on ORTs and ART have often excluded women with polycystic ovaries (78), especially women also fulfilling the diagnostic criteria of PCOS (79), although these women constitute a large subgroup in ART and indeed are a group of patients with well-preserved ovarian reserves (49, 80). Also, ORT studies have generally been designed to predict poor responses to ovarian stimulation, including risk of treatment cycle cancellation (81-86), and have established the role of ORTs as an aid to predict ovarian response (27).

Although estimates of ovarian reserve would theoretically also qualify as prognostic assessments as regards pregnancy and live birth after ART, the majority of studies have failed to convincingly establish such associations, nor have they been able to distinguish thresholds beyond which pregnancies do not occur.

ORTs can be compared with screening tests, i.e. tests used to identify the risk of ‘disease’ (such as ‘cancer’ – ‘not cancer’) usually relative to a cut-off
level and the value of a test is thus dependent on the reciprocal relationship of sensitivity (case) and specificity (non-case), varying according to which threshold or cut-off level is chosen. Previous studies suggested that cut-off levels applied to give ORTs high sensitivity for pregnancy after IVF-ICSI must be set at a level usually so far from the normal distribution that the clinical value of the test is lost as a result of loss of specificity (27). This has led to a sceptical point of view concerning the clinical usefulness of ORTs. That is, ORTs are generally considered to be poor predictors of pregnancy and live birth, owing to their poor discriminative ability in individual cases.

The predominant opinion has thus been that ORTs are mainly beneficial as regards quantitative assessment in the design of protocols for COH. This applies particularly to AFCs and to AMH, where predictive abilities as regards poor and excessive responses to COH have proved to be equally strong. Thus AFCs and AMH levels may serve as bases for individualization of COH protocols (87, 88). In the case of AMH in particular, the dose-response relationship in COH is now well recognised (4).

**ORTs, live birth rates and oocyte quality**

Some recent articles have, however, demonstrated an association between serum AMH levels and live birth rates (77, 89-91). When such an association has been found, the interpretation has been that it is mediated by the strong relationship between AMH and ovarian response, i.e. oocyte yield (92, 93). It has thus been emphasised that this association is merely quantitative; the more oocytes retrieved, the more embryos to select among for transfer (93). Thus, a major question is whether or not the predictive value of AMH levels and other ORTs as regards live birth rates is solely mediated through an association with ovarian response or whether they also reflect quality aspects of the oocytes.
Aims

The aims of the present work were to study different methods of ovarian reserve estimation, the individual importance of the methods, the relationships between them, and their associations with and prognostic capacities as regards IVF-ICSI treatment outcome and oocyte quality in large groups of unselected women, by evaluating

I whether or not menstrual cycle length (MCL) is associated with success rates in IVF-ICSI and whether or not MCL could be useful as a marker of ovarian reserve (Paper I),

II the associations between basal FSH levels, basal LH levels and combinations of FSH and LH levels, and pregnancy and live birth rates in IVF-ICSI (Paper II),

III the association between AFC and treatment outcome in a large unselected cohort of IVF-ICSI patients, covering the entire range of AFCs (paper III),

IV the association between serum AMH levels and treatment outcome, also in a large unselected cohort of IVF-ICSI patients (Paper IV), and

V the relationships between the above-mentioned ORTs and by determining which basal marker, or which combination of markers, provides the best model for prediction of ovarian response and the chance of a live birth in IVF-ICSI (Paper V).
Material and methods

Swedish legislation regulating ART

Since January 2003, Swedish legislation on ART states: “after fertilization (of eggs) outside the body, only one fertilized egg may be transferred back to the woman”. Two fertilized eggs may, however, be transferred if the risk of a twin pregnancy in an individual case is considered low. Current Swedish rules on ART are specified (in Swedish) in “Socialstyrelsens föreskrifter och allmänna råd om användning av vävnader och celler i hälso- och sjukvården och vid klinisk forskning” (http://www.socialstyrelsen.se/sosfs/2009-32).

Effects of Swedish legislation and use of a prediction model

Before 2003, two embryos (double embryo transfer, DET) were frequently transferred on a routine basis, resulting in twins in about 25% of IVF-ICSI pregnancies (94). Since January 2003 the majority of embryo transfers have been elective single transfers (eSETs) and the multiple pregnancy rate has fallen to well below 10% (95).

The choice between eSET and DET may be difficult when the pregnancy chance (and thus also the risk of a twin pregnancy) is estimated to be low as a result of suboptimal embryo quality, high female age, low ovarian response and/or several previous failed ART attempts. A prediction model based on the factors determining the probability of implantation for each embryo in each treatment and in each individual woman, and with the twin pregnancy risk in cases of DET taken into account, would thus be of help.

The Carl von Linné Clinic has developed such a prediction model (96, 97), based on the first published evidence-based embryo scoring system – the ‘Integrated Morphology Cleavage’ score (IMC score) (2). In the IMC scoring system, morphological embryo variables generate an objective score that relates to the implantation chance for each single embryo. The prediction model has been in use since 2004 and has resulted in maintained pregnancy rates (when transfers of frozen and thawed embryos are included) but with significantly reduced twin rates, also in comparison with other IVF clinics in Sweden (98). The prediction model has thus been used in connection with the majority of treatments included in the present work, with the exception of those in Paper I (see Table 1 below). In addition, the proportion
of treatments with DET has successively decreased from paper to paper; i.e. the later the study period, the fewer treatments with DET (Table 3).

Ethical considerations

All data were prospectively collected in patient files as part of the infertility work-up. Inclusion in analyses did not in any way alter the clinic’s routine IVF-ICSI protocols or result in any additional intervention at the time of treatment. Neither was there any risk of privacy violation; all data were de-identified at extraction for statistical analysis and all statistical processing was carried out at a group level.

Studies I–III started at a point in time when ethics approvals for these kinds of studies were not considered necessary. Therefore, such approvals were not applied for. However, the Regional Ethics Committee at the University of Uppsala approved Studies IV and V, and that approval retrospectively also sanctioned the first three studies. The ethics approval required that patients who could not be asked for informed consent prospectively were informed by an advertisement published in Uppsala Nya Tidning.

Patients and treatments

All data were collected from patient files at the Carl von Linné Clinic, Uppsala, Sweden. Data were prospectively collected for future evaluation in studies on IVF-ICSI treatment outcome. Patients were included regardless of duration of infertility, or whether their infertility was primary or secondary, and regardless of diagnosis or cause of infertility.

The study periods overlapped. The included patients (treatments) in the different studies are therefore to some extent the same. The study periods and numbers of patients and treatments are presented in Table 1.

<table>
<thead>
<tr>
<th>Paper</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients (n)</td>
<td>3228</td>
<td>745</td>
<td>2092</td>
<td>892</td>
</tr>
<tr>
<td>Treatments (n)</td>
<td>6271</td>
<td>1328</td>
<td>4308</td>
<td>1230</td>
</tr>
</tbody>
</table>

Figures are numbers of started treatment cycles in Papers I, III, IV and V, and per number of ovum pick-ups (OPUs) in Paper II (inclusion and data extraction were restricted to per OPU and ET, which is why figures on started cycles are missing).
Design and methods

All studies were prospective clinical cohort studies.

Mode of controlled ovarian hyperstimulation (COH) and mode and day of embryo transfer in the respective studies are given in Tables 2 and 3 below. Doses of rFSH or hMG were individualized according to the patients’ ovarian reserve status as assessed by the respective available ORTs.

Luteal phase support was given vaginally to all patients for two weeks after embryo transfer. In all studies, pregnancy was defined as visualization of a gestational sac in vaginal ultrasonography in gestational week 7+, i.e. a clinical pregnancy; biochemical pregnancies were not included. Delivery of at least one living child defined live birth. Live birth rates were used as a proxy of oocyte quality.

Statistically, all included treatments and treatment outcomes have been treated equally. That is, pregnancy rate and live birth rate were binary variables (yes/no) defined per treatment cycle, OPU or ET regardless of whether or not the embryo transfer was single or double, or if there was a singleton or a twin pregnancy at delivery.

Table 2. Mode of ovarian hyperstimulation in the respective studies

<table>
<thead>
<tr>
<th>Paper</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>GnRH agonist %</td>
<td>96.5</td>
<td>100</td>
<td>91.7</td>
<td>100</td>
</tr>
<tr>
<td>GnRH antagonist %</td>
<td>3.5</td>
<td>-</td>
<td>8.3</td>
<td>-</td>
</tr>
<tr>
<td>rFSH %</td>
<td>96</td>
<td>100</td>
<td>84.7</td>
<td>51.6</td>
</tr>
<tr>
<td>hMG %</td>
<td>4</td>
<td>-</td>
<td>15.3</td>
<td>48.4</td>
</tr>
</tbody>
</table>

GnRH = gonadotrophin-releasing hormone; rFSH = recombinant follicle-stimulating hormone; hMG = human menopausal gonadotrophin

Table 3. Mode and day of embryo transfer in the respective studies

<table>
<thead>
<tr>
<th>Paper</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>SET %</td>
<td>42.6</td>
<td>62</td>
<td>50.3</td>
<td>73.6</td>
</tr>
<tr>
<td>DET %</td>
<td>57.4</td>
<td>38</td>
<td>49.7</td>
<td>26.4</td>
</tr>
<tr>
<td>ET day 2 %</td>
<td>93.7</td>
<td>100</td>
<td>100</td>
<td>98.5</td>
</tr>
<tr>
<td>ET day 3 %</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.1</td>
</tr>
<tr>
<td>ET day 5 %</td>
<td>6.3</td>
<td>-</td>
<td>-</td>
<td>1.4</td>
</tr>
</tbody>
</table>

SET = single embryo transfer; DET = double embryo transfer; ET = embryo transfer. Day 2, 3 or 5 denotes day after ovum pick-up.

Ovarian sensitivity index

As mentioned in the Introduction, there is no consensus definition of poor response to COH; neither is there a consensus definition of normal or excessive response. In this work, the ‘Ovarian sensitivity index’ (99) (OSI) is used to express ovarian responsiveness. OSI is the ratio between oocyte
yield and administered FSH dose, i.e. the amount of oocytes retrieved divided by the given dose of FSH (or hMG) at the time of COH. Taking both the stimulus (the given dose of FSH or hMG) and the resulting effect (oocyte yield) into account, OSI defines ovarian response in a more objective and less arbitrary manner than oocyte yield alone.

**Paper I**

All patients undergoing treatment during the study period were included, with the exception of women considered chronically anovulatory (defined by cycles of >50 days). Self-reported mean menstrual length during the previous year was recorded before the initiation of treatments. Menstrual cycle length was defined as number of days from the first day of bleeding until the day before the next bleeding period. This figure was recorded and subsequently related to treatment outcome.

Menstrual cycle lengths were normally distributed and were categorized into six groups: <26 days, 26–27 days, 28–29 days, 30–31 days, 32–34 days, and >34 days. The grouped variable was used as a continuous variable in the analyses.

In the statistical models we analysed the outcomes pregnancy and live birth and continuous outcome variables (number of transferred embryos, amount of rFSH or hMG used during COH and duration of ovarian stimulation, amount of oocytes retrieved at OPU, and embryo score) versus the predictors menstrual cycle length or menstrual cycle length and age. Associations concerning frozen and thawed embryo transfers were also evaluated.

**Paper II**

Prior to the first IVF-ICSI treatment women had blood samples analysed on menstrual cycle day 2, 3 or 4 for basal levels of FSH, LH and estradiol (E2). There were no restricting limits of serum levels of FSH or LH as regards inclusion in analyses. Only fresh treatment cycles were included in the analyses.

**Assays:**

Serum levels of FSH and LH were analysed by using two methods; in 6.3% by AutoDELPHIA (Wallac OY, Åbo, Finland) and in 93.7% using ARCHITECT (Abbott Laboratories, IL, USA). (This was due to a change in methods from AutoDELPHIA to ARCHITECT initiated by the laboratory.) The total coefficients of variation for FSH were 2.4–3.7% (AutoDELPHIA) and 2.4% (ARCHITECT). The corresponding figures for LH were 2.7–10.1% (AutoDELPHIA) and 3.0–3.5% (ARCHITECT).

E2 was also analysed by using AutoDELPHIA (Wallac OY, Åbo, Finland) until November 2004, when it was changed to ARCHITECT (Abbott
Laboratories, IL, USA). The intra- and inter-assay coefficients of variation were 2.4–3.5 % and 0.8–3.1% for AutoDELPHIA, and 1.6–2.9% and <1.3% for ARCHITECT, respectively.

According to the laboratory, the respective methods correlated well with each other in all aspects and have thus been considered equivalent.

Because of skewed distributions, serum concentrations of FSH and LH were logarithmically transformed (using natural logarithms). Logarithmically transformed values of basal FSH and LH were graphically arranged in a bivariate correlation plot. The graph was then divided into four quadrants by analysing all given combinations of FSH and LH, aiming at an optimal pregnancy rate in any of the quadrants containing at least 15% of the observations. This provided ‘linked cut-off levels’ for FSH and LH, determining four groups of combinations of FSH and LH: low FSH–high LH (group 1), low FSH–low LH (group 2), high FSH–high LH (group 3), and high FSH–low LH (group 4). The results from groups 2 and 3 were equivalent and were therefore subsequently combined, thus rendering three groups for comparison:

1. Group 1: low FSH–high LH
2. Groups 2+3: low FSH–low LH or high FSH–high LH
3. Group 4: high FSH–low LH

In the statistical models we analysed basal FSH, basal LH and combinations of FSH and LH versus pregnancy and live birth rates and other treatment outcome data, with and without adjustment for age.

Paper III

Before their first treatment, all patients were scanned by means of 2D ultrasoundography for AFC. Patients were scanned regardless of day in the cycle, using a 7.5 MHz vaginal probe. AFC was recorded as the sum of all follicles of 2–10 mm in diameter. Patients were enrolled regardless of the amount of antral follicles, thus covering the extremes of AFC (i.e. both few and many) and including women with polycystic ovaries and PCOS. Only fresh treatment cycles were included in the analyses.

Polycystic ovaries were diagnosed by means of vaginal ultrasonography (78, 79). Thus, the diagnosis of ovulatory and anovulatory polycystic ovaries was restricted to the combination of 12 or more antral follicles per ovary (i.e. a total of at least 24 antral follicles) and the presence or absence of clinical signs of anovulation (amenorrhoea or irregular cycles of >35 days). Serum levels of androgens were not analysed, nor were clinical signs of hyperandrogenism recorded.
The main outcome variables were pregnancy and live birth per started stimulation, OPU and ET. AFC was stratified into four groups defined by cut-offs corresponding to pregnancy rates of 15%, 25% and 35%, and the grouped variables were used as continuous variables in the analyses. Odds ratios (ORs) for each AFC stratum with 95% confidence intervals and p-values were presented. All models were adjusted for age (continuous variable). The doses of FSH/hMG given during ovarian stimulation, the numbers of oocytes and other treatment variables were analysed by using generalized estimating equation (GEE) models for dichotomous or for continuous outcome variables, where the predictor was the four-group AFC categorisation. Pregnancy rates and live birth rates per started cycle were also compared between women with and without polycystic ovaries, unadjusted and with adjustment for age and BMI.

Paper IV
Before treatment all patients had a blood sample taken for analysis of AMH in serum. Blood was drawn regardless of day in the cycle. Specimens were frozen within 24 hours and stored at -20 °C until analysis was carried out within two weeks. A new AMH assay was performed if IVF-ICSI treatment was not commenced within 12 months, or else the treatment was excluded from analysis. Patients were enrolled regardless of their levels of AMH. The maximum female age for inclusion was 42 years. Only fresh treatment cycles were included in the analyses.

Patients underwent ultrasonography for AFC. AFC was defined as the sum of all follicles 2 to 10 mm in size. Ovaries with a total of AFC >23 were classified as polycystic (78). Polycystic ovaries combined with a/oligomenorrhoea were classified as polycystic ovary syndrome (PCOS) (79).

AMH levels were analysed by enzyme-linked immunosorbent assay (ELISA) using DSL kits (Webster, Texas, USA). Coefficients of variation were 9% (intra-assay) and 6% (inter-assay) for low levels (mean 2 ng/ml), and 6% (intra-assay) and 3% (inter-assay) for high levels (mean 8 ng/ml). The conversion factor (pmol/l to ng/ml) was 7.14 pmol/l=1ng/ml.

Power calculation
We intended to categorize the observations into four groups for internal comparison, and assumed the mean pregnancy rate per started treatment cycle to be 15% in the group with the poorest treatment outcome. Provided that all observations were independent, we needed to include a total of 1072 observations (i.e. 268 in each group) for a power of 80% and an α value of <0.05 to detect a difference in pregnancy rates of at least 10% between the poorest group and the best group.
On February 1, 2011, the lab (Karolinska Hospital Laboratory, Stockholm) switched assays from DSL to the AMH Gen II assay (Beckman Coulter). Although the two methods have equivalent precision and results correlate well with each other, the Gen II method provides values that differ from those given by the DSL method (100) and there was no reliable conversion factor available. For this reason, inclusion was restricted to patients whose AMH levels had been analysed by using DSL kits, and who underwent treatment before the end of June 2011. Since patients were allowed to contribute with more than one treatment each, we increased the included amount of observations to reduce the risk of a loss in power.

AMH values were stratified into four groups using cut-offs defined by the 25th, 50th and 75th percentiles of the observations. Although ovarian responses to COH increased as expected quartile by quartile, the two groups in the middle were alike in terms of outcome, with equal pregnancy rates (28.2% and 28.25%, respectively; \( p=0.99 \)) and live birth rates (20.0% and 21.4%, respectively; \( p=0.67 \); figures per started cycle). The two middle groups were therefore joined, thus resulting in three AMH classes according to the 25th and 75th percentiles. The grouped variable with values 1, 2 and 3 was then used as a continuous variable in the analyses. Odds ratios for each AMH group, with 95% confidence intervals (CIs) and p-values were presented. All models were adjusted for age.

Doses of rFSH or hMG given at the time of COH, and other treatment variables were analysed by using GEE models for dichotomous and for continuous outcome variables. The predictor was the three-group AMH categorization. The results are presented as means or percentages with 95% CIs per AMH group with p-values, unadjusted and adjusted for age.

In both Study III and Study IV, we checked if our findings also held true when analyses were restricted to the first IVF cycle (index treatment) for each patient.

Paper V
The study cohort was the same as in Paper IV.

Women were asked for their mean menstrual cycle lengths (MCLs) during the preceding year. Cycles longer than 40 days were considered to be anovulatory and excluded, since at the time of inclusion it was unclear if anovulatory status was hypo-, normo- or hypergonadotrophic.

Within one year before treatment all patients had had a blood sample drawn for analysis of serum AMH (DSL ELISA kit; Webster, Texas, USA) and basal (cycle day 2, 3 or 4) FSH and LH (chemiluminescent immunoassay ARCHITECT; Abbott Laboratories, IL, USA).
See Paper IV above for laboratory details concerning AMH. For FSH, intra-assay CVs were 3.3% and 3.6% for low and high levels, respectively. The corresponding CVs for LH were 4.6% and 3.7%.

Antral follicles were counted by one of three investigators before treatment by means of vaginal ultrasonography as described above (Paper III).

Mean logOSI -1 SD defined a poor response and mean logOSI +1 SD defined an excessive response (101).

Not all patients had complete data on all ORTs. See Table 4 below for details.

Table 4. ORTs and numbers of patients and started treatments in Paper V

<table>
<thead>
<tr>
<th>ORT</th>
<th>Patients (n)</th>
<th>Started treatments (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMH</td>
<td>892</td>
<td>1230</td>
</tr>
<tr>
<td>AFC</td>
<td>595</td>
<td>830</td>
</tr>
<tr>
<td>FSH</td>
<td>740</td>
<td>1015</td>
</tr>
<tr>
<td>LH</td>
<td>689</td>
<td>946</td>
</tr>
<tr>
<td>FSH–LH</td>
<td>687</td>
<td>942</td>
</tr>
<tr>
<td>MCL</td>
<td>829</td>
<td>1151</td>
</tr>
<tr>
<td>Complete data on all ORTs</td>
<td>443</td>
<td>620</td>
</tr>
</tbody>
</table>

AFC = antral follicle count, AMH = anti-Müllerian hormone, FSH = follicle-stimulating hormone, LH = luteinising hormone, MCL = menstrual cycle length, ORT = ovarian reserve test. FSH–LH is a grouped variable; see description under ‘Paper II’ above.

AFC and levels of AMH were analysed as continuous variables. FSH and LH levels were grouped into three combinations defined by the cut-offs 6.7 U/l (FSH) and 4.9 U/L (LH) as described above (Paper II). Menstrual cycle lengths were grouped into six classes: <26, 26-27, 28-29, 30-31, 32-34 and >34 days (Paper I). The grouped FSH–LH and MCL variables were then used as continuous variables in the analyses. Primary outcome was live births per started treatment cycle. Odds ratios with 95% confidence intervals and p-values were presented.

Statistics

The distributions of FSH and LH levels, AFC, AMH levels and OSIs were skewed. Where appropriate, values were log-transformed (natural logarithms) for statistical analyses.

The main statistical analyses (Papers I–V) were carried out by using generalized estimating equation (GEE) models (102) with and without adjustment for female age. GEE models account for clustered data and correlations between observations in generalized linear regression models. The same pa-
tient can thus be included at different points in time. This approach allowed us to include all available cycles.

Comments

Papers I and II
In Papers I and II, differences in continuous outcome variables between groups (number of transferred embryos, amount of rFSH or hMG used at and duration of ovarian stimulation, amount of oocytes retrieved at OPU, and embryo scores) were analysed by means of one-way analysis of variance (ANOVA); the factors were the six-group categorizations of MCL (Paper I) and the three-group categorization of combined FSH and LH levels (Paper II). Differences in dichotomous outcomes (implantation, pregnancy, live birth) were evaluated by means of logistic regression analysis. Since some couples contributed with more than one treatment, the use of these methods was not entirely correct. For this reason, Papers I and II were appropriately re-analysed using GEE models after the papers were published. All results (i.e. from logistic regression analysis and ANOVA as well as from GEE models) are presented below in the Results section (Tables 7 and 9).

The ability of basal FSH–LH combinations and age (Paper II) to discriminate between pregnancy and non-pregnancy was assessed by using Area Under the Receiver Operating Characteristic curves (AUROCs) based on figures from the GEE models.

Papers III and IV
To study the associations between logAFCs and pregnancy rates, and between logAMH values and pregnancy rates graphically, we used a generalized additive model (GAM) to plot the dependent variable logit $P_{\text{pregnancy}}$ per started cycle (Figures 8 and 10).

Analysing the outcomes restricted to the index treatment only was carried out by using logistic regression analysis for dichotomous outcomes (pregnancy, live birth) and ANOVA for continuous outcome variables.

Paper V
The internal relationships between five different ORTs (logAFC, logAMH, logFSH, logLH, and MCL) were determined by use of Pearson’s correlation test.

The discriminative capacities of the different ORTs and/or female age to predict poor response, excessive response and live birth per started stimulation were assessed by means of AUROCs.
Statistical software

The statistical package SAS (v. 8.02 (Paper I), v. 9.1 (Papers II–IV), v. 9.3 (Paper V), SAS Institute, Cary, NC, USA) was used for all calculations. A p-value <0.05 was considered significant.

R (v. 2.14.1, http://www.r-project.org) was used for construction of the logit graphs (Papers III and IV).
Results

Table 5 shows ages in the study populations and amounts and types of treatments in the different studies.

Table 5. Ages, amounts of patients and treatments, and types of treatment.

<table>
<thead>
<tr>
<th>Paper</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD)</td>
<td>35(4.3)</td>
<td>36(4.1)</td>
<td>35.3(4.2)</td>
<td>36(4.2)</td>
</tr>
<tr>
<td>Patients (n)</td>
<td>3228</td>
<td>745</td>
<td>2092</td>
<td>892</td>
</tr>
<tr>
<td>Started cycles (n)</td>
<td>6271</td>
<td>N.A.</td>
<td>4308</td>
<td>1230</td>
</tr>
<tr>
<td>OPU (n)</td>
<td>5876</td>
<td>1328</td>
<td>4004</td>
<td>1111</td>
</tr>
<tr>
<td>ET (n)</td>
<td>5528</td>
<td>1234</td>
<td>3701</td>
<td>1037</td>
</tr>
</tbody>
</table>

Amounts of treatments patients have undergone, %:

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49</td>
<td>51</td>
<td>44.7</td>
<td>69.3</td>
</tr>
<tr>
<td>2–3</td>
<td>41</td>
<td>42</td>
<td>42.7</td>
<td>30</td>
</tr>
<tr>
<td>&gt;3</td>
<td>10</td>
<td>7</td>
<td>12.6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Mean no. of treatments (SD) 1.94 (1.26) 1.78 (1.04) 2.06 (1.33) 1.38 (0.63)

Types of treatment:

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>IVF %</td>
<td>57.6</td>
<td>56.5</td>
<td>53.4</td>
<td>59.3</td>
</tr>
<tr>
<td>ICSI %</td>
<td>40</td>
<td>41.8</td>
<td>35.7</td>
<td>34</td>
</tr>
<tr>
<td>Combined IVF and ICSI %</td>
<td>2.4</td>
<td>1.7</td>
<td>10.9</td>
<td>6.7</td>
</tr>
</tbody>
</table>

ET = embryo transfer, OPU = Ovum pick-up, SD = standard deviation

Paper I

There were subtle but significant reductions in menstrual cycle lengths with increasing age, with a mean cycle length of 30 (±0.2 [SEM]) days among women of 28 years or younger, decreasing to 28.1 (±0.1) days among women of 40 years or older (Figure 5, p<0.0001).
Figure 5. The decline of mean MCL in relation to increasing age at OPU, illustrating the subtle but significant difference between the younger and older women in the study group. Analysis was carried out as regards the first treatment only. Days ± SEM. $P_{\text{trend}} < 0.0001$.

Table 7 shows study group characteristics and treatment outcome according to menstrual cycle lengths (six groups). There were highly significant positive trends as regards pregnancy and live birth rates per started cycle, OPU and ET with increasing mean menstrual cycle length. Pregnancy rates per ET with and without age adjustment in the different menstrual cycle length groups are illustrated in Figure 6.

Figure 6. Pregnancy rate per embryo transfer (ET) in relation to stratified menstrual cycle length (MCL), unadjusted and adjusted for age. Mean values ± SEM. $P_{\text{trend}} < 0.0001$ for both.
The total FSH/hMG dose required for ovarian stimulation was inversely associated with menstrual cycle length; women with short menstrual cycles thus requiring higher doses to reach full follicle maturation. Conversely, the numbers of eggs retrieved at OPU were fewer in women with shorter cycles. Combined, these two findings indicate a strong correlation between OSI and menstrual cycle length. The embryo scores (IMC scores on day 2) followed the same pattern as the oocyte yields. The number of cancelled ovarian stimulations was highest in the shortest cycle groups. All these findings were also significant after adjustment for age (p<0.01). As mentioned above (Methods section), the use of ANOVA was not correct, since some couples contributed with more than one treatment. The relevant trend tests were therefore repeated on the basis of GEE models, with essentially the same results (Table 7).

To illustrate the impact of cycle length on results, the following odds ratios (ORs) were calculated (Table 6):

Table 6. Pregnancy rates and live birth rates per ET relative to menstrual cycle lengths (six groups) with and without age adjustment.

<table>
<thead>
<tr>
<th></th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pregnancy/ET</td>
<td>1.17</td>
<td>1.11–1.23</td>
</tr>
<tr>
<td>Pregnancy/ET, age-adjusted</td>
<td>1.11</td>
<td>1.06–1.17</td>
</tr>
<tr>
<td>Live birth/ET</td>
<td>1.14</td>
<td>1.08–1.20</td>
</tr>
<tr>
<td>Live birth/ET age-adjusted</td>
<td>1.08</td>
<td>1.02–1.14</td>
</tr>
</tbody>
</table>

ET = embryo transfer, 95% CI = 95% confidence interval, OR = odds ratio.

Live births per ET relative to age, adjusted for menstrual cycle length, gave an OR of 0.91 [95% CI 0.90–0.92]. These results thus suggest that the probability of live birth decreases by 7.4% (1.08⁻¹=0.926; 1-0.926=0.074) for each menstrual cycle length group downward from >34 days, given similar ages and the fact that ET was achieved. Correspondingly, live birth rates decrease by 9% (1-0.91=0.09) per year of advancing age, given similar menstrual cycle length.

Comparing the two extreme groups of menstrual cycle length, women with cycles > 34 days had almost twice the chance of conceiving vs. those with cycles of less than 26 days.

Pregnancy rates as regards transfers of frozen and thawed embryos showed an increasing trend with longer menstrual cycle length (OR 1.170 [95% CI 1.029–1.329]) from 24% (SEM 6.1) in the shortest cycle group to 38.8% (SEM 6.0) among women with cycles of more than 34 days (p=0.016; Table 5). The figures remained significant after adjustment for age at OPU (OR 1.160 [95% CI 1.021–1.318], p=0.023).
Table 7. Distribution of menstrual cycle lengths, study group characteristics and treatment data

<table>
<thead>
<tr>
<th>MCL, days</th>
<th>n</th>
<th>&lt;26</th>
<th>26-27</th>
<th>28-29</th>
<th>30-31</th>
<th>32-34</th>
<th>&gt;34</th>
<th>P^d</th>
<th>P_{GEE}^e</th>
<th>P_{GEE \text{ age adj}}^f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution, %</td>
<td>100%</td>
<td>7</td>
<td>20.4</td>
<td>46.5</td>
<td>14.7</td>
<td>5.7</td>
<td>5.7</td>
<td>NA</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age</td>
<td>6271</td>
<td>35.7 (0.2)</td>
<td>35.4 (0.1)</td>
<td>35.0 (0.1)</td>
<td>33.9 (0.1)</td>
<td>33.0 (0.2)</td>
<td>32.6 (0.2)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Pregnancies/stim start %</td>
<td>6271</td>
<td>20.1 (1.9)</td>
<td>26.2 (1.2)</td>
<td>29.7 (0.8)</td>
<td>33.2 (1.5)</td>
<td>35.6 (2.5)</td>
<td>39.9 (2.6)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Live births/stim start %</td>
<td>5876</td>
<td>22.7 (2.1)</td>
<td>28.6 (1.3)</td>
<td>31.2 (0.9)</td>
<td>35.6 (1.6)</td>
<td>37.7 (2.6)</td>
<td>41.9 (2.7)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Live births/ET %</td>
<td>5528</td>
<td>16.9 (2.0)</td>
<td>23.7 (1.3)</td>
<td>25.3 (0.8)</td>
<td>28.2 (1.5)</td>
<td>31.3 (2.6)</td>
<td>30.9 (2.6)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.0013</td>
</tr>
<tr>
<td>Embryos to ET (n)</td>
<td>5528</td>
<td>1.61 (0.01)</td>
<td>1.57 (0.01)</td>
<td>1.58 (0.01)</td>
<td>1.57 (0.01)</td>
<td>1.56 (0.02)</td>
<td>1.49 (0.02)</td>
<td>&lt;0.01</td>
<td>0.039</td>
<td>0.204</td>
</tr>
<tr>
<td>FSH or hMG dose, IU</td>
<td>6271</td>
<td>3138 (65)</td>
<td>2895 (37)</td>
<td>2497 (23)</td>
<td>2099 (36)</td>
<td>1828 (54)</td>
<td>1733 (43)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Oocytes at OPU</td>
<td>5851</td>
<td>8.1 (0.2)</td>
<td>8.9 (0.1)</td>
<td>10.2 (0.1)</td>
<td>11.0 (0.1)</td>
<td>12.3 (0.3)</td>
<td>11.2 (0.3)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.0013</td>
</tr>
<tr>
<td>Embryo score 1-10a</td>
<td>5137b</td>
<td>8.8 (0.1)</td>
<td>8.6 (0.1)</td>
<td>8.9 (0.0)</td>
<td>9.0 (0.1)</td>
<td>9.0 (0.1)</td>
<td>9.2 (0.1)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.0013</td>
</tr>
<tr>
<td>FSH-stimulation days</td>
<td>5879c</td>
<td>11.2 (0.1)</td>
<td>11.0 (0.0)</td>
<td>11.0 (0.0)</td>
<td>10.97 (0.1)</td>
<td>10.8 (0.1)</td>
<td>10.9 (0.1)</td>
<td>0.001</td>
<td>0.111</td>
<td>0.144</td>
</tr>
<tr>
<td>Cancelled treatments %</td>
<td>6271</td>
<td>18.9 (1.9)</td>
<td>14.8 (1.0)</td>
<td>10.4 (0.6)</td>
<td>10.2 (1.0)</td>
<td>10.3 (1.6)</td>
<td>10.1 (1.6)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.0013</td>
</tr>
<tr>
<td>Pregnancies/FET %</td>
<td>1023</td>
<td>24.0 (6.1)</td>
<td>25.1 (3.4)</td>
<td>25.6 (1.9)</td>
<td>30.3 (3.6)</td>
<td>33.3 (5.9)</td>
<td>38.8 (6.0)</td>
<td>0.016</td>
<td>0.0198</td>
<td>0.031</td>
</tr>
<tr>
<td>Live births/FET %</td>
<td>1023</td>
<td>16.0 (5.2)</td>
<td>16.2 (2.9)</td>
<td>17.9 (1.7)</td>
<td>24.8 (3.4)</td>
<td>18.2 (4.8)</td>
<td>25.4 (5.4)</td>
<td>0.062</td>
<td>0.064</td>
<td>0.107</td>
</tr>
</tbody>
</table>

n = 6,271 treatment cycles in 3,228 women. Mean values ± standard error of the mean (SEM).
ET = embryo transfer, FET = transfer of a frozen and thawed embryo, FSH = follicle-stimulating hormone, hMG = human menopausal gonadotrophin, IU = international units, OPU = ovum pick-up, stim start = started controlled ovarian hyperstimulation

aIntegrated Morphology Cleavage (IMC) score.
bDifference in n_{embryo score} and n_{embryos to ET} due to day-5 transfers not scored according to IMC.
cDifference in number of observations of days of stimulation vs. n_{OPU} due to missing values.
dP_{trend} using ANOVA for continuous outcome parameters and logistic regression analysis for dichotomous outcome parameters.
eP_{trend} using GEE for dichotomous and continuous outcome parameters.
fP_{trend} using GEE for dichotomous and continuous outcome parameters, adjusted for age.
Table 8 shows the basal levels of serum FSH and LH in the study population.

<table>
<thead>
<tr>
<th>Gonadotrophin</th>
<th>Range</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSH</td>
<td>1.9–23.0 U/l</td>
<td>7.23 (2.3)</td>
</tr>
<tr>
<td>LH</td>
<td>0.8–19.0 U/l</td>
<td>5.21 (2.56)</td>
</tr>
</tbody>
</table>

FSH = follicle-stimulating hormone. LH = luteinising hormone

No significant linear or log-linear effects on treatment outcome were found as regards FSH, LH or the LH:FSH ratio. Dichotomous univariate cut-off levels to distinguish between groups with the largest significant differences in pregnancy rates were: for FSH, 6.7 U/l (p = 0.007), for LH, 4.2 U/l (P =0.006) and for the LH:FSH ratio, 0.545 (p = 0.016). Pregnancy rates were higher with low FSH and high LH levels or a high LH:FSH ratio. After age adjustment, dichotomous FSH and LH remained significant, whereas the LH:FSH ratio did not; neither as a continuous variable, nor stratified into five levels.

Figure 7 shows the bivariate plot of log levels of basal serum FSH and LH. The ‘linked cut-off levels’ for distinction between the groups with the highest (15%) and lowest pregnancy rates in the analysis of combined FSH and LH levels turned out to be similar to the univariate cut-off levels; 6.7 U/l for FSH but slightly higher for LH (4.9 U/l). Using the ‘linked cut-off levels’, the study population was thus divided into four groups of FSH–LH. The four groups demonstrate the mutual dependency of FSH and LH: group 1 (low FSH–high LH) showed superior treatment outcome, whereas group 4 (high FSH–low LH) showed the opposite. Groups 2 and 3 (low FSH–low LH and high FSH–high LH, respectively) were not different from each other (P=0.7) and both these groups showed intermediate pregnancy rates compared with groups 1 and 4. For further analyses, groups 2 and 3 were combined (Table 9). There were no differences in levels of estradiol between groups, neither before (p = 0.53) nor after (p = 0.14) the two middle groups were combined.

Importantly, in a GEE model use of the grouped FSH–LH combinations resulted in a loss of significance as regards levels of both dichotomous FSH and dichotomous LH.

Live birth rates and other variables associated with treatment success followed the same pattern as the pregnancy rates: the numbers of oocytes retrieved at OPU were positively associated with lower FSH–higher LH levels; the contrary was the case for total FSH doses required for ovarian stimu-
lations. There was thus a strong positive association between OSI and low FSH–high LH.

The best discriminative model for prediction of pregnancy included FSH–LH groups 1, 2+3, 4 and female age (AUROC 0.63).

Figure 7. Bivariate correlation plot of logarithmically transformed basal FSH and LH levels. By analysing all given combinations of FSH and LH levels and aiming at an optimal pregnancy rate in one of the quadrants, the ‘linked cut-off levels’ for FSH (6.7 U/L) and LH (4.9 U/L) were provided and four groups of FSH–LH combinations could be determined. The mean pregnancy rate (%) is given in each group (quadrant).

Groups 2 and 3 (lower left and upper right quadrants) did not differ from each other in terms of treatment outcome and were joined in further analyses.
Table 9. FSH–LH groups according to the bivariate plot above (Figure 7). Groups 2 and 3 (low FSH–low LH and high FSH–high LH) did not differ from each other in terms of treatment outcome and are presented together.

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Group 1 (n)</th>
<th>Groups 2&amp;3 (n)</th>
<th>Group 4 (n)</th>
<th>P&lt;sup&gt;d&lt;/sup&gt;</th>
<th>P age adj&lt;sup&gt;d&lt;/sup&gt;</th>
<th>P&lt;sub&gt;gee&lt;/sub&gt;</th>
<th>P&lt;sub&gt;gee age adj&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations (n)</td>
<td>1328</td>
<td>211 (15.9%)</td>
<td>817 (61.5%)</td>
<td>300 (22.6%)</td>
<td></td>
<td></td>
<td>0.0169</td>
<td>NA</td>
</tr>
<tr>
<td>Age</td>
<td>1328</td>
<td>35.3 ± 0.3</td>
<td>36.1 ± 0.1</td>
<td>36.6 ± 0.2</td>
<td>0.0006</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Implantation %</td>
<td>1234</td>
<td>31.4</td>
<td>23.1</td>
<td>17.6</td>
<td>&lt;0.0001</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Pregnancies/OPU %</td>
<td>1328</td>
<td>39.3 ± 3.4</td>
<td>27.7 ± 1.6</td>
<td>22.3 ± 2.4</td>
<td>&lt;0.0001</td>
<td>0.0004</td>
<td>0.0009</td>
<td>0.0038</td>
</tr>
<tr>
<td>Pregnancies/ET %</td>
<td>1324</td>
<td>40.1 ± 3.4</td>
<td>29.9 ± 1.7</td>
<td>24.7 ± 2.6</td>
<td>0.0004</td>
<td>0.0026</td>
<td>0.0009</td>
<td>0.0038</td>
</tr>
<tr>
<td>Live births/OPU %</td>
<td>1328</td>
<td>32.7 ± 3.2</td>
<td>20.7 ± 1.4</td>
<td>17.7 ± 2.2</td>
<td>0.0001</td>
<td>0.0011</td>
<td>0.0004</td>
<td>0.0024</td>
</tr>
<tr>
<td>Live births/ET %</td>
<td>1324</td>
<td>33.3 ± 3.3</td>
<td>22.4 ± 1.5</td>
<td>19.6 ± 2.4</td>
<td>0.0008</td>
<td>0.0005</td>
<td>0.0018</td>
<td>0.0084</td>
</tr>
<tr>
<td>FSH cd3 U/l</td>
<td>1328</td>
<td>5.6 ± 0.1</td>
<td>7.3 ± 0.1</td>
<td>8.2 ± 0.1</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>LH cd3 U/l</td>
<td>1328</td>
<td>7.2 ± 0.2</td>
<td>5.3 ± 0.1</td>
<td>3.7 ± 0.0</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>LH/FSH cd3</td>
<td>1328</td>
<td>1.4 ± 0.1</td>
<td>0.7 ± 0.0</td>
<td>0.5 ± 0.0</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Estradiol cd3 pmol/l</td>
<td>1328</td>
<td>167.5 ± 9.3</td>
<td>165.0 ± 4.2</td>
<td>159.9 ± 3.6</td>
<td>NA</td>
<td>NA</td>
<td>0.53</td>
<td>0.49</td>
</tr>
<tr>
<td>Menstrual cycle length</td>
<td>1304</td>
<td>32.9 ± 0.9</td>
<td>28.6 ± 0.2</td>
<td>28.4 ± 0.4</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>0.0031</td>
<td>0.0032</td>
</tr>
<tr>
<td>BMI kg/m&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1278</td>
<td>23.8 ± 0.2</td>
<td>24.4 ± 0.1</td>
<td>24.2 ± 0.2</td>
<td>0.5079</td>
<td>0.7394</td>
<td>0.633</td>
<td>0.806</td>
</tr>
<tr>
<td>Given FSH dose, IU</td>
<td>1328</td>
<td>1801 ± 71.1</td>
<td>2573 ± 46.2</td>
<td>3000 ± 81.4</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Oocytes at OPU</td>
<td>1328</td>
<td>10.1 ± 0.3</td>
<td>9.1 ± 0.2</td>
<td>8.0 ± 0.3</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Embryo score 1-10&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1223</td>
<td>9.1 ± 0.1</td>
<td>8.9 ± 0.1</td>
<td>8.8 ± 0.1</td>
<td>0.0461</td>
<td>0.0502</td>
<td>0.082</td>
<td>0.086</td>
</tr>
<tr>
<td>No ET %</td>
<td>1328</td>
<td>1.9 ± 0.9</td>
<td>7.5 ± 0.9</td>
<td>9.7 ± 1.7</td>
<td>0.0013</td>
<td>0.0018</td>
<td>0.0004</td>
<td>0.0007</td>
</tr>
<tr>
<td>AFC</td>
<td>1253</td>
<td>23.5 ± 0.9</td>
<td>16.9 ± 0.3</td>
<td>14.8 ± 0.4</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Ovarian area cm&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1243</td>
<td>11.4 ± 0.4</td>
<td>8.6 ± 0.2</td>
<td>7.4 ± 0.2</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>OSI (oocytes/given dose)</td>
<td>1322</td>
<td>7.5 ± 0.3</td>
<td>4.9 ± 0.1</td>
<td>3.7 ± 0.2</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<sup>n</sup> = 1,328 treatment cycles in 745 women. Mean values ± standard error of the mean (SEM).

FSH = follicle-stimulating hormone, LH = luteinising hormone, OPU = ovum pick-up, OSI = ovarian sensitivity index (amount of oocytes retrieved/given dose of rFSH).

IMC score; *Difference in n vs. n<sub>ET</sub> due to missing values; *Difference in n vs. n<sub>OPU</sub> due to missing values; *<sub>趋势</sub> using ANOVA for continuous outcome parameters and logistic regression analysis for dichotomous outcome parameters, without and with age adjustment; *<sub>趋势</sub> using GEE for dichotomous and continuous outcome parameters, without and with age adjustment.
Paper III

Antral follicle counts ranged from three to eighty. The mean overall AFC (±SD) was 19.2 (11.7). Pregnancy rate and live birth rate per started treatment cycle were positively associated with AFC in a log-linear way (Figure 8). The pregnancy rate intervals 0–15%, 16–25%, 26–35% and >35% corresponded to the AFC strata 0–5, 6–11, 12–23 and >23, respectively (Figure 8). Above AFC ~30 (approximately 25% of all treatments) there were no further increases in pregnancy and live birth rates.

Treatment outcomes according to the four AFC groups are shown in Table 10. There were significant associations between AFC strata and the amount of retrieved oocytes, and given doses of FSH/hMG (inverse relationship). There was no interaction between age and AFC with regard to pregnancy rates and live birth rates. Analysing the outcomes restricted to the index (first) treatment only did not substantially change the results (Table 10).

Figure 9 illustrates live birth rates/OPU at different age groups within each AFC stratum (for ease of interpretation, age was here split into three groups; <34, 34–38 and >38 years). Comparing the two extremes of AFCs, women with an AFC of >23 had at least a 50% higher chance of a live birth after treatment than those with an AFC of <5, when age was taken into account. The OR for live birth/OPU relative to stratified AFC was 1.48 [95% CI 1.33–1.64] unadjusted and 1.30 [1.17–1.45] after age adjustment, implying a 30% increased chance of delivering a child for each AFC stratum upward from >5, given the same age.

The influence of AFC on pregnancy rates and live birth rates/OPU was studied in a GEE model with adjustment for age and the number of retrieved oocytes. AFC as well as the amount of oocytes and age remained highly significant as regards both pregnancy rates (p<0.0001) and live birth rates (p<0.001, no interaction between AFC and age or between AFC and amount of oocytes). The live birth rate/OPU odds ratio as regards AFC adjusted for age and amount of oocytes was 1.23 [95% CI 1.10–1.37], indicative of an increased chance of delivering a child of 23% per AFC stratum upward, when both age and oocyte yield are taken into account. The corresponding figure for pregnancy/OPU was 1.25 [1.13–1.38]. These figures imply that AFC also provides qualitative information on ovarian reserve.

Polycystic ovaries were found in 634 women. Of these, 519 were ovulatory (undergoing 931 treatment cycles) and 115 were anovulatory (304 cycles; 24.6% of the treatment cycles in the polycystic ovaries group). Pregnancy and live birth rates (both p<0.0001) were higher among women with polycystic ovaries compared with women without polycystic ovaries, with differences also remaining highly significant after adjustment for age and BMI. There were no significant differences in pregnancy rates (p = 0.76) or
live birth rates \(p = 0.95\) between ovulatory and anovulatory women with polycystic ovaries.

**Figure 8.** Pregnancy rate (%) per started stimulation (solid line) relative to log-transformed antral follicle counts (AFCs), with 95% confidential interval (dotted lines). Density denotes the distribution after log transformation. The AFC numbers given along the x-axis have been back-transformed for ease of interpretation. The corresponding graph for live birth rate showed the same pattern.

AF: antral follicles.

**Figure 9.** Live birth rates/OPU (%) relative to different ages (<34, 34–38, >38 years) within each AFC stratum (<5, 6–11, 12–23 and >23 antral follicles). Mean values ± SEM
Table 10. IVF-ICSI outcome in relation to the antral follicle count (AFC) stratified into four groups.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$n$</th>
<th>AFC ≤ 5</th>
<th>AFC 6-11</th>
<th>AFC 12-23</th>
<th>AFC &gt; 23</th>
<th>$p_{GEE}$</th>
<th>$p_{GEE}$ age adj</th>
<th>$p_{index}$</th>
<th>$p_{index}$ age adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations (n)</td>
<td>4308</td>
<td>129 (3.0%)</td>
<td>893 (20.7%)</td>
<td>2051 (47.6%)</td>
<td>1235 (28.7%)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age</td>
<td>4308</td>
<td>38.1 (37.4 - 38.9)</td>
<td>37.6 (37.4 - 37.9)</td>
<td>35.3 (35.1 - 35.4)</td>
<td>33.5 (33.2 - 33.7)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Antral Follicle Count</td>
<td>4308</td>
<td>4.5 (4.4 - 4.6)</td>
<td>8.4 (8.3 - 8.5)</td>
<td>15.9 (15.8 - 16.0)</td>
<td>33.8 (33.2 - 34.4)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Pregnancy/SS %</td>
<td>4308</td>
<td>13.2 (7.3 - 19.1)</td>
<td>19.8 (17.2 - 22.4)</td>
<td>29.7 (27.8 - 31.7)</td>
<td>36.4 (33.7 - 39.0)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Pregnancy/OPU %</td>
<td>4004</td>
<td>16.7 (9.3 - 24.0)</td>
<td>22.4 (19.5 - 25.3)</td>
<td>31.2 (29.1 - 33.2)</td>
<td>38.9 (36.1 - 41.7)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Pregnancy/ET %</td>
<td>3701</td>
<td>20.2 (11.5 - 29.0)</td>
<td>25.1 (21.9 - 28.3)</td>
<td>33.5 (31.3 - 35.7)</td>
<td>41.2 (38.3 - 44.1)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Live births/SS %</td>
<td>4308</td>
<td>10.1 (4.8 - 15.3)</td>
<td>13.6 (11.3 - 15.8)</td>
<td>22.1 (20.3 - 23.9)</td>
<td>27.4 (24.9 - 29.9)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Live births/OPU %</td>
<td>4004</td>
<td>12.8 (6.2 - 19.3)</td>
<td>15.3 (12.8 - 17.8)</td>
<td>23.2 (21.3 - 25.0)</td>
<td>29.3 (26.7 - 31.9)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Live births/ET %</td>
<td>3701</td>
<td>15.5 (7.6 - 23.4)</td>
<td>17.0 (14.2 - 19.8)</td>
<td>24.9 (22.9 - 26.9)</td>
<td>31.0 (28.3 - 33.8)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>FSH or hMG dose, IU$^a$</td>
<td>4308</td>
<td>4109 (3882 - 4336)</td>
<td>3684 (3599 - 3770)</td>
<td>2475 (2424 - 2527)</td>
<td>1587 (1545 - 1628)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Oocytes at OPU</td>
<td>3990$^c$</td>
<td>5.3 (4.6 - 6.0)</td>
<td>6.8 (6.6 - 7.1)</td>
<td>9.7 (9.5 - 9.9)</td>
<td>11.5 (11.2 - 11.8)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Embryo score 1-10$^c$</td>
<td>3706</td>
<td>8.5 (8.1 - 8.9)</td>
<td>8.6 (8.4 - 8.7)</td>
<td>8.9 (8.9 - 9.0)</td>
<td>9.1 (9.0 - 9.2)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Embryos to ET (n)</td>
<td>3699$^d$</td>
<td>1.7 (1.6 - 1.8)</td>
<td>1.6 (1.6 - 1.7)</td>
<td>1.5 (1.5 - 1.5)</td>
<td>1.4 (1.4 - 1.4)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Canceled treatments %</td>
<td>4308</td>
<td>34.9 (26.6 - 43.2)</td>
<td>20.9 (18.3 - 23.6)</td>
<td>11.2 (9.8 - 12.6)</td>
<td>11.7 (9.9 - 13.5)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>FSH cd3 U/l</td>
<td>2722$^e$</td>
<td>10.0 (9.0 - 11.0)</td>
<td>8.5 (8.3 - 8.8)</td>
<td>7.2 (7.1 - 7.3)</td>
<td>6.0 (5.9 - 6.1)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>LH cd3 U/l</td>
<td>2498$^b$</td>
<td>5.1 (4.4 - 5.8)</td>
<td>4.9 (4.7 - 5.1)</td>
<td>5.0 (4.9 - 5.1)</td>
<td>5.9 (5.7 - 6.2)</td>
<td>0.0004</td>
<td>0.0015</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Menstrual cycle length</td>
<td>4222$^b$</td>
<td>27.0 (26.6 - 27.4)</td>
<td>28.0 (27.1 - 28.9)</td>
<td>28.4 (28.1 - 28.8)</td>
<td>40.7 (38.2 - 43.2)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>BMI kg/m2</td>
<td>4300$^b$</td>
<td>23.2 (22.6 - 23.8)</td>
<td>23.4 (23.1 - 23.6)</td>
<td>23.0 (22.9 - 23.2)</td>
<td>23.5 (23.3 - 23.7)</td>
<td>0.50</td>
<td>0.30</td>
<td>0.63</td>
<td>0.13</td>
</tr>
</tbody>
</table>

AFC was stratified into four groups (0–5, 6–11, 12–23 and >23 antral follicles) based on pregnancy rate intervals (0–15%, 16–25%, 26–35%, and >35%).

$n = 4,308$ treatment cycles in 2,092 women. Mean values ± 95% confidence intervals.

AFC = antral follicle count, BMI = body mass index, cd = cycle day, ET = embryo transfer, FSH = follicle-stimulating hormone, IU = international units, LH = luteinising hormone, OPU = ovum pickup, SS = start of ovarian stimulation.

$^a$FSH or hMG total dose at stimulation; $^b$Difference in observations versus $n_{min}$ due to missing values; $^c$Difference in observations versus $n_{opt}$ due to missing values; $^d$Difference in observations versus $n_{rev}$ due to missing values; $^e$Integrated morphology cleavage embryo score, IMC; $^f$P values denote trend test based on generalized estimating equation (GEE) models, without and with age adjustment; $^g$P$_{trend}$ without and with age adjustment for the index (first) treatment only.
Paper IV

Serum concentrations of AMH ranged from 0.06 to 26.3 ng/ml and were log-normally distributed with a median of 1.6 ng/ml and a mean (SD) of 2.3 (2.5) ng/ml. The pregnancy rate per started treatment cycle was positively and linearly associated with logAMH up to ~AMH 5 ng/ml and then levelled (Figure 10). Figure 11 demonstrates the linear association between AMH and live birth rates, which was highly significant even after adjusting for age.

Figure 10. Pregnancy rate (%) per started cycle (solid line) relative to log-transformed Anti-Müllerian hormone (AMH) levels with 95% confidence intervals (dotted lines). ‘Density’ denotes the distribution after log transformation. The AMH values along the x-axis and the pregnancy rates along the y-axis have been back-transformed for ease of interpretation. n = 1230 IVF-ICSI treatment cycles.

Figure 11. Live-birth rate per started cycle, without (mean values ± SEM, black) and with age adjustment (grey).

Group 1: AMH<0.84 ng/ml (25th percentile); Group 2: AMH 0.84–2.94 ng/ml; Group 3: AMH >2.94 ng/ml (75th percentile).

n = 1230 IVF-ICSI treatment cycles. AMH = Anti-Müllerian hormone.

There were no differences in pregnancy rates (p = 0.22) and live birth rates (p = 0.23) according to diagnosis (unexplained, tubal infertility, endometriosis, male factor, or combinations thereof). There was no AMH level below
which no pregnancies occurred. Below an AMH level of 0.2 ng/ml (all ages) only 18 out of 53 started cycles (34%) resulted in embryo transfer and only three cycles ended in a pregnancy and a live birth.

The 25th and 75th percentiles corresponded to AMH levels of 0.84 and 2.94 ng/ml, respectively. Stimulation and treatment outcomes for the three AMH groups are presented in Table 11. Analyses of the index treatments followed the same pattern as the GEE analyses.

Apart from pregnancy rates and live-birth rates, oocyte yield, embryo score, AFC and mean menstrual cycle length were positively associated with AMH levels. There was an inverse relationship between AMH and the given rFSH/hMG dose at stimulation, cancellation rate and basal FSH levels. Mean AFC in group 3 (AMH > 2.94 ng/ml) was 29, thus comprising women with polycystic ovaries (78). These patients had long menstrual cycles (Table 11) and ¼ of them were anovulatory.

GEE models were used to analyse the influence of AMH levels on pregnancy rates and live birth rates/OPU. The odds ratio [with 95% CI] for live birth/OPU as regards AMH was 1.75 [1.42–2.16] unadjusted and 1.57 [1.26–1.95] after age adjustment. This implies that for each step upwards in AMH group, live birth rates increase by 57% given the same age and provided that OPU is reached. The odds ratio for live birth/OPU for AMH adjusted for oocyte yield was 1.55 [1.22–1.96], p = 0.0003, with oocyte yield remaining significant in the analysis.

In multivariate GEE models controlling for both female age and oocyte yield, AMH was independently significant as regards both pregnancy rate (p<0.001) and live birth rate (p = 0.004). The odds ratio for a live birth/OPU in connection with AMH adjusted for both oocyte yield and age was 1.42 [1.12–1.8], p = 0.004. This indicates a 42% increase in the chance of a live birth for each AMH stratum upward, when both oocyte yield and age are unchanged. There was no interaction between age and AMH, between AMH and oocyte yield or between age and oocyte yield.
Table 11. IVF-ICSI outcome in relation to Anti-Müllerian hormone (AMH) level stratified into three groups.

<table>
<thead>
<tr>
<th>n</th>
<th>Group 1 AMH &lt;0.84</th>
<th>Group 2 AMH 0.84–2.94</th>
<th>Group 3 AMH &gt;2.94</th>
<th>( \text{P}_{\text{GEE}} )</th>
<th>( \text{P}_{\text{GEE, age adj}} )</th>
<th>( \text{P}_{\text{index}} )</th>
<th>( \text{P}_{\text{index, age adj}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observations (n)</strong></td>
<td>309 (25.1%)</td>
<td>613 (49.8%)</td>
<td>308 (25.0%)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>1230</td>
<td>37.1 (36.7 - 37.5)</td>
<td>36.5 (36.2 - 36.8)</td>
<td>33.9 (33.4 - 34.4)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>AMH ng/ml</strong></td>
<td>1230</td>
<td>0.5 (0.5 - 0.5)</td>
<td>1.7 (1.7 - 1.8)</td>
<td>5.4 (5.0 - 5.8)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Pregnancies/SS %</strong></td>
<td>1230</td>
<td>15.5 (9.1 - 17.7)</td>
<td>22.0 (18.6 - 25.4)</td>
<td>32.5 (27.1 - 37.9)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Pregnancies/OPU %</strong></td>
<td>1111</td>
<td>19.5 (14.5 - 24.5)</td>
<td>30.0 (26.3 - 33.8)</td>
<td>39.8 (34.1 - 45.5)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Pregnancies/ET %</strong></td>
<td>1037</td>
<td>22.5 (16.9 - 28.2)</td>
<td>31.9 (28.0 - 35.9)</td>
<td>40.8 (35.0 - 46.6)</td>
<td>&lt;0.0001</td>
<td>0.0002</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>Live births/SS %</strong></td>
<td>1230</td>
<td>10.7 (7.2 - 14.1)</td>
<td>20.7 (17.5 - 23.9)</td>
<td>30.8 (25.7 - 36.0)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Live births/OPU %</strong></td>
<td>1111</td>
<td>13.4 (9.1 - 17.7)</td>
<td>22.0 (18.6 - 25.4)</td>
<td>32.5 (27.1 - 37.9)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Live births/ET %</strong></td>
<td>1037</td>
<td>15.5 (10.6 - 20.4)</td>
<td>23.4 (19.8 - 27.0)</td>
<td>33.3 (27.8 - 38.9)</td>
<td>&lt;0.0001</td>
<td>0.0006</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>FSH or hMG dose, IU</strong></td>
<td>1229b</td>
<td>4108 (3965 - 4251)</td>
<td>2789 (2692 - 2886)</td>
<td>1530 (1454 - 1605)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Oocytes at OPU</strong></td>
<td>1104c</td>
<td>5.9 (5.5 - 6.3)</td>
<td>9.4 (9.0 - 9.7)</td>
<td>12.3 (11.7 - 12.9)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Embryo score (1-10)</strong></td>
<td>1035d</td>
<td>8.9 (8.7 - 9.1)</td>
<td>9.2 (9.0 - 9.3)</td>
<td>9.4 (9.2 - 9.5)</td>
<td>0.0006</td>
<td>0.0002</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>Embryos to ET (n)</strong></td>
<td>1036d</td>
<td>1.3 (1.3 - 1.4)</td>
<td>1.3 (1.3 - 1.3)</td>
<td>1.1 (1.1 - 1.2)</td>
<td>&lt;0.0001</td>
<td>0.065</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Cancelled treatments %</strong></td>
<td>1230</td>
<td>31.1 (25.9 - 36.3)</td>
<td>11.6 (9.0 - 14.1)</td>
<td>8.4 (5.3 - 11.6)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>AFC, no.</strong></td>
<td>830b</td>
<td>10.0 (9.4 - 10.5)</td>
<td>15.8 (15.1 - 16.5)</td>
<td>29.0 (27.4 - 30.7)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>FSH cd3 (IU/L)</strong></td>
<td>1015b</td>
<td>9.3 (8.8 - 9.8)</td>
<td>7.1 (6.9 - 7.3)</td>
<td>6.0 (5.8 - 6.2)</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>LH cd3 (IU/L)</strong></td>
<td>946b</td>
<td>5.1 (4.5 - 5.6)</td>
<td>4.5 (4.3 - 4.7)</td>
<td>5.8 (5.4 - 6.2)</td>
<td>0.09</td>
<td>0.1</td>
<td>0.0822</td>
</tr>
<tr>
<td><strong>Menstrual cycle length</strong></td>
<td>1179b</td>
<td>27.1 (26.8 - 27.3)</td>
<td>29.1 (27.8 - 30.3)</td>
<td>36.9 (32.6 - 41.1)</td>
<td>0.0006</td>
<td>0.0007</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>BMI kg/m²</strong></td>
<td>1190b</td>
<td>23.5 (23.0 - 23.9)</td>
<td>23.5 (23.2 - 23.9)</td>
<td>23.4 (22.9 - 23.8)</td>
<td>0.8</td>
<td>0.5</td>
<td>0.7512</td>
</tr>
</tbody>
</table>

Serum concentrations of AMH were stratified into four groups according to the 25th, 50th and 75th percentiles. The two middle groups were alike in terms of outcome and therefore joined, rendering three AMH groups according to the 25th and 75th percentiles: Group 1, AMH <0.84, group 2, AMH 0.84–2.94 and group 3, AMH >2.94 ng/ml. n = 1,230 treatment cycles in 892 women. Mean values ± 95% confidence intervals.

AFC = antral follicle count, AMH = Anti-Müllerian hormone, BMI = body mass index, cd = cycle day, ET = embryo transfer, FSH = follicle-stimulating hormone, IU = international units, LH = luteinising hormone, OPU = ovum pickup, SS = start of ovarian stimulation.

aIntegrated morphology cleavage embryo score, IMC; bDifference in observations versus n SS due to missing values; cDifference in observations versus n OPU due to missing values; dDifference in observations versus n ET due to missing values; ePtrend, GEE models; fPtrend for index (first) treatment only ± age adjustment.
Paper V

Details on AFC, levels of AMH, LH and FSH, and menstrual cycle lengths are presented in Table 12. Pearson’s correlation coefficients between basal ORTs are presented in a correlation matrix (Table 13) and graphically in Figure 12. The strongest correlation was found between AFC and AMH (r = 0.71, p < 0.0001).

Table 12. Characteristics of the included ORTs.

<table>
<thead>
<tr>
<th>ORT</th>
<th>Patients (n)</th>
<th>Lowest</th>
<th>Highest</th>
<th>Mean (SD)</th>
<th>Median</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFC (no.)</td>
<td>595c</td>
<td>3</td>
<td>70</td>
<td>17.8 (10.9)</td>
<td>15.0</td>
<td>Log-normal</td>
</tr>
<tr>
<td>AMH (ng/ml)</td>
<td>892</td>
<td>0.06</td>
<td>26.3</td>
<td>2.3 (2.5)</td>
<td>1.6</td>
<td>Log-normal</td>
</tr>
<tr>
<td>FSH (IU/l)</td>
<td>740b</td>
<td>1.0</td>
<td>28.0</td>
<td>7.3 (3.1)</td>
<td>6.75</td>
<td>Log-normal</td>
</tr>
<tr>
<td>LH (IU/l)</td>
<td>689a</td>
<td>0.2</td>
<td>18</td>
<td>5.0 (3.1)</td>
<td>4.5</td>
<td>Log-normal</td>
</tr>
<tr>
<td>MCL (days)b</td>
<td>829a</td>
<td>21</td>
<td>38</td>
<td>28.2 (2.3)</td>
<td>28.0</td>
<td>Normal</td>
</tr>
</tbody>
</table>

AFC = antral follicle count, AMH = anti-Müllerian hormone, FSH = follicle-stimulating hormone, IU = international units, LH = luteinising hormone, MCL = menstrual cycle length.
aDifferences in nobs due to missing data.
bCycles >40 days were excluded.

Table 13. Correlation coefficients (Pearson’s r; bold) with corresponding p-values and n patients included in the analyses.

<table>
<thead>
<tr>
<th>logAFC</th>
<th>logAMH</th>
<th>logFSH</th>
<th>logLH</th>
<th>MCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>r = 0.71</td>
<td>p&lt;0.0001</td>
<td>n = 595</td>
<td>n = 892</td>
<td></td>
</tr>
<tr>
<td>r = -0.34</td>
<td>p&lt;0.0001</td>
<td>n = 489a</td>
<td>n = 740</td>
<td></td>
</tr>
<tr>
<td>r = 0.14</td>
<td>p=0.003</td>
<td>n = 462</td>
<td>n = 689</td>
<td></td>
</tr>
<tr>
<td>r = 0.51</td>
<td>p&lt;0.0001</td>
<td>n = 572</td>
<td>n = 829</td>
<td></td>
</tr>
<tr>
<td>r = 0.46</td>
<td>p&lt;0.0001</td>
<td>n = 829</td>
<td>n = 692</td>
<td></td>
</tr>
<tr>
<td>r = -0.15</td>
<td>p&lt;0.0001</td>
<td>n = 692</td>
<td>n = 646</td>
<td></td>
</tr>
<tr>
<td>r = 0.16</td>
<td>p&lt;0.0001</td>
<td>n = 646</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Differences in n due to missing data or, for AFC, due to ultrasonographic criteria (see Methods).
AFC = antral follicle count, AMH = anti-Müllerian hormone, FSH = follicle-stimulating hormone, LH = luteinising hormone, MCL = menstrual cycle length.
Figure 12. Correlation plots of AFC vs. AMH, FSH, LH and MCL. AFC = antral follicle count, AMH = anti-Müllerian hormone, FSH = follicle-stimulating hormone, LH = luteinising hormone, MCL = menstrual cycle length.

Table 14 summarizes the results of the univariate and stepwise multivariate GEE analyses and presents the corresponding AUROCs where appropriate. (Six hundred and twenty cycles in 443 patients who had complete data on all ORTs were used for the multivariate GEE analyses.) Live birth rates were negatively associated with age and positively associated with both AMH levels and AFCs. In univariate analyses, female age and AMH were similarly predictive of live birth (AUROC 0.61). The model with the best power to predict live birth included female age and AMH (AUROC 0.64). The increase in the AUROC value from 0.61 to 0.64 by adding AMH to age (or age to AMH) was significant (p<0.05). There was no significant interaction between AMH and age.

The abilities of the ORTs to predict poor and excessive ovarian responses are also presented in Table 14. The best model for the prediction of a poor or excessive response included AFC, AMH and age (AUROC 0.89).
Table 14. Outcome of GEE analyses and AUROCs.

<table>
<thead>
<tr>
<th>ORT</th>
<th>n°</th>
<th>Odds ratio</th>
<th>95% CI</th>
<th>AUROC Live birth</th>
<th>AUROC poor response</th>
<th>P-value</th>
<th>AUROC excessive response</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCL</td>
<td>1151</td>
<td>0.08</td>
<td>0.08–0.07</td>
<td>NA</td>
<td>0.70</td>
<td>&lt;0.0001</td>
<td>0.66</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>MCL+ Age</td>
<td>1151</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.75</td>
<td>&lt;0.0001</td>
<td>0.74</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>FSH–LH</td>
<td>942</td>
<td>-0.15</td>
<td>0.15–0.98</td>
<td>NA</td>
<td>0.60</td>
<td>&lt;0.0001</td>
<td>0.63</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>FSH–LH+ Age</td>
<td>942</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.69</td>
<td>&lt;0.0001</td>
<td>0.72</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>LogAFC</td>
<td>830</td>
<td>1.64</td>
<td>1.22–2.12</td>
<td>0.58</td>
<td>0.85</td>
<td>&lt;0.0001</td>
<td>0.84</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>LogAFC+ Age</td>
<td>830</td>
<td>1.40</td>
<td>1.03–1.91</td>
<td>0.62</td>
<td>0.86</td>
<td>&lt;0.0001</td>
<td>0.85</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>logAMH</td>
<td>1230</td>
<td>1.46</td>
<td>1.26–1.71</td>
<td>0.61</td>
<td>0.86</td>
<td>&lt;0.0001</td>
<td>0.86</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age</td>
<td>1230</td>
<td>0.92</td>
<td>0.89–0.95</td>
<td>0.61</td>
<td>0.67</td>
<td>&lt;0.0001</td>
<td>0.69</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>logAMH+ Age</td>
<td>1230</td>
<td>1.35</td>
<td>1.15–1.57</td>
<td>0.64</td>
<td>0.87</td>
<td>&lt;0.0001</td>
<td>0.87</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>logAFC+ logAMH+ Age</td>
<td>759</td>
<td>-</td>
<td>-</td>
<td>0.89</td>
<td>&lt;0.0001</td>
<td>0.89</td>
<td>&lt;0.0001</td>
<td>0.03</td>
</tr>
</tbody>
</table>

AFC = antral follicle count, AMH = anti-Müllerian hormone, AUROC = area under the receiver operating characteristic curve, 95% CI = 95% confidence interval, FSH = follicle-stimulating hormone, LH = luteinising hormone, FSH–LH is a grouped variable (see Methods), MCL = menstrual cycle length, ORT = ovarian reserve test.

AUROC denotes the discriminative capacities of the different ORTs to predict live birth/started cycle, and poor and excessive ovarian responses/ovum pick-up (OPU).

Poor response: mean logOSI* -1 SD

Excessive response: mean logOSI* +1 SD

*OSI: ovarian sensitivity index (number of oocytes retrieved/given dose of FSH or hMG at time of stimulation).

Stepwise multivariate GEE analyses were carried out in connection with 631 cycles on which there was complete data. The outcome was then verified using the 1230 cycles (892 patients) with data on AMH (marked light grey); and for OSI (per OPU) 759 cycles with data on AMH+AFC (marked dark grey).

a) n denotes number of started treatments. Results for response are per OPU (nMCL = 1051, nFSH-LH = 836, nAFC = 759, nAMH = 1111.) Differences in n are due to missing data.
b) P-value for analyses re. AUROC poor response
c) P-value for analyses re. AUROC excessive response
d) n treatments per OPU with data on both AMH and AFC.
Oocyte quality?

Antral follicle counts (Paper III) and serum concentrations of AMH (Paper IV) had significant impacts on live birth rates when controlling for both age and oocyte yield. That is, when age and the numbers of oocytes retrieved were held constant, differences in AFCs and/or AMH levels were still associated with differences in treatment outcome. These findings were interpreted as meaning that AFC and AMH provide information not only on quantitative ovarian reserve, but also oocyte quality (using live births as a proxy of oocyte quality).

To see if these findings also hold true for MCL (Paper I) and FSH–LH combinations (Paper II), similar analyses were applied to them in their respective cohorts. In multivariate GEE models controlling for both female age and oocyte yield, both MCL and FSH–LH combinations were independently significant as regards live-birth rates. The odds ratios [with 95% CI] for a live birth/OPU adjusted for both oocyte yield and age were:

I 1.06 [1.01–1.21], p = 0.03, for MCL (both age and oocyte yield remained significant), and

II 1.47 [1.15–1.86], p = 0.0016, for FSH–LH combinations (age but not oocyte yield remained significant).

There was no interaction between MCL and age or MCL and oocyte yield, or between FSH–LH combinations and age or FSH–LH combinations and oocyte yield.

Thus, holding both age and oocyte yield constant, for each step upwards in (grouped) MCL or towards a more advantageous FSH–LH combination, live birth rates increased by 6% and 47%, respectively. Together with the previous findings, these figures indicate that the ORTs MCL and FSH–LH combinations give information on both oocyte quantity and oocyte quality, to some extent.
Discussion

Ovarian reserve, measured by means of ORTs, was strongly associated with pregnancy rates and live birth rates in IVF-ICSI, as well as with other treatment outcome variables, also when controlling for female age. Treatment success increased gradually with improving ovarian reserve. The associations were generally log-linear. Moreover, ORTs contribute with information not only on oocyte quantity but also on oocyte quality, an association previously not recognised.

Instead of aiming at delineating suitable cut-off levels for negative surrogate outcomes after IVF-ICSI for different ORTs, the general approach was to obtain graded information on ovarian reserve by evaluating (the continuous) associations with treatment outcome using large cohorts, and with pregnancy rates and live birth rates as primary outcomes. By including unselected patients with all kinds of ovarian characteristics we were able to assess the whole spectrum of ovarian reserve.

In the sub-fertile populations, ovarian reserve was log-normally distributed, reflecting a continuum of ovarian types from small ‘oligofollicular’ ovaries to large ovaries with PCO morphology. Importantly, the findings clearly show that polycystic ovaries seamlessly qualify as one extreme in this (log) Gaussian distribution. Thus, women with polycystic ovaries – who consistently had the best treatment outcomes, also after adjustment for age and BMI – should be included in studies on ART outcome and ovarian reserve.

All ORTs showed various degrees of intercorrelations. The strongest was found for AMH and AFC. Outright prediction models for live birth rates based on ORTs were generally modest, but clearly significant, and the strongest model included AMH and age. Prediction models for ovarian response, i.e. poor and excessive response, included also AFC and showed high prediction power, suggesting that ovarian response may serve as a proxy for ovarian reserve.

Methodological considerations and comments

To some extent, patients (treatments) included in the different studies are similar, as data have been collected over a long period of time. Any potential changes in treatment outcome as a result of general treatment improvements
over time have presumably influenced the study population equally, regardless of different treatment conditions between patients.

All studies were clinical studies, with their inherent weaknesses. Thus, they were not standardized for COH protocols, the patients’ BMI or infertility duration. Patients were included in an unselected manner regardless of their pregnancy history or cause of infertility, e.g. anovulation, tubal factor infertility, male factor infertility, etc. This may possibly be considered to be a drawback. However, there were no differences in pregnancy rates or live birth rates according to diagnosis (Paper IV), which is why confounding consequences of mixing infertility causes seem unlikely in this material.

The majority of studies on ART outcome and BMI point toward an increased risk of miscarriage among the obese (103-105), without a link to embryo quality (106), and also an increased miscarriage risk among underweight women (107). Bodyweights were checked at the time of OPU. There was no difference in BMI distribution between groups of patients (i.e. treatments; Tables 9–11), which is why there was no reason to adjust for BMI. Moreover, no patient had a BMI above 35 kg/m² and an effect of BMI on treatment outcome was absent among women with polycystic ovaries (Paper III).

In Study I we excluded patients considered to be chronically anovulatory, i.e. they were not included in the group of patients with the longest cycles. The same was done for the MCL variable in Study V, since the treatment prognosis for patients with anovulation differs greatly depending on the cause of anovulation. That is, the purpose was to avoid a possible adverse influence on the results by including hypogonadotrophic (WHO I) or hypergonadotrophic (WHO III) anovulatory women among those with long ovulatory cycles.

Statistics

Statistically, all included treatments and outcomes have been treated similarly; pregnancy rate and live birth rate were binary variables (yes/no) regardless of whether or not the embryo transfer was single or double, or if there was a singleton or a twin pregnancy at delivery. The different studies describe similar patterns in terms of outcomes, so the potential impacts of SET or DET, or fewer double embryo transfers as a result of use of the prediction model, are likely to be small.

Some women contributed with multiple treatment cycles, which is why the data were analysed by using generalized estimating equation (GEE) models (102). GEE models account for multiple treatments in the same subject and appropriately address error estimates in the context of correlated observations. Using GEE models allowed us to properly model all available cycles, instead of restricting analyses to the index treatment only. This was considered an advantage: the use of pregnancy rates and live birth rates as endpoints in repeated instead of single treatments probably reflects ovarian
reserve and female fertility (at treatment) more accurately (27). In addition, the great number of patients and treatments available provided great statistical power.

The principle for stratification differed between studies:

**Paper I**
MCL was grouped into six classes. There was a need to address within-woman variability in MCL. Figures on mean MCL were patient-reported and not based on patient records in a consistent way. We chose to categorise MCL rather than to analyse “exact” days, as this would be difficult or impossible for the patient to state as well as to depend on as a reliable figure. (Also, the rationale for stratification was raised as an issue at peer review, which is why MCL was re-analysed as a continuous variable controlling for age in GEE models and it remained highly significant as regards both PR and LBR/started cycle).

**Paper II**
No significant linear effects on treatment outcome were found for levels of FSH, LH or the LH:FSH ratio. Basal FSH concentration (36) has been recognized as an ORT for a long time and has generally been evaluated dichotomously; correspondingly, a cut-off value for FSH could be established. However, treatment outcome was superior among women with a high LH:FSH ratio, which motivated the development of a model incorporating LH as well. Four groups of ‘combined FSH and LH levels’ based on dichotomized FSH and LH levels could be established, as described in the Results section.

**Papers III and IV**
AFCs (Paper III) were grouped according to clinically useful pregnancy rates. This procedure deserves some attention. The intention was to facilitate a general interpretation of the information on AFC. If stratifying by quartiles, for example, the AFC strata would depend on the distribution of AFCs in the studied cohort. The cohort was not necessarily representative of a universal distribution among the infertile population, which could pose a possible risk of selection bias. Different clinics may also have different policies for approving patients for therapy and the association between AFC and treatment outcome may thus vary between different clinics and (distributions of) populations. By stratifying by treatment outcome we could also conclude that determined pregnancy intervals were associated with determined AFC ranges.

Although statistically reasonable, the approach of stratifying the independent variable according to the dependent variable was subject to criticism and therefore thoroughly discussed (but eventually accepted) at peer review.
After consideration, we thus used a more conventional statistical method in Study IV and stratified AMH levels by quartiles.

**Paper V**

The initial intention was to compare and evaluate the different ORTs relative to one another by applying the respective methods used in Papers I–IV. However, the different means of stratification made that approach illogical, particularly as regards AFC and AMH. AFC and AMH were therefore analysed as (log) continuous variables. The grouped variable ‘combined FSH–LH’ (Paper II), however, could not be handled differently than as categorial.

**Laboratory methods**

Although the study period for Paper II lasted almost a decade, the number of included observations is relatively small in comparison with those in Papers I and III. Patients who were referred for treatment only and did not undergo infertility work-up at the Carl von Linné Clinic had their serum FSH and LH levels analysed by local laboratories, using a diversity of methods. Even though the different laboratory methods were likely to be equivalent, we could not obtain confirmation that this was the case. To reduce the risk using incomparable values we chose to restrict the inclusion of observations to those with results only from the laboratory we employ.

The background regarding analysis of AMH levels (Paper IV) by way of a method no longer in use is described in the Methods section. The AMH levels given in this study are thus not equivalent to those that are obtained today, which is a disadvantage. Confirming studies are thus warranted to reproduce our findings and to define appropriate AMH cut-off levels using the (only currently available) Beckman Coulter Gen II assay.

**Ovarian reserve tests and treatment outcome**

In all papers, the transition from poor to well-preserved ovarian reserve was a continuum. There were no delineations of separate entities of patients. All ORTs demonstrated strong associations with pregnancy and live birth rates, also after adjustment for female age.

Menstrual cycle length (Paper I), log AFC (Paper III) and log AMH (Paper IV) were positively associated with treatment success, whereas basal LH levels were found to modulate the information provided by basal FSH with a high LH level mitigating the impact of a high FSH level (Paper II). Conversely, all ORTs, including after adjustment for age, were associated with poor treatment outcome among those with a reduced ovarian reserve. Analogous associations were found as regards ORTs and ovarian response.
The associations were generally log-linear (i.e. linear after log-transformation), most evidently for AFC (Paper III) and AMH (Paper IV). The log-linear relationships provide important clinical information and deserve to be emphasised: differences, unlogged – i.e. in absolute figures – in the low range of ovarian reserve had greater impact on treatment outcome than the corresponding differences in the higher range. In this context it must also be underlined that none of the studied ORTs delineated any cut-off level beyond which pregnancies did not occur. As for AMH, pregnancies were very scarce at immeasurable levels; as soon as AMH levels were quantifiable, PRs and LBRs increased log-linearly.

Menstrual cycle length (Paper I) has previously not been recognised as a marker of ovarian reserve. The age-dependent shortening of MCL is due to a shortened follicular phase, which is associated with decreasing levels of inhibin B and increases in those of FSH, whereas in general the length of the luteal phase is preserved (8, 9, 108, 109). As the length of the whole menstrual cycle thus mainly depends on the length of the follicular phase, MCL can be considered to be a surrogate measure of the length of the follicular phase. The association with treatment outcome is thus presumably primarily a follicular phase phenomenon. If it were possible to obtain a more reliable measure of the follicular phase per se, such a measure would probably reflect ovarian reserve better than MCL.

The positive association of MCL with favourable combinations of FSH and LH as well as with AFC and AMH (Tables 9–11) could be explained by ‘overlap of cycles’. It has been demonstrated that with decreasing cycle lengths, the follicular phase starts before the end of the preceding luteal phase (11), with a rise in FSH levels as a consequence of falling levels of inhibins, in turn due to few antral follicles. Accordingly, in these women AFC is low and so is the ovarian reserve. This overlap causes a ‘shift to the left’ of the coming follicular phase and could thus possibly contribute to an explanation of the gonadotrophin pattern among women with a reduced ovarian reserve, i.e. high FSH levels and low LH levels on cycle day 3 (Paper II).

Conversely, long cycle (follicular phase) lengths among those with a well-preserved ovarian reserve display the opposite: higher inhibin B levels in the early follicular phase may inhibit FSH secretion and lead to a delayed start of follicular growth, later selection of the dominant follicle and thus a longer follicular phase. A prolonged follicular phase, especially among women with polycystic ovaries, is also consistent with an increase in LH levels. High LH levels stimulate the theca cells to synthesize androgens, which may contribute to the arrest of follicular growth (110).

Measuring basal FSH levels prior to IVF-ICSI is widespread. A high level reflects a reduced ovarian reserve and is associated with poor IVF treatment.
outcome (36). We also found (Study II) that high basal FSH levels were associated with low pregnancy rates, whereas the opposite was the case for basal LH levels. However, ‘combined FSH and LH levels’ as a model was superior to both univariate gonadotrophin levels and the LH:FSH ratio. Low FSH levels combined with high LH levels were associated with the highest pregnancy and delivery rates and the reciprocal combination was associated with the poorest treatment outcomes. Thus, LH levels added significant information and modified the information given by assay of FSH alone. Mean levels of estradiol were low and did not differ between the FSH–LH groups, verifying early follicular phase in all groups at the time of venepuncture.

The influence of LH has been studied previously. Several investigators have reported a negative impact of a low LH:FSH ratio on treatment outcome (42-45, 47, 111), which is in accordance with the findings reported in Paper II, although at some variance. LH levels can be considered to be a modulator of the effect of any FSH level, high or low, as demonstrated by the different pregnancy rates in the different FSH–LH groups. Moreover, age-adjusted LH:FSH ratios were not associated with pregnancy rates, regardless of whether they were analysed dichotomously, continuously or stratified. The explanation for this may be the possibility of obtaining similar LH:FSH ratios in all FSH–LH groups except Group 4 (high FSH–low LH). For example, the LH:FSH ratio could be two in Groups 1, 2 and 3 (6/3=2, 4/2=2 and 14/7=2, respectively), whereas mean pregnancy rates differed between Group 1 and Groups 2 and 3. The LH:FSH ratio does not take into account the absolute values of FSH or LH levels. The results in Paper II thus suggest that it is more informative to regard FSH and LH in absolute figures grouped in different combinations (as defined by the plot, Figure 7) instead of as ratios with given cut-off levels. Additionally, ratios tend to increase statistical errors and are generally more difficult to interpret than absolute figures (112).

The impact of a favourable FSH–LH combination may be due to lower FSH and higher LH levels reflecting a larger number of antral follicles, the extreme cases being polycystic ovaries and PCOS (113). Women with polycystic ovaries and those with PCOS generally exhibit high LH:FSH ratios. Polycystic ovaries have a higher follicle density than ‘normal’ ovaries (21), and thus a superior ovarian reserve (49, 80) and they produce the highest levels of AMH ((57) and Paper IV).

Given that polycystic ovaries represent the high extreme of the ovarian reserve spectrum, small ‘oligofollicular’ ovaries may represent the other extreme. Oligofollicular ovaries contain few antral follicles and the serum gonadotrophin pattern exhibits higher FSH than LH levels in the early follicular phase (Papers II and III) and low levels of AMH (Paper IV). Accordingly, serum concentrations of LH and FSH (inversely) correlate with antral
follicle counts and AMH, over the entire range from oligofollicular ovaries to polycystic ovaries.

Women with polycystic ovaries showed the best treatment outcome, independently of ovulatory status and also when controlling for age and BMI (Paper III), suggesting that women with polycystic ovaries – and PCOS – should be included to cover the whole range of ovarian reserve to yield an evaluation as dynamic as possible in analysis of the associations with IVF-ICSI treatment outcome. Thus, ovarian reserve (in a population) should be regarded as a (log-) continuum from poor to well preserved, and the inclusion of polycystic ovaries is essential to describe the full normal range.

An interesting finding in Studies III and IV was that pregnancy rates levelled out above AFC 30 and AMH levels above 5 ng/ml. This is in line with a recent report on the association between LBR and oocyte yield (92). The reason for this effect is unclear. It may to some extent be iatrogenic, as (elective) single embryo transfers were more frequent in these groups (Tables 10 and 11). Indeed, the levelling off effect was absent and the association between AFC and LBR was strictly (log-) linear over the whole AFC range in preliminary findings from 2005 (114). Treatments included in those analyses were unaffected by the clinic’s prediction model; at that time DET was performed on a routine basis also among high responders. Moreover, endometrial receptivity may be impaired as a result of high estradiol levels (115), often reached among highly responding patients, with premature induction of progesterone receptors, resulting in an advanced endometrial condition (19).

Study V was aimed at evaluating the correlations between the different ORTs and determining which basal marker, or which combination of basal markers, is best for estimation of ovarian response and the chance of a live birth in IVF-ICSI. Most studies on ORTs have been focused on only one or two markers.

The two ORTs with less predictive power (Paper V) were MCL and the FSH–LH combinations, perhaps partly because of the smaller population used in this study compared with those studied in Papers I and II. Although both these markers were associated with live birth rates in large populations (Papers I and II), they are clearly less accurate than AMH and AFC. The lower prediction power for MCL is hardly surprising, given its dependence on the observational data provided by the patients. For FSH–LH, the lower accuracy may be due to their indirect relationship with the ovaries and the resting oocyte pool, being dependent on a complex pituitary feedback system. Previous studies have also shown an inferior predictive power as regards basal FSH compared with AFC (116), although FSH as part of a grouped variable together with LH has never been tested versus AMH or AFC before.
The AFC and serum AMH concentrations were strongly correlated and both were univariately associated with live births. However, in multivariate analysis with AMH and age, AFC was rendered insignificant. This could partly be explained by the reduced AMH secretion per follicle with age due to a reduced number of functioning granulosa cells per follicle, in turn also affecting oocyte quality (89). This specific link to oocyte quality cannot be captured by AFC and may thus contribute to making AFC the less sensitive ORT of the two. In the light of this, one would however expect an interaction between AFC and AMH with age, but no such interaction was observed.

However, AFC could be included together with AMH and age for optimal prediction of poor and excessive ovarian responses. It thus seems that AFC and AMH may reflect slightly different qualities of the pool of antral follicles and their ability to respond to various doses of exogenous FSH/hMG. Moreover, AFC has inherent disadvantages being a visual method and is dependent on consistency in counts, which may vary between investigators and contribute to making AFC less accurate as an ORT than AMH.

Early studies on ORTs were mainly focused on predicting ovarian responses to FSH/hMG stimulation, and they established that poor ORT results indicate a risk of poor response and risk of cycle cancellation (116-118). Our data confirm those findings and also show that the associations between ORTs and ovarian responses (i.e. OSIs) are continuous.

Ovarian sensitivity indexes (OSIs) exhibited a (log) Gaussian distribution in an unselected dataset of around 7500 consecutive IVF-ICSI treatments (101). Hence, defining normal, poor and excessive responses according to OSI is reasonably evidence-based as opposed to most conventional definitions of such subgroups, which commonly state only the number of retrieved oocytes with cut-off levels for high and low response based on varying arbitrary grounds. OSIs also take the given stimulation doses into account and they thus serve as a dynamic surrogate measure of ovarian integrity.

Predictions of low and high responses according to OSI data were both very powerful, with AUROC values unusually high (0.89; Paper V) as regards a relationship between basal biological markers and a stimulus-response effect. In addition, OSI has been shown to exhibit prediction levels for live birth rates of the same magnitude as found here in connection with AMH (101), suggesting that OSI and AMH may reflect a similar mechanism of importance for pregnancy and live birth.

The lower levels for discrimination of live birth rates, with the best model incorporating AMH and age, reflect the naturally more complex combination of multiple factors determining the ultimate success of assisted reproduction.
Discriminative capacity was assessed via AUROC (also known as the c-statistic). ROC curves are obtained by comparing the proportion of couples who achieve a pregnancy with those who do not, relative to all given probability thresholds. The AUROC thus expresses how well a model can, for example, (dichotomously) discriminate couples who will achieve a pregnancy from those who will not within a given time frame (119).

AUROCs must be interpreted relative to a figure of 0.5, which is the same as no discrimination (i.e. pure guessing). The univariate AUROC for age in connection with basal prediction of a live birth was 0.61 (Paper V). The univariate AUROC for logAMH was also 0.61, similar to previous findings for AMH (89), suggesting that (unadjusted) AMH and age have comparable impacts on treatment outcome. An AUROC value of 0.61 would thus indicate a relative increase in predictive power of 22% (relative to AUROC = 0.5; i.e. 0.61/0.5=1.22). Combining AMH and age significantly increased the AUROC to 0.64, thus improving discrimination by 27% on top of or relative to female age alone (0.61-0.5 = 0.11; 0.64-0.5 = 0.14; 0.14/0.11 = 1.27).

Do ovarian reserve tests predict oocyte quality?

Largely, opinion has been that ORTs give only quantitative information on the remaining oocyte pool and do not reflect oocyte quality (89, 120-123). This may partly depend on under-powering of studies, the predominance of limiting study inclusion to the index treatment only, and/or limiting the ovarian reserve spectrum by, for example, not including the group of women with PCOS. Some previous studies, however, have shown an association between serum AMH concentrations and live birth rates after assisted reproduction (77, 89-91), although further statistical analyses suggested this association could be mediated and explained by the relationship between AMH and (quantitative) oocyte yield. In this context, it has been argued that AMH cannot predict oocyte quality per se. Any positive influence of a well-preserved ovary would thus solely depend on the correlation with oocyte yields: the more eggs, the more embryos to choose from (93).

Transfer of a euploid embryo far from always results in a healthy pregnancy and a child, but the birth of a child is dependent on the transfer of an embryo derived from a healthy oocyte. Moreover, aneuploidy has very high specificity for non-pregnancy. Therefore, live birth rate may be used as a proxy of euploidy.

Controlling for both age and oocyte yield, both AFC (Paper III) and AMH (Paper IV) still had considerable impact on live birth rates. This association has previously not been recognised; in previous reports on this specific subject, all except one single study (91) have failed to find this possible information on oocyte quality provided by AMH. However, when the responsible author of that report later discussed their findings in a review pa-
per (as a co-author), the general conclusions were negative on this point (93). Corresponding analyses were also carried out in retrospect as regards MCL and combined FSH–LH, with the same result for these ORTs. It is reasonable to conclude (and difficult to interpret differently) that these results are consequences of the ORTs reflecting not only quantitative but to some extent also qualitative aspects of the oocytes. Further supporting this, a recent study revealed higher fecundability (the probability of a natural conception in a menstrual cycle) among women with high AMH levels compared with those with low AMH levels, also after adjusting for age (124). Moreover, in an insemination programme, groups of women with high serum AMH levels had significantly more pregnancies than women with low AMH levels when controlling for age (125). Nevertheless, we need to await thorough genetic analyses for a definite conclusion on this plausible relationship between ovarian reserve markers and oocyte quality.

Considering the above, it is likely that the proportion of good quality oocytes (subsequently embryos) is higher if ovarian reserve is well preserved than if it is not. If age and oocyte yield were similar in two women, the patient with a well-preserved ovarian reserve would thus have a greater likelihood of having a euploid oocyte retrieved than the woman with a low ovarian reserve. If ORTs therefore reflect the proportion of remaining normal oocytes in the oocyte pool, it is a logical finding that the number of oocytes has an impact per se as a greater yield increases the chance that at least one retrieved oocyte will be euploid. Accordingly, in patients with well-preserved ovarian reserve, mild stimulation (126) resulting in low-level oocyte recovery is often sufficient, as the proportion of euploid oocytes is likely to be high. At the other end of the AMH spectrum, both factors – i.e. reduced oocyte recovery and a low proportion of normal oocytes – compromise the chances of pregnancy.

**Do ovarian reserve tests predict pregnancy and live birth?**

Taken together, treatment success increased gradually with improving ORT results. Conversely, all ORTs showed poor results among those with a reduced ovarian reserve.

Several models for the prediction of IVF-ICSI treatment outcome have been presented since the mid 1990s. Only a few have concerned live birth rates (69, 91, 127). Variables such as age, infertility duration, various causes of infertility and pregnancy history have all been shown to be significant as regards treatment outcome; all of them, however, being inconsistent between studies with the exception of female age (91). Although female age is acknowledged to be the main limiting factor of IVF-ICSI treatment success, the association between live births and oocyte yield is also strong (92). In
recent years, it has been argued that the link between ORT results and live birth rates is mediated solely through this quantitative relationship (93).

It is often claimed that ORTs can predict neither oocyte quality nor pregnancy, since some women with extremely poor ovarian reserves become pregnant (regardless of age). This was also true in Studies I–IV; none of the studied ORTs delineated any cut-off level beyond which pregnancies did not occur. However, pregnancy is a dichotomous outcome whereas ORTs are continuous variables and may only predict chances of pregnancy and live birth at a group level.

The reasoning regarding the prognostic impact of female age on pregnancy is similar. In groups of women aged 30–35, the probability of pregnancy occurring (per month or year) is much higher than in groups of women of 40–45 years of age. Nevertheless, some older women do in fact become pregnant and give birth to a healthy child.

However, nobody would question the general predictive value, at a group level, of female age as regards pregnancy and live birth. Similarly, there are numerous reports of pregnancies in women with very poor ORT results (such as AMH levels below the detection limit). However, again at a group level, such cases are definitely fewer than in groups of women with normal ORT results. In statistical terms it is difficult to delineate a cut-off level for age over which the prediction of non-pregnancy is 100% (it would have to be set at a very high age level), in spite of the obvious basic importance of female age. The results of the present work suggest that the predictive values of ORTs are of similar nature, but they may add significant information to what is given by age itself.

The predictive capacity of a prognostic model can be assessed by calibration and discrimination (128). Discrimination (i.e. the use of AUROC), is dichotomous in its nature, dependent on the underlying probability distributions (119), and may not be optimal in assessing models that predict future risk (or chance) or stratifying individuals into risk (or chance) categories (128). As for calibration, in a well-calibrated model the predicted outcome is adequately coherent with the observed outcome and it is thus possible to correctly classify observations into clinically useful groups (119).

For individual IVF-ICSI treatments, the discriminative and predictive capacities of ORTs are limited, probably because of overlap of probability distributions between women who will conceive and those who will not (119), and because pregnancy and live birth after IVF-ICSI depend on multiple and complex factors. Moreover, cut-off levels applied to give ORTs high sensitivity must be set at such a level that the clinical value of the test is lost as a result of low specificity (27).

To some extent, however, ORTs may predict pregnancy and live birth. The best discriminative capacity for live births for an ORT model as given by AUROCs was 0.64 (Paper V), similar to findings in previous reports (89,
In individual cases a predictive power of that level is modest and of limited value, but may provide powerful calibrations of prediction models at a group level (130) and may thus serve as a basis for a reasonable chance estimate for an individual before treatment (119). Accordingly, at counselling before treatment, probable outcomes in groups of patients of similar ages and with comparable ovarian reserves will help to set expectations at a reasonable level, besides being an important aid in the design of COH regimens. Of the investigated ORTs, assay of serum AMH – together with age – provided the best basal estimate of the chance of a live birth.

Polycystic and oligofollicular ovaries

Hardly surprising, the strongest associations between the ORTs studied and treatment outcome were the two variables with a direct ovarian source, AFC and AMH. Our results clearly show a (log) Gaussian distribution of these two variables. One should be cautious to extrapolate the crude figures to the normal population; it is likely that the distribution of these variables is slightly different in a population of normal fecundity. However, there is little reason to dispute that the endowment of oocytes is normally distributed albeit the distribution in a subfertile population, as the present one, could contain more women with a reduced ovarian reserve (i.e. oligofollicular ovaries). On the other hand, our subfertile population is also likely to contain a high proportion of women with polycystic ovaries, especially those who are also anovulatory (PCOS). As these women are positioned in the other end of the ovarian reserve spectrum, the resulting distribution in the present population may be more spread to the extremes compared with a population with normal or proven fertility.

Thus, there is reason to discuss these two groups of ovarian types specifically. Women with polycystic ovaries and PCOS have often been excluded from studies of the present type, which has contributed to lack of spread in AFC and AMH levels. Given also small study populations, it may not be surprising that results have been often negative as regards the important endpoints pregnancy and live birth rates. Is there solid reason to exclude women with polycystic ovaries, or even those who are anovulatory and thus qualify for PCOS? Our data do not support that. On the contrary, both ovulating and anovulatory polycystic ovaries fit as the end of the Gaussian curve (Papers III and IV), clearly seamlessly, without indications of a qualitative or step-wise change in any of the measured variables of ovarian reserve.

There are now a number of reasons to see polycystic ovaries as a variant of the normal and that both the clinical and endocrine profiles in these women are directly linked to their increased ‘functional ovarian mass’, most likely determined in early life (possibly fetal stage) (21). Thus, although the threshold level for classifying an ovary as polycystic by transvaginal ultra-
sonography is defined as twelve antral follicles (79), recently suggested to be increased to >19 (131) (corresponding to AMH of >5 ng/ml, strikingly similar to the level we found for AFC 20 x 2; Paper V), both the risk for anovulation (Paper III) and hyperandrogenism (132) increase with increasing number of antral follicles (or AMH) above this cut-off level. Consequently, it seems reasonable that AMH – as a single blood test – with time will be considered sufficient for a diagnosis of polycystic ovaries and, combined with signs of anovulation, sufficient for a diagnosis of PCOS.

With age, as the ovarian ‘functional mass’ decreases, the phenotypic expression of polycystic ovaries is attenuated, with an increase in ovulations in those who were previously anovulatory, and with a shortening of previous long menstrual cycles. In parallel, women with normal ovaries also experience shortened follicular phases with increasing age. Moreover, surgical reduction of ovarian volume after wedge resection or ovarian drilling in women with PCOS also results in restitution of ovulatory cycles. Thus, reduction of the ovarian ‘functional mass’, either by age or by surgery, will result in polycystic ovaries becoming morphologically – and functionally – more ‘normal’.

Our results show that the high LH:FSH ratio, ‘typical’ for PCOS, also exhibits a seamless increase in parallel with AFC and AMH, again arguing against a qualitative difference between normal and polycystic ovaries.

Interestingly, women with the oligofollicular ovarian type with a low AFC and low AMH levels exhibit the opposite clinical and hormonal profile to that of women with polycystic ovaries and PCOS, with a low LH:FSH ratio and short menstrual cycles. Again, differences to normal ovaries were seamless. It seems likely that with time, but at different ages depending on the initial endowment of oocytes, all ovaries become oligofollicular with the clinical and endocrine profile typical of that ovarian type.

The ORTs and corresponding ‘ovarian types’ are illustrated in Figure 13. In summary, the results of this work support that the basis for the variation in ovarian reserve should be sought in the variation in ovarian morphology at any given age, with polycystic ovaries as one extreme and small oligofollicular ovaries as the other extreme at the opposite end of the ovarian reserve spectrum (Figure 13).
Figure 13. Arbitrary illustration showing the (log) Gaussian distribution of ovarian morphological types, from ‘oligofollicular’ ovaries to the left, to polycystic ovaries to the right, with the corresponding findings of cycle length (Paper I), FSH–LH (Paper II), AFC (Paper III), AMH (Paper IV) and treatment outcome. Adapted from Holte, J. in Revelli, A., Tur-Kaspa, I., Holte, J., and Massobrio, M. Biotechnology of Human Reproduction, Parthenon Publishing, 2003.

Limitations

All ORTs have inherent weaknesses. Study I was based on patient-stated menstrual cycle lengths. In individual cases there is a risk of error in recalling the exact number of cycle days, which is why there is a potential error in relying on self-reported cycle lengths instead of menstrual diaries. However, consistency of recall error would cancel out the significance and any such error is thus unlikely. The variable MCL is also subject to intra-individual variation of menstrual intervals. Factors other than the amount of
antral follicles and the length of time from menstrual onset to ovulation may affect cycle length, such as function of the corpus luteum and hence the duration of the luteal phase. The clinical utility of MCL also has another limitation, since anovulatory women cannot be included. In Study I, extreme cases of anovulation were excluded from the analyses because of unknown causes of anovulation. Most of them belonged to the normogonadotrophic (WHO II) group, with high AMH levels, high AFCs and with a favourable FSH–LH combination – indeed, patients with generally favourable treatment conditions. Hence, if the reason for anovulation had been exclusively WHO II, those patients could have been included in the group with the longest cycles.

Serum levels of gonadotrophins could be influenced by a non-functioning pituitary feedback mechanism, as discussed above, or the presence of receptor polymorphism. Patients with receptor anomalies may present with elevated basal FSH levels (40). Despite the possibility of a well-preserved oocyte pool in these patients, doses used at the time of COH may vary depending on receptor genotype, with a demand for higher doses in women with certain receptor variants (133). Basal gonadotrophin levels in these cases would thus more profoundly reflect ovarian responsiveness rather than the actual oocyte pool. By influencing COH response and maybe also pregnancy outcome (134), the potential (and unknown) occurrence of FSH receptor polymorphism might have influenced the results to some extent. Such an influence, however, is likely to be marginal (135).

Although AFC has a great clinical advantage in its simplicity and non-invasive approach, it is dependent on the technical qualities of the ultrasonographic equipment, the skill of the ultrasonographer and various problems in visualising the ovaries as a result of anatomic variations and obesity, for example. It has been demonstrated that the risk of inter-observer variation is higher when AFC is high (136). As treatment outcome levelled out in the high AFC range (Paper III) such inter-observer variation is likely to be of minor significance. Moreover, to minimize possible inter-observer variability, study inclusion was restricted to patients scanned by only two (Paper III) or three (Papers IV and V) experienced investigators. We did not, however, evaluate the consistency of AFC assessment between investigators, which is a shortcoming.

Unlike AFC, serum AMH concentrations have the advantage of being investigator-independent. The low intra- and inter-cycle variability of AMH levels is well recognized (59-61), but has nevertheless recently been questioned (137). The literature on AMH is also confusing, since different laboratory kits have been used (100) and there are no acknowledged conversion factors between methods. The only commercially available AMH kit since 2011 is the Beckman Coulter Gen II ELISA, which has been reported to provide higher values compared with the (former) DSL kit (138) used by ourselves (Papers IV and V). However, a recent report indicates that the Gen
II ELISA kit is also subject to questionable reliability (139). Hence, one must appreciate the laboratory uncertainties and interpret AMH data with some caution. The strong correlation between AMH levels and AFCs does, however, indicate that AMH is at least comparable with AFC. Conformity between the two markers in individual cases would thus indicate a reliable clinical assessment of ovarian reserve.

Ovarian stimulations were conducted with individual doses of FSH or hMG. This might have influenced the results, as the doses were decided upon partly according to ORT results and the patients’ anticipated responses. On the other hand, giving all patients the same dose would not have been clinically reasonable and would probably also have resulted in attenuation of the associations between the investigated variables and treatment outcome. Moreover, OSIs exhibit prediction levels for live birth rates comparable to those obtained via assay of AMH (101). The OSI takes the given dose into account. Given the close relationship between OSI and AMH ((99) and Paper V), it seems unlikely that adapting the dosage at the time of COH according to ORT data would have significantly affected the results.

Most women with the highest AMH levels met the criteria of polycystic ovaries (>11 antral follicles per ovary) and many of them also PCOS (a/oligoamenorrhea) (79) (Table 11). However, we did not test our study cohort for hyperandrogenism, nor were clinical signs of hyperandrogenism recorded. Thus, we have not studied PCOS per se and our conclusions cannot be extrapolated to the very few – if any – patients who could qualify for a PCOS diagnosis based on hyperandrogenism and anovulation but without polycystic ovaries. This was beyond the scope of the present studies. The single feature of PCOS of interest here is the fact that polycystic ovaries exhibit the greatest number of antral follicles. Nevertheless, it could be of interest to investigate whether varying androgen profiles could affect the general positive outcome for this group of patients. Positive associations between levels of AMH and androgens in women with PCOS (140) and normogonadotrophic anovulatory women (WHO II) (141) have, however, been reported. AFC and AMH have also been found to be more specific markers of PCOS than serum androgen levels (132). Thus, it seems likely that women with PCOS, due to a well-preserved ovarian reserve on the whole, are a group with good chances in IVF-ICSI treatment, regardless of means of PCOS diagnosis. This assumption is strongly supported by vast clinical evidence that PCOS without polycystic ovaries (i.e. a high number of antral follicles) is an extremely rare finding, and when found, is possibly often the result of an ultrasonographic misdiagnosis.

We did not record ethnicity, which is why we could not control for this possible confounder. There might be an effect on treatment outcome with regard to ethnicity, since, for example, lower success rates among black Afro-
American women compared with white women have been reported (142). Ethnicity-dependent variations in AMH levels have also been reported (143). The vast majority of patients included in the present studies were Caucasian white, which is why ethnicity is unlikely to have significantly affected the results.
General conclusions

- Pregnancy rates and live birth rates after IVF-ICSI, as well as other treatment outcome variables, were strongly associated with ovarian reserve in a non-dichotomous fashion. For all ORTs, including after adjustment for female age, treatment success gradually increased with improving ORT results. Pregnancy and live birth rates decreased with age, but were partly compensated for by a well-preserved ovarian reserve.

- The associations were generally log-linear. Thus, differences (in absolute figures) among those with poor ovarian reserve have greater impact on treatment outcome than corresponding differences among those with a well-preserved ovarian reserve.

- There were no separate entities of patients and the transition from patients with poor to those with good treatment outcome was clearly seamlessly continuous. In groups of patients, ovarian reserve should be regarded as a gradual continuum where women with polycystic ovaries – and those with PCOS – are the group with the best ovarian reserve and the best chances as regards IVF-ICSI treatment.

- ORTs provide information not only on oocyte quantity, but also on oocyte quality, a relationship previously not recognised. A marker indicating that ovarian reserve is well preserved probably reflects a high proportion of euploid oocytes.

- All ORT results correlated significantly with each other. The strongest correlation was between AFCs and serum AMH concentrations.

- ORTs are more accurate in discriminating ovarian responses than they are in discriminating live births. The best model for the prediction of ovarian response is AMH level and AFC together with age.

- Apart from chronological age, an objective estimate and specific information on ovarian reserve is a prerequisite for adequate individualisation of COH and is also of importance for basal prediction at counselling before IVF-ICSI treatment.
The discriminative and predictive capacity of ORTs as regards live birth is limited, but they provide reasonable probability estimates before treatment. Of the investigated ORTs, AMH – together with age – provides the best model for estimating the chance of a live birth, where AMH and age seem to contribute with predictive powers of comparable strength.
Sammanfattning på svenska


Ovarialreserv kan inte mätas exakt men kan skattas med hjälp av olika markörer. Det finns flera s.k. ovarialreservtestar beskrivna såsom mätning av gonadotropiner (styrhormonerna FSH och LH från hypofysen). FSH och LH mäts basalt, d.v.s. tidigt i menencykeln på cykeldag 2, 3 el 4. Andra vanliga tester är antimüllerskt hormon (AMH) vilket mäts i blodprov, samt antral-follikelräkning vilket innebär skattning av antalet s.k. antrala folliklar (äggblåsor) med vaginalt ultraljud (eng. antral follicle count, AFC).

De flesta tidigare studier angående ovarialreserv och kopplingen till utfall vid infertilitetsbehandling har utformats för att studera negativa kvantitativa samband, som till exempel att identifiera gränsvärden för ett dåligt eller uteblivet svaret på hormonestimulering och/eller risken för att behandlingscykeln kan komma att avbrytas. På så vis har de olika ovarialreservtesterna fr.a. kommit att betraktas som hjälpmedel vid individualisering av hormonestimulering, samt för att förutsäga svaret på stimulering. Även om ovarialreservmätt teoretiskt också skulle kunna bidra till att prognosticera chans till graviditet och barn efter IVF-ICSI, har majoriteten av studier inte förmått visa det på ett övertygande sätt varför den kliniska nyttan av skattning av ovarialreserv har ifrågasatts.

I mitt projekt har jag studerat sambanden mellan olika ovarialreservtestar och behandlingsresultat vid IVF-ICSI: menstruationscykeln längd (studie I), basala nivåer av FSH, LH och fr.a. kombinationer av dem (studie II), AFC (studie III), AMH (studie IV) samt de olika metodernas inbördes relation (studie V). Primära utfall var graviditetsfrekvens och levande födda efter behandling. Sekundära utfall var andra behandlingsparametrar såsom antal erhållna oocyter och givna doser vid hormonestimulering.
Alla studier var prospektiva Kohortstudier som inkluderade ett stort antal oselekterade patienter som genomgick en eller flera IVF-ICSI-behandlingar. För att korrigera för eventuellt beroende mellan flera behandlingar hos samma kvinna gjordes de viktigaste statistiska analyserna med s.k. ”generalized estimating equation” (GEE) modeller. Samma patient kunde därigenom inkluderas vid olika tidpunkter och den strategin möjliggjorde inkluision av alla tillgängliga behandlingscyklar. Analyserna justerades för kvinnlig ålder.

Syftet med avhandlingsarbetet var att studera olika metoder för skattning av ovarialreserv, respektive metods individuella betydelse, deras inbördes relationer respektive deras prognostiska förmåga för en framgångsrik IVF-ICSI behandling.

**Studie I:**
Studien omfattade 3228 kvinnor som genomgick 6271 IVF-ICSI behandlingar. Genomsnittlig menstruationscykellängd förkortades med två dagar med stigande ålder (p_trend <0.0001). Sambandet mellan cykellängd och graviditets- och förlossningsfrekvens var positivt och linjärt, även efter åldersjustering (p_trend <0.0001). Chansen att få barn efter IVF-ICSI var nästan fördublad för kvinnor med en menscykellängd >34 dagar jämfört med kvinnor med cykellängd <26 dagar. Det förelåg också ett motsvarande positivt samband mellan menscykellängd och embryo kvalitet respektive svaret på hormonestimulering.

**Studie II:**
Nivåerna av FSH och LH mättes hos 745 kvinnor som genomgick 1328 behandlingar. Av fördelningsskäl log-transformerades värdena (naturlig logaritm). Kombinationer av FSH- och LH-värden gav signifikant bättre information än kvoten mellan FSH och LH, respektive FSH och LH var för sig. Högsta genomsnittliga graviditetsfrekvensen (39%) noterades om lågt FSH (<6.7 U/l) kombinerades med högt LH (>4.9 U/l), medan den var lägst (22%) om högt FSH (>6.7 U/l) kombinerades med lågt LH (<4.9 U/l), och intermediär (27-28%) om FSH och LH båda var låga eller höga (p_trend=0.0004). Motsvarande samband förelåg med förlossningsfrekvens, stimuleringssvar och embryo kvalitet. Bästa modellen för prediktion av graviditet utgjordes av ålder tillsammans med grupperade värden på FSH och LH (i fyra kombinationer enl. gränsvärdena ovan; prediktiv kapacitet utifrån analys med s.k. Area under ROC-kurva, AUROC, med värde på 0.63).

**Studie III:**
Studien omfattade 2092 kvinnor som genomgick 4308 behandlingar. AFC varierade mellan 3 och 80 och var log-normalt fördelade med ett medelvärde (SD) på 19.2 (11.7). AFC var log-linjärt och positivt associerat med både graviditets- och förlossningsfrekvens som dock planade ut vid AFC över 30. Behandlingsutfallet var bäst bland kvinnor med s.k. polycystiska äggstockar,
oberoende av om de ovulerade eller ej. Det var också signifikanta samband mellan AFC och övrigt stimuleringsutfall.

I en multivariat GEE-modell med justering för både ålder och mängden erhållna ooeuter förblev AFC högsignifikant för både gravidiets- och förlossningsfrekvens, vilket talar för att AFC innehåller information även om den kvalitativa sidan av ovarialreserv.

**Studie IV:**
I studien ingick 892 kvinnor som genomgick 1230 IVF-ICSI-cykler. AMH-värdena var log-normallt fördelade (liksom AFC, studie III) med ett medelvärde (SD) på 2,3 (2,5) ng/ml. Levande födda per påbörjad cykel (% [95% konfidensintervall]) ökade log-linjärt från 10,7% [7,2-14,1] för AMH <0,84 ng/ml (25: e percentilen) till 30,8% [25,7-36,0] för AMH >2,94 ng/ml (75: e percentilen, p
trend <0.0001). Även i denna population var förlössningsfrekvensen högst bland kvinnor med polycystiska äggstockar. Motsvarande samband med övriga behandlingsutfall, respektive utfallet av analysen avseende äggkvalitet, fanns för AMH på samma vis som för AFC.

**Studie V:**
Studien baserades på samma population som studie IV. MCL, FSH och LH, kombinationer av FSH och LH, samt AFC och AMH undersöktes för inbördes korrelationer samt samband med levande födda och stimuleringsvar i både univariata och multivariata statistiska modeller. Den starkaste korrelationen förelåg mellan AFC och AMH (r = 0.71, p <0.0001). I univariata analyser var ålder, AMH och AFC associerade med antal levande födda/startad behandling. I multivariat analys förblev dock enbart ålder och AMH oberoende signifikanta (båda p <0.001). Denna modell, d.v.s. ålder tillsammans med AMH, var också den som bäst predicerade barn (AUROC 0.64); univariat var AUROC 0.61 för både AMH och ålder.

Bästa modellen för prediktion av dåligt respektive (alltför) kraftigt stimuleringsvar inkluderade ålder, AMH och AFC (AUROC 0.89).

**Slutsatser:**
Ovarialreserv är starkt relaterat till gravidiets- och förlossningsutfall men också till andra behandlingsvariabler vid IVF-ICSI. Alla skattningstekoder, även åldersjusterade, visade successivt ökande behandlingsutsiker med tilltagna ovarialreserv. Ålder tillsammans med AMH och AFC predicerade äggstockarnas svar på hormonstimulering med stor noggrannhet.

Sambanden var generellt log-linjära, vilket innebär att en skillnad hos kvinnor med nedsatt ovarialreserv hade större inverkan på behandlingsutfallet än motsvarande skillnad bland kvinnor med bra ovarialreserv.

Alla markörerna visade sig inrymma information om såväl den kvantitativa som den kvalitativa sidan av ovarialreserv, vilket tidigare inte visats på ett övertygande sätt. Sannolikt innebär detta att om ett test påvisar bra ovari-
alreserv, är proportionen euploida (kromosomalt normala; kvalitativa) oocytter större än då testet visar det omvända.

Ovarialreserv (i den subfertila populationen) är log-normalt fördelad utan klara skiljelinjer mellan nedsatt, bra och väl bevarad ovarianreserv, d.v.s. övergången inom gruppen från dålig till bra är ett kontinuum och kvinnor med polycystiska äggstockar utgör den grupp som har bäst bevarad ovarianreserv.

Ovarialreservmarkörer har på individnivå begränsad prediktiv kapacitet för chansen till graviditet och barn efter IVF-ICSI. Utifrån graviditets- och förlössningsfrekvens på gruppnivå kan dock information om ovarianreserv användas vid rådgivning för skattning av chansen även i det enskilda fallet. Av de undersökta markörerna utgör AMH tillsammans med åldern det bästa underlaget för sådan rådgivning, där AMH och ålder tycks bidra med prediktiv information av jämförbar styrka.
Acknowledgements

This work was carried out at Carl von Linné Kliniken, Uppsala, and the Department of Women’s and Children’s Health, Uppsala University, Sweden. I would like to thank each and everyone who helped, supported and encouraged me and who has contributed to the completion of this thesis. To those whom I owe especially much, I express my sincere gratitude in my mother tongue:


Alla ni sköterskor, barnmorskor, laboratoriepersonal och sekreterare på Carl von Linné Kliniken. Ni är alla guld värda! Varje dag gör ni till något bra fast det ofta är snärjigt och stressigt. Det är en glädje att få arbeta med er!

Ove Axelsson. Professor. Du välkomnade mig så tidigt att det sedan tog så lång tid att du hann sluta innan jag kom till skott och blev registrerad som doktorand på KBH.


Johan Botling. Studiekamrat. Alla timmarna… allt nötande… alla bra planer… alla tentorna…

Alla ni kollegor, sköterskor, barnmorskor och all personal på KK, Mälersjukhuset, för alla år av kliniskt arbete, er vänskap, för att ni är ni och för att ni gjorde mig till den jag var innan jag slutade och sedan blev den jag är. Jag vill särskilt nämna:

Stefan Hammarbäck. Tack för att du anställde mig och för allt du lärde mig under åren tillsammans på KK.

Bo Möller. Tack för din välvilliga entusiasm, uppmuntran och för att du befordrade mig fast jag inte ville.

Olof Birck. Utan dig hade jag nog inte blivit anställd på KK. Vem vet var jag hade hamnat då?

Ann Blom. Uroterapeut och ”religionsinstiftare”. Tack för alla roliga samtaler samt bra samarbete under åren på KK. Vi möts i poolen?

Chatarina ”Chatta” Andersson. Spindeln i nätet på fertilitetsmottagningen MSE. Tack för all hjälp under åren på KK.


Våra fina barn Johanna och Max för alla skratt och för att ni aldrig slutar att utmana och ifrågasätta. Ni betyder allt, ni är livsinnehåll, glädje och ger perspektiv på tillvaron. Lev livet medan det pågår!

The work within this thesis was supported by Carl von Linné Kliniken, Uppsala; the Department of Women’s and Children’s Health, Uppsala University; the Department of Obstetrics and Gynaecology, Mälarsjukhuset, Eskilstuna; FoU-centrum, Landstinget Sörmland; and by grants from the Family Planning Fund in Uppsala, Uppsala University, Uppsala, Sweden.
References

13. Danielsson KG 2000 [The interplay between embryo and endometrium is important for successful implantation. Increased knowledge can result in new contraceptive methods and better treatment of infertility]. Lakartidningen 97:1660-1662, 1665-1666
31. Richardson SJ 1993 The biological basis of the menopause. Baillieres Clin Endocrinol Metab 7:1-16
47. Liu KE, Greenblatt EM 2008 Elevated day 3 follicle-stimulating hormone/luteinizing hormone ratio >or= 2 is associated with higher rates of cancellation in in vitro fertilization-embryo transfer cycles. Fertil Steril 90:297-301
52. Chang MY, Chiang CH, Hsieh TT, Soong YK, Hsu KH 1998 Use of the antral follicle count to predict the outcome of assisted reproductive technologies. Fertil Steril 69:505-510
57. Piltonen T, Morin-Papunen L, Koivunen R, Perheentupa A, Ruokonen A, Tapanainen JS 2005 Serum anti-Mullerian hormone levels remain high until late reproductive age and decrease during metformin therapy in women with polycystic ovary syndrome. Hum Reprod 20:1820-1826

60. La Marca A, Stabile G, Artenisio AC, Volpe A 2006 Serum anti-Mullerian hormone throughout the human menstrual cycle. Hum Reprod 21:3103-3107


62. Li HW, Wong CY, Yeung WS, Ho PC, Ng EH 2011 Serum anti-mullerian hormone level is not altered in women using hormonal contraceptives. Contraception 83:582-585


64. Kallio S, Puurunen J, Ruokonen A, Vaskivuo T, Piltonen T, Tapanainen JS 2012 Antimullerian hormone levels decrease in women using combined contraception independently of administration route. Fertil Steril


71. Lawson R, El-Toukhy T, Kassab A, Taylor A, Braude P, Parsons J, Seed P 2003 Poor response to ovulation induction is a stronger predictor of
early menopause than elevated basal FSH: a life table analysis. Hum Reprod 18:527-533
72. El-Toukhy T, Khalaf Y, Hart R, Taylor A, Braude P 2002 Young age does not protect against the adverse effects of reduced ovarian reserve--an eight year study. Hum Reprod 17:1519-1524
80. Nikolaou D, Gilling-Smith C 2004 Early ovarian ageing: are women with polycystic ovaries protected? Hum Reprod 19:2175-2179
85. Erdem M, Erdem A, Gursoy R, Biberoglu K 2004 Comparison of basal and clomiphene citrate induced FSH and inhibin B, ovarian volume and
84

antral follicle counts as ovarian reserve tests and predictors of poor ovarian response in IVF. J Assist Reprod Genet 21:37-45

86. **Hendriks DJ, Broekmans FJ, Bancsi LF, de Jong FH, Looman CW, Te Velde ER** 2005 Repeated clomiphene citrate challenge testing in the prediction of outcome in IVF: a comparison with basal markers for ovarian reserve. Hum Reprod 20:163-169

87. **Broer SL, Mol BW, Hendriks D, Broekmans FJ** 2009 The role of antimullerian hormone in prediction of outcome after IVF: comparison with the antral follicle count. Fertil Steril 91:705-714


97. **Holte J, Bergh T, Tilley J, Pettersson H, Berglund L** 2004 The construction and application of a prediction model to minimize twin implantation rate at a preserved high pregnancy rate. In: ESHRE. Berlin


100. Nelson SM, La Marca A 2011 The journey from the old to the new AMH assay: how to avoid getting lost in the values. Reprod Biomed Online 23:411-420


114. **Holte J, Hadziosmanovic N, Bergh T, Berglund L** 2005 Ovarian reserve is directly proportional to antral follicle counts over the entire range, from oligo follicular to polycystic ovaries: results of a prospective study in 2568 IVF/ICSI cycles. Abstract. In: ESHRE. Copenhagen


121. **Smeenk JM, Sweep FC, Zielhuis GA, Kremer JA, Thomas CM, Braat DD** 2007 Antimullerian hormone predicts ovarian responsiveness, but not embryo quality or pregnancy, after in vitro fertilization or intracytoplasmic sperm injection. Fertil Steril 87:223-226


123. **Melo MA, Garrido N, Alvarez C, Bellver J, Meseguer M, Pellicer A, Remohi J** 2009 Antral follicle count (AFC) can be used in the prediction of ovarian response but cannot predict the oocyte/embryo quality or the in vitro fertilization outcome in an egg donation program. Fertil Steril 91:148-156

125. Li HW, Yeung WS, Lau EY, Ho PC, Ng EH 2010 Evaluating the performance of serum antimullerian hormone concentration in predicting the live birth rate of controlled ovarian stimulation and intrauterine insemination. Fertil Steril 94:2177-2181


A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine.