Assessment of current and future ovarian reserve status

Simone Broer

Assessment of current and future ovarian reserve status

Thesis, Utrecht University, The Netherlands, with a summary in Dutch.
Proefschrift, Universiteit Utrecht, met een samenvatting in het Nederlands.

ISBN: 978-94-610-8158-2
Author: Simone Broer
Cover design: Jeroen de Lange, YellowCliff Media
Lay-out: Roy Sanders
Printed by: Gildeprint Drukkerijen B.V. Enschede, The Netherlands

The author gratefully acknowledges financial support for printing this thesis by: Division of Woman and Baby University Medical Center Utrecht, Beckman Coulter, BMA BV (Mosos), Ferring BV, Goodlife, Gynotec BV, JE Jurriaanse Stichting, Medical Dynamics, Memidis Pharma BV, Merck Serono, Merck Sharp & Dohme BV, Origio Benelux BV, Pfizer BV and Will Pharma.

Assessment of current and future ovarian reserve status

Beoordeling van de huidige en toekomstige ovariële reserve status
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 14 juni 2011 des middags te 4.15 uur

door

Simone Louise Broer
geboren op 3 juli 1985 te Naarden

Promotoren:	Prof.dr. F.J.M. Broekmans Prof.dr. B.C.J.M. Fauser
Co-promotor:	Dr. M.J.C. Eijkemans

Aan mijn ouders

Contents

Chapter 10

Chapter 1

General introduction

Ovarian ageing and ovarian reserve

The term *ovarian ageing* represents the age related decline of the quantity and quality of the oocytes residing within the follicles present in the ovarian cortex. Each woman receives an endowment of oocytes during fetal development. At the fourth month of fetal development, the ovaries contain some 6-7 million oocytes surrounded by a layer of flat granulosa cells to form the primordial follicle pool (1-3). Due to a rapid loss of the great majority of the primordial follicles via apoptosis in the second half of fetal life, at birth only 1-2 million primordial follicles remain (4). After birth, this high rate of follicle loss slows down somewhat, so that at menarche (first menstrual period) at least 300,000 to 400,000 primordial follicles remain (2;5). During the reproductive years the gradual decline will continue. Not only does the quantity of the follicles and oocytes decline, but oocyte quality also demonstrates changes with increasing female age. This becomes apparent in increased aneuploidy rates, responsible for the increased miscarriage rates and increased rates of infertility observed at older age (6). Underlying mechanisms may involve differences between germ cells at the time they are formed during fetal life, accumulated damage of oocytes in the course of a woman's life, or age-related changes in the quality of the granulosa cells surrounding the oocyte (7).

The changes in quantity and quality will lead to four milestones in the reproductive lifespan; the first two will easily occur unnoticed: the onset of decreasing fertility and the subsequent loss of natural fertility (capacity of creating a viable ongoing pregnancy leading to the birth of a child). As the decline continues and follicle numbers fall below a critical threshold of a few thousand, the menstrual cycle pattern becomes irregular, marking the onset of a life period referred to as the menopausal transition (8). Finally, when fewer than 1,000 follicles are left, the final cessation of menses (menopause) will occur (9-11) (*Figure 1*). By definition, menopause can only be assessed retrospectively, as a period of amenorrhea of at least 12 months after the final menstruation must have been surpassed (12). The normal process of ovarian ageing varies considerably among women. This implies that some women remain highly fertile until the fifth decade of life, whereas others already face the loss of natural fertility in their mid-thirties. This variation also becomes evident from the large variation in age at menopause. Menopause occurs at a mean age of 51 years (13), with a range of variation between 40 and 60 years (13-19). Despite the variation with age, it is hypothesized that a fixed temporal relationship is present among the four reproductive events, with the occurrence of the end of natural fertility some 10 years before menopause (7) (*Figure 2*).

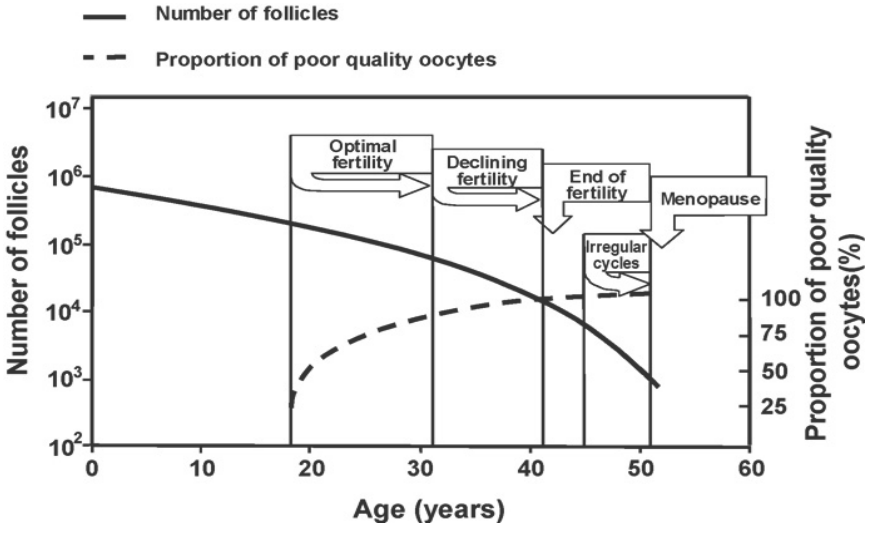


Figure 1. Schematic representation of ovarian ageing
Schematic representation of the number of primordial follicles present in the ovaries and the chromosomal quality of oocytes in relation to female age and corresponding reproductive events. Graph was drawn after Hansen et al. (10), and de Bruin et al. (237).

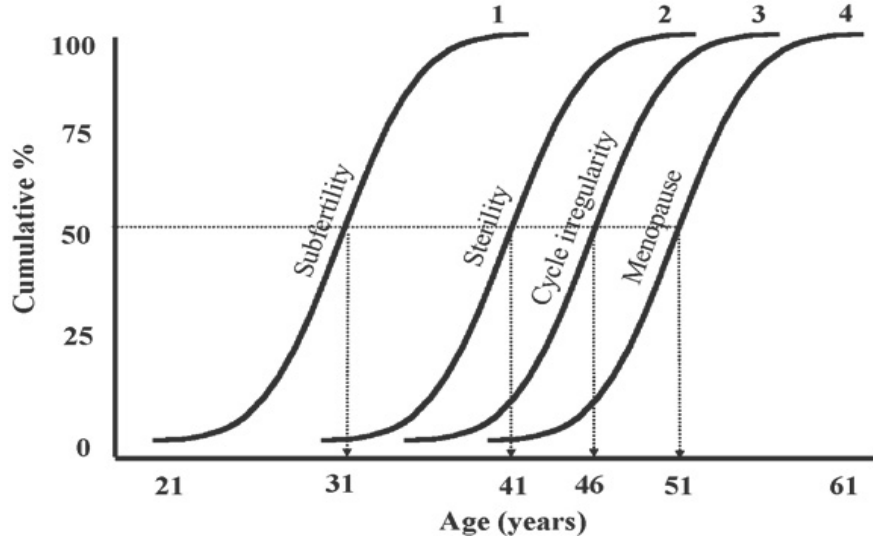


Figure 2. Age variation and stages of female reproductive ageing
Schematic representation of the age variations of the various stages of female reproductive ageing, depicted in a cumulative fashion. A fixed time interval between the subsequent cycles is assumed. Redrawn after te Velde and Pearson (7). Mean ages for the events are depicted on the x-axis.

Evidence for this hypothesis mainly stems from cross sectional observations (20), while longitudinal data establishing such relationship for individuals are scarce (21). Several reports have indicated that early loss of natural fecundity, as evidenced by repeated poor response to ovarian hyperstimulation for in vitro fertilization (IVF), leads to early occurrence of the menopausal transition (22-24). Also, the length of the time period of cycle irregularity preceding menopause has appeared to be independent of the age at which menopause subsequently is established (7;25). The relationship between age at onset of cycle irregularity and age at menopause has further corroborated the 'fixed interval' hypothesis (21).

Timing of the menopausal transition and menopause predominantly relate to the number of follicles present in the ovaries at a given time. Because of the individual variability, the development of tests that correctly forecast these follicle numbers, also called *the ovarian reserve*, has been a field of interest in the last decade (7;20).

The purpose of ovarian reserve testing: assessment of current and future fertility

Due to the availability of contraceptive methods, together with a growing economical wealth, which provided the opportunity for women to increase their level of education and to participate in the labor force (26), there is a general tendency to postpone childbearing in the Western societies. This has distinct implications; a growing proportion of women attempting to conceive will fail in achieving this goal within a time frame of 12 months, a condition referred to as female infertility. Each year approximately 15,000 young women in their twenties will be destined to have a much shorter reproductive lifespan than they imagine due to genetic and other uncontrollable factors and some 15% of all women will be sterile before the age of 40 years. Since women do not know the precise age at which their fertility begins to decline, it makes it difficult to modify behavior and the safeguarding of fertility. Consequently, these women and their partners will heavily depend on Assisted Reproduction Technology (ART) in order to achieve pregnancy. Assisted reproduction technology, such as IVF, will provide solutions for a mere 50% of these women, since ART will only be able to compensate for the decreased natural fertility to a limited extent (27;28). Delayed childbearing has also greatly contributed to the reduced number of children born per woman. With increasing longevity, population composition will change dramatically towards a clear preponderance of elderly people (29;30), and these changes will have immense social and economic implications.

Furthermore, in the Netherlands today, 16,000 IVF treatments are carried out with estimated annual costs of 35 million euros. Such expenditure may well be greatly reduced by preventing age related infertility through preventive management. This may also positively affect the number of children born per woman and thereby help to counteract the trend towards an increasingly aged population. Therefore, there is an urgent need for the development of tools that can assess the ovarian reserve and thus mirror current and future fertility status.

Assessment of current fertility

Current fertility assessment relates to predicting the chances for live birth in natural exposure or in infertility treatment conditions. Outcomes of interest in infertility treatment mainly relate to IVF or introcytoplasmatic sperm injection (ICSI) treatment, such as the response to ovarian hyperstimulation and the chances of ongoing pregnancy. There is a specific urge to identify women of relatively young age with clearly diminished reserve, as well as older women with still an adequate ovarian reserve. Based on the test result, management could be individualized, for instance by stimulation-dose adjustment, by counseling against initiation of IVF treatment, or by indicating the necessity of early initiation of treatment before the ovarian reserve has diminished too far. Individualization of patient management could be more cost-effective, as it could increase the efficacy and reduce the costs of the fertility treatment. So far many studies have been conducted on ovarian reserve tests (ORTs) in IVF/ICSI outcome, but they show contradictory results for both response prediction and pregnancy prediction.

Conventional meta-analysis should be performed to summarize the available evidence. But in conventional meta-analysis the major problem, which is heterogeneity between studies based on differences in the patient populations, stimulation protocols, hormone assays and other variables, remains an issue. It also became evident that the clinical value was dependent on the consequences related to the test result. Moreover, female age is also most frequently omitted as contributor in multivariable models. It has therefore not been assessed properly to what extent these tests add value to age and other patient characteristics. Also, it remains to be unraveled whether these tests perform the same in different subgroups. Conventional meta-analysis does not have the ability to address these issues properly; an individual patient data meta-analysis however, could do so. This will allow for the assessment of the true value of these tests. The true clinical value however, will depend on the effect on clinical patient management. Will it solely be used for counseling or will it have consequences like adaptations in the stimulation protocol or even the refusal of treatment?

Assessment of future fertility

The time interval until natural sterility will have set in, referred to as the reproductive lifespan, will mirror the period in which fertility may be optimal. For the assessment of this reproductive lifespan, tools that closely relate to the future age at menopause may be developed into useful long term predictors. Seen the fixed temporal interrelationship between end of fertility and menopause, correct prediction of menopause may provide valuable information on the individual level. This could open new avenues for the primary prevention of female infertility. Moreover, menopausal age is also related to women's health in general (31). Predicted early menopause could emphasize the need for timely prevention of bone demineralization, and cardiovascular and neurological disease (32-34), while the prediction of late menopause would open options for preventive management of breast and intestinal cancer (35). So far, the time relationship between ovarian reserve tests and menopause has been shown in cross-sectional studies and short term follow up studies (36-38). Long term follow up studies will allow insight into the feasibility of future fertility forecasting at those stages of life, where relevant decisions on preventive management are realistic.

Ovarian Reserve Tests

Actual direct measurement of the primordial follicle pool is impossible, but it has been shown that the number of the antral follicles in the ovaries is proportionally related to the size of the primordial follicle stock from which they were recruited (39). A marker correctly reflecting the number of antral follicles is therefore potentially suitable for the prediction of ovarian senescence. Current candidate markers for such purpose are early follicular Follicle Stimulating Hormone (FSH) concentration (40), the Antral Follicle Count (AFC) as measured by transvaginal ultrasound (41), and Anti-Müllerian Hormone (AMH) levels (42).

Follicle Stimulating Hormone

FSH is an indirect marker of the primordial follicle pool. FSH levels increase with advancing age, by a reduction in the release of inhibin B, and possibly also estradiol, thereby reducing the negative feedback on FSH release form the pituitary (43). High FSH levels therefore indirectly represent a small antral and primordial follicle cohort size.

Antral Follicle Count

Ovarian follicles of at least 2 mm in diameter can be detected using ultrasonography. Follicles in the size range 2-5 or 2-10 mm correlate well with the size of the primordial follicle pool. Antral follicles that are >2 mm in diameter are

highly responsive to gonadotrophins, but some in this size range may be in the early stages of atresia. The AFC entails the count of the follicles of 2-5 or 2-10 mm. The AFC has to be carried out in the early follicular phase of the cycle, although variation of counts across the cycle may be very modest (44). The AFC correlates well with the number of primordial follicles (45).

Anti-Müllerian Hormone

Anti-Müllerian Hormone has been identified as a dimeric glycoprotein and a member of the Transforming Growth Factor Beta (TGFß) family of growth and differentiation factors (46). The gene encoding for AMH is located on chromosome 19p13.3. Until recently, AMH was predominantly known for its role in male sexual differentiation (47;48). AMH is produced by Sertoli cells at the time of testicular differentiation and induces regression of the Müllerian ducts. In the ovaries of female fetuses, AMH can first be detected at 32 weeks of gestation (49). The absent production of AMH from primitive granulosa cells in the early stages of female fetal development will allow the Müllerian ducts to develop into the uterus, fallopian tubes and the upper part of the vagina (50;51). AMH has an inhibitory role in the initial recruitment and thereby aids in regulating the number of follicles remaining in the primordial follicle pool. Secondly, AMH has an inhibitory effect on follicular sensitivity to FSH and could therefore play a role in the process of dominant follicle selection (52) (*Figure 3*).

Serum AMH is produced from the cohort of ultrasonically visible antral follicles up to 7 mm, but follicles below the sensitivity limits of ultrasonography may also contribute to serum levels. This is based on the observation that AMH serum levels do not fall to zero when FSH sensitive antral follicles (2-5 mm) are stimulated into larger, dominant follicles during ovarian hyperstimulation for IVF and interrupt their AMH production (53) (*Figure 4*).

The independence of menstrual cycle stage (54-56) and the proportional relationship with the primordial and antral follicle cohort (45;53;57) make AMH a good candidate for assessment of ovarian reserve status in the female.

Other ovarian reserve tests

There are several other ovarian reserve tests known, such as Inhibin B, basal estradiol, CCCT (Clomiphene Citrate Challenge Test), GAST (gonadotrophin releasing hormone agonist stimulation test), EFORT (exogenous FSH ORT) and OVVOL (ovarian volume). These test are inferior to the ovarian reserve tests FSH, AFC and AMH because of a lower accuracy, too high false positive rate or the complexity of the test (58).

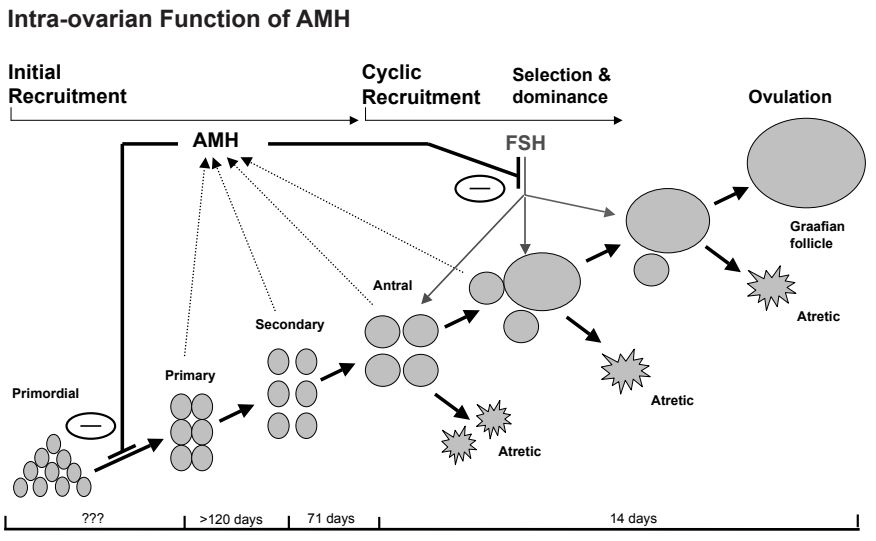


Figure 3. Intra-ovarian function of AMH in folliculogenesis
First, AMH has an inhibitory role in the initial recruitment and thereby aids in regulating the number of follicles remaining in the primordial follicle pool. Second, AMH has an inhibitory effect on follicular sensitivity to FSH and could therefore play a role in the process of dominant follicle selection.
AMH = Anti-Müllerian Hormone; FSH = Follicle Stimulating Hormone

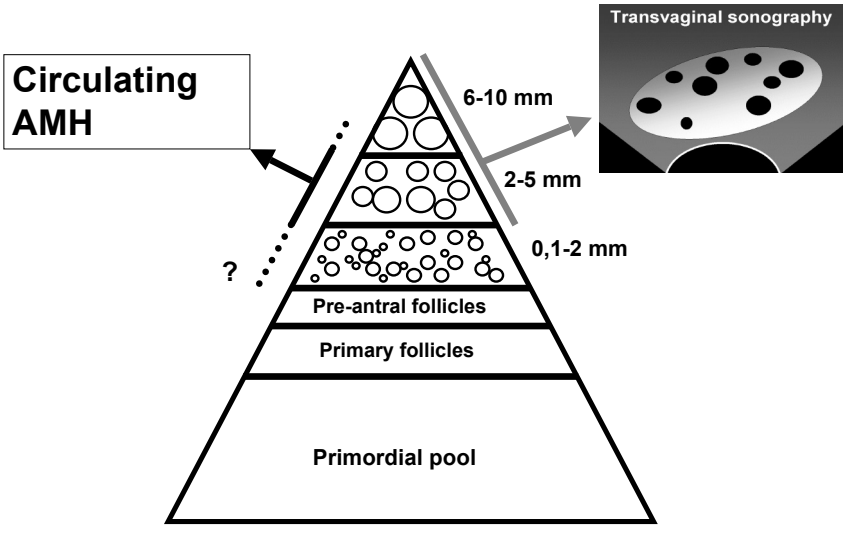


Figure 4. The source of AMH that enters the blood circulation.
Serum AMH is produced from the cohort of ultrasonically visible antral follicles up to 7 mm. Moreover, follicles below the sensitivity limits of ultrasonography may also contribute to serum levels. This is based on the observation that serum AMH levels do not fall to zero when FSH-sensitive antral follicles (2-5mm) are stimulated into larger, dominant follicles during ovarian hyperstimulation for IVF and interrupt their AMH production. The black line and dots represent the stages of antral follicles that contribute to serum AMH. The grey line represents the ultrasonically visible antral follicles.
AMH = Anti-Müllerian Hormone.

Aims and outline of the thesis

The studies presented in this thesis focus on further assessment of the real value of ovarian reserve tests (ORTs) in predicting current fertility status in ART treatment. Also, extended follow up in normal female volunteers was used to demonstrate the value of these tests in forecasting menopause.

The aims of the work can be listed as follows:

1. Study the accuracy of ovarian reserve tests in the prediction of outcome in ART
2. Study the added accuracy of ovarian reserve tests in the prediction of outcome in ART, when baseline patient characteristics such as female age are taken into account
3. Assess the added value of ovarian reserve tests in the prediction of age at menopause

Outline

Chapter 2 gives a systematic overview and meta-analysis of the existing literature on AFC and AMH in the prediction of poor response and/or ongoing pregnancy.

Chapter 3 describes the results of an individual patient data (IPD) meta-analysis (IMPORT study) which studied the added value of ORTs on patient characteristics and the value of multivariable prediction models in the prediction of a poor response and ongoing pregnancy.

Chapter 4 describes the results of an IPD meta-analysis (IMPORT study) which assessed the accuracy of the ORTs in the prediction of a poor response and ongoing pregnancy in several clinical subgroups, defined by age, BMI and duration of subfertility.

Chapter 5 gives a systematic overview and meta-analysis of the literature on AFC and AMH in the prediction of an excessive response.

Chapter 6 describes the results of an IPD meta-analysis (EXPORT study) which studied the added value of ORTs on patient characteristics and the value of multivariable prediction models in the prediction of an excessive response. The accuracy of the ORTs in the prediction of an excessive response was also studied in several clinical subgroups defined by age, BMI and duration of subfertility.

Chapter 7 studies the accuracy of ORTs in menopause prediction. A long term follow-up study was conducted in healthy normo-ovulatory women to asses the relationship of ORTs and age at menopause.

Chapter 8 summarizes the results of the studies presented and discusses the implications for clinical practice and future research.

Chapter 2

The role of Anti-Müllerian Hormone in the prediction of outcome after IVF: comparison with the Antral Follicle Count

S.L. Broer, B.W. Mol, D. Hendriks, F.J.M. Broekmans

Fertil Steril. 2009;91(3):705-14

Abstract

Objective

To assess the value of Anti-Müllerian Hormone (AMH) as a test to predict poor ovarian response and pregnancy occurrence after IVF and compare it to the performance of the Antral Follicle Count (AFC).

Design

A systematic review of existing literature and meta-analysis were carried out. After a comprehensive search, studies were included if 2x2 tables for outcome poor response and pregnancy in IVF patients in relation to AMH or the AFC could be constructed.

Setting

Academic referral center for tertiary care.

Patients

Cases indicated for IVF.

Interventions

None.

Main outcome measures

Poor response and non pregnancy after IVF.

Results

A total of 13 studies were detected reporting on AMH and 17 studies on AFC. Due to heterogeneity among studies, calculation of a summary point estimate for sensitivity and specificity was not possible. However, for both tests summary receiver operating characteristic (ROC) curves for the outcome measures poor response and non-pregnancy could be estimated and compared. The curves for the prediction of poor response indicated no significant difference between the performance for AMH and the AFC (P = 0.73). For the prediction of non-pregnancy, a poor performance for both AMH and AFC (P = 0.67) was found.

Conclusions

In this meta-analysis it was illustrated that AMH has at least the same level of accuracy and clinical value for the prediction of poor response and non-pregnancy as the AFC.

Introduction

Increasing female educational levels and participation in the labor force has resulted in a clear rise in the mean age at which women deliver their first child in Western-style societies (26). As natural fertility starts to decline after the age of 30, many women therefore will be faced with unexpected problems in becoming pregnant due to decreased ovarian reserve (59-61). It has been shown that the rate of the ovarian reserve decline varies considerably between individual women, making it a challenge to design tests that estimate an individual's remaining reproductive lifespan at a given age (61).

Ovarian reserve relates to both the quantity and quality of the ovarian follicle pool. The number of primordial follicles that are left in the ovary at a given age is therefore an important indicator for ovarian reserve and dictates reproductive events such as the age at menopause (61). Although direct measurement of the primordial follicle pool is impossible, it has been shown that the number of antral follicles in the ovaries is proportionally related to the size of primordial follicle stock from which they were recruited (39). Therefore, the Antral Follicle Count (AFC) is believed to represent the quantitative aspect of ovarian ageing (58). Unfortunately, markers that may directly reflect oocyte quality are clearly lacking at the moment. Consequently, the age related decrease in fertility cannot be determined through a direct test. Only through measurement of the quantity of the oocytes can information on the quality aspects of ovarian reserve be obtained (62;63).

Ovarian response to ovarian hyperstimulation in IVF is another way in which the quantitative ovarian reserve may come to expression. Although poor response may be considered a sign of diminished ovarian reserve it may also be caused by other factors, such as underdosing in obesity or in certain FSH receptor polymorphisms (64). Assessment of the true nature of a poor ovarian response may help to direct the management of the patient (62;65;66). Additionally, correct identification of poor responders, especially in older patients before entering an in vitro fertilization program is important, as this could help in proper management regarding gonadotrophin dosing and denial of treatment. For this purpose, the tests of choice are currently the AFC or basal FSH as was shown in a comparative review (67).

Anti-Müllerian Hormone (AMH), member of transforming growth factor-β family, is produced in the granulosa cells (68). The highest level of AMH expression is present in granulosa cells of secondary, preantral and small antral follicles up to 6 mm in diameter (69), while in follicles growing into dominance this expression ceases (53;70). AMH is barely detectable at birth and reaches the high-

est values after puberty, then decreases progressively with age and becomes undetectable at menopause (71;72). Serum AMH levels have been shown to strongly correlate with the number of antral follicles (73;74), and have appeared to be cycle independent (54;56). From several studies AMH has emerged as a predictor of ovarian response to hyperstimulation (75;76) and possibly even of the chance of becoming pregnant after IVF (77).
The aim of the present systematic review is to assess the true accuracy of AMH as a prognostic factor for the outcome of IVF/ICSI treatment in comparison with the Antral Follicle Count, which was shown the best predictor of poor response after IVF (67).

Methods

In the present review, studies were enrolled that addressed the evaluation of the AFC and AMH as predictors of the outcome poor response and pregnancy after IVF or ICSI treatment. No preset criteria for the definition of poor ovarian response or pregnancy were used. Poor ovarian response definition included cycle cancellation, number of dominant follicles at ultrasound or oocytes at retrieval below a certain threshold or combinations of the above. Pregnancy definition included both clinical and ongoing pregnancy. Also, any cut off or set of cut offs for an abnormal test result were included in the review and analysis.
A systematic search of MEDLINE was carried out using the keywords 'in vitro fertilization' or 'in vitro fertilisation', or 'assisted' or 'intracytoplasmatic' or 'intracytoplasmic', in combination with 'Anti-Müllerian Hormone' or 'müllerian inhibiting substance' or 'müllerian inhibiting factor'. A period including all years until December 2006 was covered by the search. The abstracts of all the studies identified were read by one researcher (SB). Any article that could possibly be of value for the association between AMH and the IVF outcomes poor ovarian response or pregnancy was pre-selected. In the next step, two researchers (SB and DH) carefully read and judged all pre-selected articles independently. If it was judged possible to construct 2x2 tables, where test result at a certain cut off was related to the outcome parameters poor response and/or pregnancy, the study was selected for final recordings and analysis. In the event of any disagreement between the two researchers, the opinion of a third author (FB) was final. The authors of studies that related test result to IVF outcome without the possibility to construct a 2x2 table were contacted by email and asked to provide the necessary data for the construction of such a table. If adequate data were obtained this way the study was added to the selection. In every selected study the reference list was scanned to identify studies that could possibly be included in the selection and then processed as described.

Each selected study was further scored by the researchers (SB and DH) on the following study quality characteristics: (1) sampling (consecutive versus other), (2) data collection (prospective versus retrospective), (3) study design (cohort study versus case-control study), (4) blinding (present or absent), (5) selection bias, (6) verification bias, (7) analysis on one or multiple cycles per couple, (8) stimulation (GnRH-agonist or GnRH-antagonist). Also, data on the cut off levels used were recorded, as well as the assay used for AMH measurement.
For the comparison of AMH and AFC we updated the recently published meta-analyses (63;67) on the performance of the AFC. The period to be covered by the systematic search for studies reporting on the AFC in the prediction of poor response and non-pregnancy after IVF was extended to December 2006. The same basic series of keywords was used as listed above, in combination with 'Antral Follicle Count' or 'antral follicle number'. If by this search new studies were found and judged suitable for processing according to the previously described procedure they were added to the already analyzed AFC studies. If a study on both AMH and the AFC was located by any of the search strategies, this study was used for both review groups.
As this review used only published data from the literature no approval from a institutional review board was required.

Analysis

First, for each study finally included, we calculated sensitivity and specificity from the 2x2 tables. Sensitivity-specificity points were plotted in a ROC-curve. Homogeneity of the sensitivity and specificity points was tested by means of the chi square test. A summary point estimate of sensitivity-specificity points and 95% confidence interval was calculated if homogeneity for both parameters could not be rejected. In case of heterogeneity for one or both parameters, logistic regression was used to evaluate whether the study characteristics were associated with the discriminatory capacity. If one of the study characteristics was found to have a statistically significant impact on the performance of the test, further analysis was performed in subgroups of patients. If not, it was explored whether the differences in sensitivity and specificity combinations were the result of the use of different threshold levels. For that purpose, a Spearman correlation coefficient was calculated for the association between sensitivity and specificity. In case of a negative correlation as defined by a correlation coefficient of -0,5 or less, a summary ROC-curve was estimated, using a random-effects regression model (78-80) and assuming that studies were heterogeneous because of the use of different threshold levels. The same procedures were followed for studies on AFC from the updated search.

Table 1. Characteristics of the 13 included studies for Anti-Müllerian Hormone

Author	Consecutive	One cycle per couple	Data per cycle	Cohort / case-control	Prospective/ Retrospective
Ebner 2006	Yes	Yes	Yes	Cohort	Prospective
Eldar-Geva 2005	Yes	Yes	Yes	Cohort	Prospective
Ficicioglu 2006	Yes	Yes	Yes	Cohort	Prospective
Fréour 2007	No	Yes	Yes	Cohort	Prospective
Kwee 2007	Yes	Yes	Yes	Cohort	Prospective
La Marca 2006	Yes	Yes	Yes	Cohort	Prospective
McIlveen 2006	No	Yes	Yes	Cohort	Prospective
Muttukrishna 2004	No	Yes	Yes	Cohort	Prospective
Muttukrishna 2005	Yes	Yes	Yes	Cohort	Retrospective
Penarrubia 2005	Yes	Yes	Yes	Case-control	Retrospective
Smeenk 2006	No	Yes	Yes	Cohort	Prospective
Tremellen 2005	Yes	Yes	Yes	Cohort	Prospective
Van Rooij 2002	Yes	Yes	Yes	Cohort	Prospective

BC = Beckman – Coulter, DSL = Diagnostic System Laboratories, NS = not stated

The constructed summary ROC curves for AMH and AFC were tested for statistically significant differences with a linear regression model, similar to the model used to evaluate the impact of study characteristics.

Results

Systematic review

The systematic MEDLINE search produced 742 hits, from which we selected 24 studies based on the abstract reading. We were able to create 2x2 tables from 9 studies. We contacted the authors from the remaining studies of whom four provided us with the necessary data to construct the 2x2 tables. Through this search and selection strategy, a final number of 13 studies reporting on the capacity of AMH to predict ovarian response and/or non-pregnancy after IVF and considered suitable for data extraction and meta-analysis were identified (74;81-92). Five studies reported on both poor response and pregnancy, 1 study on pregnancy alone and 7 studies on poor response alone. The characteristics of the included studies are listed in *Table 1*. From this table it was shown that all studies presented data for one cycle per couple and that the majority used a prospective cohort design. Also, definitions for poor response were quite uniform. However, selection bias was judged to be present in quite a number of studies.

Table 1. Continued

Blinding	Selection Bias	Verification Bias	Agonist / Antagonist	Definition Poor Response Prediction; Pregnancy	AMH-assay
No	Yes	Yes	Both	<4 oocytes; Clinical	BC
No	Yes	No	Agonist	Ongoing	BC
No	No	No	Agonist	<5 oocytes	DSL
No	Yes	No	NS	<6 oocytes	BC / DSL
No	Yes	No	Agonist	<6 oocytes	DSL
No	No	Yes	Agonist	<4 oocytes or cancellation; Ongoing	BC
No	Yes	No	Agonist	≤4 oocytes	BC
No	Yes	No	Both	<4 oocytes or cancellation; Ongoing	BC
No	No	No	Agonist	≤4 oocytes	BC
No	No	No	Agonist	Cancellation; Clinical	BC
Yes	Yes	No	Agonist	≤4 oocytes; Ongoing	BC
No	Yes	No	Agonist	≤4 oocytes	BC
No	Yes	No	Agonist	<4 oocytes	BC

For the AFC, the updated systematic search and selection revealed no additional studies eligible for analysis. Consequently, a total of 17 studies on AFC was available.

Accuracy of poor response prediction

Sensitivities and specificities for the prediction of poor ovarian response, as calculated from each study reporting on AMH, are summarized in *Table 2*. The sensitivity varied between 40% and 91%, and the specificity between 64% and 100%. Homogeneity for both sensitivity and specificity had to be rejected (p value for the χ^2-test for sensitivity and specificity 0.04 and 0.001, respectively). For this reason, the calculation of a single summary point estimate for sensitivity and specificity was not meaningful.

Logistic regression analysis showed that none of the study characteristics recorded had a statistically significant impact on the reported predictive performance of AMH. For example, whether the design of the study was retrospective or prospective, no influence was made on the prognostic capacity of AMH as estimated by the studies. A plot of sensitivity-specificity points in a ROC space is shown in *Figure 1*.

Table 2. Performance of Anti-Müllerian Hormone (AMH) in the prediction of a poor response and nonpregnancy in IVF patients and shift from pre-test to post test probability of poor response and nonpregnancy for patients with abnormal test result

Author	Cycles (n)	AMH cut-off value	Prediction Characteristics			Pretest probability (%)	Posttest probability (%)	Proportion with abnormal test (%)
			Sensitivity	Specificity	LR+			
Poor response								
Van Rooij 2002	119	0.10 ug/l	0.49	0.94	8.2	29	77	18
	119	0.20 ug/l	0.54	0.90	5.4	29	70	23
	119	0.30 ug/l	0.60	0.89	5.5	29	70	25
Muttukrishna 2004	69	0.10 ng/ml	0.76	0.88	6.3	25	68	28
Muttukrishna 2005	108	0.20 ng/ml	0.87	0.64	2.4	#	#	#
Penarrubia 2005	80	4.90 pmol/l	0.40	0.92	5.0	25	62	16
Ebner 2006	141	1.66 ng.ml	0.69	0.86	4.9	26	63	28
Ficicioglu 2006	44	0.25 pg/ml	0.91	0.91	10.1	25	77	30
La Marca 2006	48	0.50 ng/ml	0.85	0.82	4.7	25	63	33
	48	0.75 ng/ml	0.80	0.93	11.4	25	77	27
	84	1.25 ng/ml	0.58	0.75	2.3	57	76	44
Smeenk 2006	80	1.40 µg/l	0.62	0.73	2.3	16	31	33
Tremellen 2005	75	8.10 pmol/l	0.80	0.85	5.3	27	67	32
Fréour 2007	69	1.30 µg/l	0.44	1.00	∞	13	100	6
Kwee 2007	104	0.80 µg/l	0.55	0.94	9.2	27	75	19
	104	1.00 µg/l	0.66	0.94	11.0	27	78	22
	104	1.20 µg/l	0.69	0.88	5.8	27	68	27
	104	1.40 µg/l	0.76	0.86	5.4	27	66	31
	104	1.60 µg/l	0.79	0.78	3.6	27	56	38

Table 2. Continued

Pregnancy								
Van Rooij 2002	106	0.10 ng/ml	0.22	0.89	2.0	75	85	19
	106	0.20 ng/ml	0.27	0.85	1.8	75	84	24
	106	0.30 ng/ml	0.28	0.81	1.5	75	81	25
Eldar-Geva 2005	56	18 pmol/l	0.67	0.69	2.2	54	71	50
Penarrubia 2005	80	Not stated	0.62	0.55	1.4	66	73	56
Ebner 2006	132	1.66 ng/ml	0.19	0.69	0.6	51	39	25
Smeenk 2006	80	1.40 µg/l	0.38	0.73	1.4	50	58	33
Kwee 2007	104	1.40 µg/l	0.34	0.79	1.6	77	84	31

Note: if a study reported on multiple cut-off values, data for all cut-off values are shown. LR+= likelihood ratio for a positive test result. # not possible to calculate.

The Spearman correlation coefficient for sensitivity and specificity was –0.31 which was judged to be sufficient to estimate a summary ROC curve (*Figure 1*).

Accuracy of non pregnancy prediction

For the prediction of non-pregnancy, the sensitivities and specificities of each study are summarized in *Table 2*. Similar as for ovarian response, homogeneity for sensitivity had to be rejected. However, specificity appeared to be homogeneous (χ^2-test: p value 0.11). Sensitivity varied between 19% and 66%, whereas specificity varied between 55% and 89%. As for the estimation of one summary point for sensitivity and specificity, statistical homogeneity for both test parameters was required. Consequently, this solution was abandoned. A plot of sensitivity-specificity points in ROC space is shown in *Figure 1*.

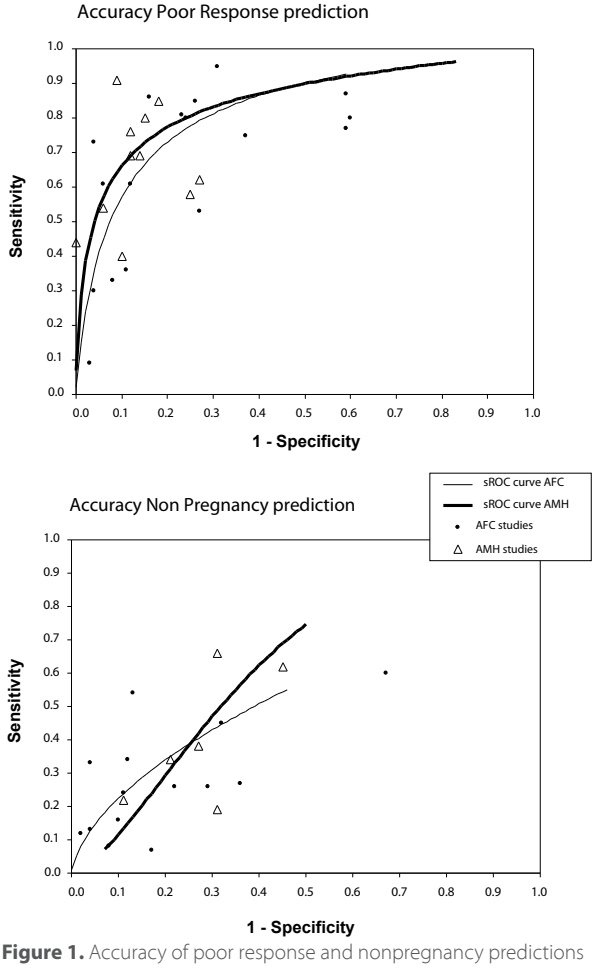


Figure 1. Accuracy of poor response and nonpregnancy predictions
AFC = Antral Follicle Count, AMH = Anti-Müllerian Hormone, ROC = Receiver Operating characteristic

The Spearman correlation coefficient for sensitivity and specificity was -0.71 for the prediction of non-pregnancy, which was judged to be sufficient to estimate a summary ROC curve (*Figure 1*).

Clinical value

Based on the summary ROC curves depicted in *Figure 1*, a range of positive likelihood ratios was calculated corresponding to various sens/spec points at this ROC curve. For each of these likelihood ratio values the pre-AMH test probability of poor response or non pregnancy (set at 20% and 80%, respectively) were converted into a post-AMH-test probability. *Table 3* depicts a series of likelihood ranges and the probability of obtaining an AMH test corresponding to this likelihood ratio range, as well as the post test probability of poor response and non pregnancy. At a maximum positive likelihood ratio of ~8, the post-AMH test probability of poor response will approximate 65%, if the pre-AMH-test probability is assumed to be as high as 20%. The probability of obtaining a test result for AMH with a likelihood ratio ~8 is high enough to consider the AMH as a clinically valuable test for poor response prediction.

For prediction of non pregnancy, the extremely low AMH cut off level that is necessary to obtain a moderate positive likelihood ratio of ~5, leading to a post test pregnancy rate of less than 5% based on a pre test rate of 20%, occurs only in an extremely limited number of patients (*Table 3*). The summary ROC curve runs not far from the line of equality indicating that most of the ROC curve is uninformative (likelihood ratio ~1).

Comparison of AMH to AFC

In the analysis of the 17 available studies, sensitivity and specificity for AFC in the prediction of poor response and non-pregnancy both showed heterogeneity. After excluding the necessity of subgroup analysis from the study characteristic analysis, the Spearman correlation coefficients between sensitivity and specificity for both poor response and non-pregnancy were judged to be sufficient to estimate summary ROC curves (-0.63 and -0.67 respectively). These curves are drawn in *Figure 1*, and show quite a high accuracy for the prediction of poor ovarian response, but very limited accuracy for non-pregnancy prediction.

Comparison of the estimated summary ROC curves for the prediction of poor response showed no significant improvement in the performance for AMH compared to the AFC (p value 0.73). The overall accuracy for predicting non-pregnancy was poor for both tests. There was no significant difference between the ROC curves for the prediction of non-pregnancy between both tests (p value: 0.67).

Clinical value as outlined in *Table 3* indicates a slightly better performance for basal AMH compared to the AFC. Especially the course of the ROC curve along the y-axis suggests that many cases of poor response can be identified with only a limited number of false positives. If more false positives are accepted sensitivity can amount up to 70% with only a false positive rate of 10% and this test performance will imply a realistic number of abnormal tests.

Table 3. Occurrence of both Ant-Müllerian Hormone (AMH) and Antral Follice Count (AFC) results within a specified likelihood ratio (LR) range and the concomitant posttest probabilities of poor response and nonpregnancy, given a prevalence of poor response of 20% and nonpregnancy of 80%.

Prediction of poor response (pre-test probability = 20%)			
LR range	Occurrence of test results in this range (%)		Posttest probability of poor response (%)
	AMH	AFC	
0-1	66	68	<20
1-2	7	10	20-33
2-3	5	4	33-43
3-4	7	6	43-50
4-5	1	0	50-56
5-6	1	0	56-60
6-7	0	0	60-64
7-8	0	0	64-67
>8	13	12	>67
Prediction of non-pregnancy (pre-test probability = 80%)			
0-1	75	77	<80
1-2	15	16	80-89
2-3	6	5	89-93
3-4	1	0	93-94
4-5	3	2	94-95
5-6	0	0	95-96
6-7	0	0	96-96.5
7-8	0	0	96.5-97
>8	0	0	>97

For a high level of LR (i.e., ~8) the probability of producing a poor response is 70%. The chance of obtaining such a test result at the cut-off level for AMH used would be ~13%. At the same high level of positive LR the chance of not becoming pregnant is ~97%. The probability of measuring AMH at that low cut-off, however, is close to zero.

Discussion

Main findings

This meta-analysis summarizes the available evidence on the accuracy of AMH in the prediction of poor ovarian response after stimulation for IVF in comparison to the AFC. The ROC curves do not suggest a clearly better predictive ability for AMH compared to the AFC, and indeed the difference was not significant statistically (p=0.73). This implies that the best poor response predictor to date, the AFC (67), has obtained company from a test that may have some crucial advantages. Application of this test does not need to be carried out on a specific day of the cycle, as AMH levels have shown to fluctuate only marginally and prediction by samples of any cycle day will be equally accurate (56;88;93). Blood sampling often is part of preparation for IVF treatment and therefore extra vena puncturing will not be necessary. Currently, the availability of the AMH assay may present some problems but surely this test system soon will become part of one of the large automated platforms, with inherent validity checks and limited assay variation. In contrast, the AFC necessitates skilled ultrasound operators who carefully identify, measure and count ovarian follicles. Although observer bias may be limited technically (94;95), a new source of bias may arise from the fact that the ultrasound operator is aware of the cut off for test judgment and may become influenced by the consequences of the test for the treatment of the couple. Such test inflation has recently been suggested from a study in older IVF patients who were allowed or refused IVF treatment on the basis of this test (66). Also, the AFC is to be carried out in the early follicular phase of the cycle, although variation of counts across the cycle may also be very modest (96).

The performance for non-pregnancy prediction is clearly poor for both AMH and the AFC. This comes as no surprise as AMH, like the AFC, is strongly believed to only represent the size of the cohort of FSH-sensitive follicles continuously present in the ovaries. Response to ovarian hyperstimulation will be directly linked to this cohort size (97). The relation between quantity and oocyte and embryo quality is much less clear. Indeed, the chance of pregnancy after IVF depends on many more factors than the cohort size alone, like embryo quality, transfer technique and endometrial receptivity (98). Also, over the past decades, not a single ovarian reserve test has been evaluated in a series of subsequent IVF cycles. It is likely that only by studying several consecutive cycles a true representation of a female's remaining reproductive capacity will be obtained. Only one study has demonstrated a certain predictive value regarding the occurrence of pregnancy in a selected group of cases with normal FSH and

AFC levels (77). This study, however, could not be included in the present meta-analysis due to the lack of data to produce a contingency table.

Limitations

Our meta-analysis has possible weaknesses. First, across studies poor response has been defined in various ways. Most definitions are based upon outcome parameters of the IVF treatment, such as cycle cancellation for absent or very limited follicle growth, the number of oocytes obtained or the number of mature follicles at ultrasound. Besides the definition of poor response, the cut-off value of poor response also differs among studies, for example the number of oocytes retrieved may vary between <4 and <7 oocytes. This may lead to heterogeneous study groups and therefore potential difficulties in pooling of data. However, as studies appeared quite homogenous regarding the quality characteristics analyzed, the spread across the ROC diagram indicates that cut off values and definitions for the outcome variable used are very likely to be the cause for this variation. This is also exemplified by the fact that a summary ROC curve could be fitted to the studies.

It should be remembered that the purpose of any ORT is the identification of women with poor ovarian reserve for their age. This meta-analysis assessed the performance of AMH in a univariate context, independent of female age, while female age is the most important predictor in a priori prospects for IVF outcome (99). Thus, clinical studies in which the performance of AMH is compared in a multivariable analysis taking into account its interaction with female age are needed before the true applicability of AMH can be established.

Finally, no international assay standard for AMH measurements exists, possibly contributing to the discordance between different studies and therefore making comparison between laboratories difficult (91). Also, there is a moderate intercycle and interobserver variability in the Antral Follicle Count (94). Currently, the role of these factors cannot be separately analyzed.

Implications for clinical practice

The question is whether the fairly good predictive ability of AMH regarding the occurrence of poor response to hyperstimulation has clinical value. An ideal ovarian reserve test should identify a substantial percentage of IVF indicated cases, with a practically zero chance of becoming pregnant in a series of treatment cycles due to the adverse effects of diminished ovarian reserve. Those cases can be refrained from entering the program, as a quite high patient burden with only disappointing outcome can thereby be prevented. Also, high costs for only minimal results could be avoided. Accurate prediction of poor response

could therefore have clinical value if the pregnancy prospects are so unfavorable that a predicted poor responder would be denied treatment.
Here we face two problems as follows: one is the fact that poor response prediction is not fully reliable and that women with false positive test results may incorrectly be refrained from IVF. From the ROC curve in *Figure 1* it can be read that at a desired level of sensitivity of 70-80% a false positive rate of 10-20% can be expected. If predicted poor responders indeed have a very poor prognosis for pregnancy and should be refused for treatment extreme cut-off values would be used to prevent false positives. This would implicate that only minor percentages of abnormal tests will be found and many poor responders will pass unrecognized. Secondly, many poor responders achieve a pregnancy, although prospects indeed are less optimal compared to normal responders (100). Especially poor responders at a young age have a different prognosis compared to older poor responders (65). It is in fact the lack of a direct relation between quantity of response and quality of the oocytes that makes identification of very poor prognosis cases so difficult. Finally, the valuation of the weight of both false positive and false negative predictions should be considered. If patients are interviewed on the incorrect withholding IVF as compared to incorrect performing IVF, they consider the first much worse than the second, thereby implicating that currently available tests for ovarian reserve have in itself insufficient accuracy to withhold IVF (101).
Apart from the predictive meaning for the occurrence of pregnancy after IVF, the prediction of poor ovarian response is also potentially important for individual adjustment of the dose of gonadotrophins prior to IVF. Patients with a poor expected response are believed to benefit from a starting dose of 225 instead of 150 IU/day. A randomized trial on the subject showed that an individual dose regimen in a well-defined 'standard' patient population increased the proportion of appropriate ovarian responses and decreased the need for dose adjustment during controlled ovarian stimulation (102). In contrast, a randomized trial in predicted poor responders based on a prior AFC showed no benefit for response or pregnancy rates of a stimulating dose of 300 units compared to a dose of 150 IU (103).
In fact, the point may be raised as to whether there is any proven effective management for poor responders. In short, there are two strategies. The first method functions by using higher amounts of gonadotrophins, whereas the second method functions on the belief that with the addition of medication, ovarian sensitivity may improve. Although high doses of gonadotrophins have been used by the vast majority of authors, results have been controversial and prospective randomized studies have shown little or no benefit. Adjuvant ther-

apy with growth hormone (GH) or GH-releasing factors did not result in significant improvement. The use of corticosteroids and nitric oxide donors has shown encouraging results but confirmation studies are lacking. Finally, natural cycle IVF has produced results which are comparable with those obtained with stimulated cycles in true poor responders. Well-designed, large-scale, randomized, controlled trials are needed to assess the true efficacy of these different management strategies (104).

To date, basal FSH is the most commonly used test for ovarian reserve estimations. The accuracy and clinical value of this test has been much debated in a recent review (58). In a comparison with the Antral Follicle Count, basal FSH appeared inferior in poor response prediction (67). Therefore, based on the present results, AMH may become a test for quantitative ovarian reserve that is to be preferred over basal FSH. Apart from the cycle instability of FSH (56) compared to AMH, FSH levels may become elevated due to other causes, like familial dizygosity or FSH receptor polymorphisms (64;105;106).

Based on the current status of ovarian reserve tests it can be proposed that IVF may be initiated without any ovarian reserve test carried out. The response in the first cycle could then serve as a first line test and if it is poor, it may necessitate the application of an ORT like the AFC or AMH. If such a test would confirm the existence of a poor ovarian reserve, then the prognosis can be considered poor and further treatment refrained (107). If normal, a dose adaptation may be considered worthwhile and continuation of treatment justified.

Future research

Knowledge regarding the processes that dictate reproductive aging is still limited. We understand that follicle numbers decline with age but lack knowledge on how follicle reserve builds up in the fetal ovaries and is subsequently wasted. Thus, we cannot explain inter-individual variation in this reduction process. We recognize that oocytes lose the competence to produce viable embryos with advancing age, but fail to understand the mechanisms behind this process. Two decades of research on ovarian reserve has not delivered a highly accurate endocrine or imaging test that makes a clear clinical difference in patient management. Identifying genetic markers of the processes that regulate follicle quantities and oocyte quality (108) as well as longitudinal studies on the relationship between these markers and the occurrence of menopause (109) appear needed to truly advance the field of assessing ovarian aging and predicting reproductive potential on an individual basis.

In summary, this meta-analysis has shown that AMH has at least the same level of accuracy and clinical value for the prediction of poor response and non-pregnancy compared to the AFC. Clinical applicability ultimately depends on the way abnormal test results will alter patient management.

Acknowledgements

We gratefully thank the following persons for their cooperation by providing us with the necessary data:
Prof. Dr. Ebner, IVF-Unit, Landes- Frauen- und Kinderklinik, Linz, Austria.
D. Masson, Laboratory of Clinical Chemistry, CHU Nantes, France.
J.M.J. Smeenk. M. D., Department of Obstetrics and Gynecology, Radboud University Nijmegen Medical Center, The Netherlands.
Dr. T. Eldar-Geva, IVF-Unit, Institute of Hormone Research and Ultrasound Unit, Department of Obstetrics and Gynecology, Shaare-Zedek Medical Center, Ben Gurion University of the Negev, Jerusalem, Israel
J. Kwee, Division of Reproductive Endocrinology and Fertility and the IVF Centre, department of Obstetrics and Gynaecology, Vrije Universiteit Medical Centre, Amsterdam, the Netherlands.

Chapter 3

Added value of ovarian reserve testing on patients characteristics in the prediction of poor ovarian response and ongoing pregnancy after IVF: an individual patient data approach

The IMPORT* studygroup

S.L. Broer§, J. van Disseldorp§, K.A. Broeze, B.C. Opmeer, R.A. Anderson, M. Ashrafi, L. Bancsi, E. Caroppo, A.B. Copperman, T. Ebner, T. Eldar-Geva, M. Erdem, E.M. Greenblatt, K. Jayaprakasan, N. Raine-Fenning, E. Klinkert, J. Kwee, A. La Marca, M. McIlveen, L.T. Merce, S. Muttukrishna, S.M. Nelson, H.Y. Ng, B. Popovic-Todorovic, J.M.J. Smeenk, C. Tomás, P.J.Q. Van der Linden, I.K. Vladimirov, P. Bossuyt, M.J.C. Eijkemans, B.W. Mol and F.J.M. Broekmans

** Individual Meta-analysis of Patient data for Ovarian Reserve Testing*

§ both authors contributed equally

Submitted

Abstract

Context

Ovarian reserve tests (ORT) are frequently used prior to IVF-treatment for outcome prediction, however it is unclear whether they add prognostic value to readily available patient characteristics such as age.

Objective

To assess the added value of ORT to patient characteristics in the prediction of poor ovarian response and pregnancy after IVF.

Data sources

Individual patient data from previously published studies.

Study selection

Studies on FSH, AMH or AFC in women undergoing IVF, published until December 2009.

Data extraction

We used random intercept logistic regression prediction models to correct for between-study heterogeneity in estimating the added prognostic value of ORT on patient characteristics.

Results

We received 28 study-databases from 24 authors, regarding 5,705 women undergoing IVF. For the prediction of a poor response age had an area under the ROC curve (AUC) of 0.61. Both AFC and AMH clearly and significantly added prognostic value to age (p-value for each <0.001). A model with age, AFC and AMH had an AUC of 0.80. Similar accuracy was also reached by AMH or AFC, when used in isolation (AUCs 0.78 and 0.76, respectively). Combining the two tests did not improve the prediction of poor response (p=0.19). Age was the best single predictor of an ongoing pregnancy (AUC 0.57) here ORT did not have added value.

Conclusions

This IPD meta-analysis demonstrates that AFC and AMH clearly add to age in the prediction of poor response to ovarian stimulation for IVF and that these ORT in themselves can predict poor response well. In contrast, ORT do not add any information to the limited capacity of female age to predict ongoing pregnancy after IVF. This implies that claims on the usability of ORT prior to IVF must be limited to prediction of ovarian response, while identifying zero prognosis patients remains illusive.

Introduction

The incorporation of ovarian reserve tests (ORT) in IVF started after initial publications indicated a potential role for basal FSH in predicting pregnancy outcome after IVF and the usefulness in counseling patients (109;110). Since these first publications, a large body of additional work on basal FSH and several other tests has been published, often with inconsistent findings of the magnitude and direction of the predictive effect. It became evident that the clinical value of previously published prediction models was highly dependent on the consequences related to the prediction (i.e. counseling versus refraining from treatment). Moreover, female age, itself strongly related to IVF outcome, was frequently omitted as a serious contributor in the prediction models (58;111).

Overall, individual studies have shown considerable variation in predictive capacity of ovarian reserve tests. The conventional way to summarize the available evidence would be to perform a systematic review and meta-analysis of the sensitivity and specificity of ovarian reserve tests, as reported in the published studies (112). A major problem in interpreting the published studies is the striking heterogeneity in for example individual patient populations, stimulation protocols, hormone assays and ultrasound techniques. Conventional meta-analysis of the accuracy of tests cannot easily account for this heterogeneity, nor does it respect the continuous nature of ovarian reserve test data, or the statistical dependence between related tests and variables: the results of ORT are related to female age, and both are predictive of IVF outcome (113). To arrive at summary estimates of the added value of ORT in women undergoing IVF, we undertook a meta-analysis with original individual patient (IPD) data. By collecting test results, age and other patient characteristics and IVF outcome in each individual patient, we would be able to respect the continuous nature of ORT data and to study the added value of ORT to basic patient characteristics in predicting IVF outcome. Our aim was to answer the question whether the most widely used ORT, Follicle Stimulating Hormone (FSH), Antral Follicle Count (AFC) and Anti-Müllerian Hormone (AMH) added significantly and substantially to baseline female characteristics, such as age, in predicting the outcome of IVF treatment.

Methods

Data acquisition

We started with a literature search to identify studies of the value of FSH, AFC and AMH in predicting IVF outcome. We built on searches performed in previous, conventional systematic reviews on the subject (58;114). A systematic literature search was performed in PubMed to identify additional eligible papers, published until December 2009 (*Figure 1*). Eligible for the current review were studies presenting data on at least one ORT and at least one patient characteristic and IVF outcome, in terms of ovarian response to stimulation, clinical or ongoing pregnancy, or both.

Keywords used were synonyms for in vitro fertilization (IVF, controlled ovarian stimulation, in vitro fertilisation) and synonyms for the respective ORT (FSH, Follicle Stimulating Hormone, AFC, Antral Follicle Count or number, AMH, Anti-Müllerian Hormone, Müllerian inhibiting substance). All titles and abstracts were evaluated for eligibility by two authors (SB, JvD) and if necessary the opinion of a third author was decisive (FB). All authors of identified eligible primary studies were informed about this IPD meta-analysis project and invited to share their data in a collaborative project. If authors were inclined to participate, they were provided with a data request form, informing them on the format of the data requested.

After data acquisition, all data were carefully examined and when possible converted into a single format. Any issues or inconsistencies were checked with the original author. For more detailed description of IPD meta-analysis methodology the reader is referred to previous papers of our group (113;115).

A comparison was made between the studies that were and were not included. If possible, sensitivity and specificity of the ORT in the prediction of a poor response or ongoing pregnancy were calculated for the included and not included studies. For these two groups (included and not included studies) a Spearman correlation was calculated for every ORT and outcome measure, to test whether the differences in sensitivity and specificity were the result of different threshold levels and therefore to study the association between sensitivity and specificity. The Spearman correlations for each ORT and outcome were then compared between these groups, to see whether these groups were comparable.

Statistical analysis

All analyses were performed both for poor response as well as for ongoing pregnancy. Poor response was defined as the yield of 4 or less oocytes at follicle aspiration or a cancelled cycle due to poor ovarian response (less than 3-4 dominant follicles (>12 mm diameter)), since this is a common used definition for poor response (114). Ongoing pregnancy was defined as a visible gesta-

tional sac on ultrasound with heartbeat at a gestational age of at least 9 weeks. Duration of subfertility was defined as the period from the cessation of oral contraceptive use or start of unprotected intercourse until the first IVF attempt. A missing value analysis on the ORT and patient characteristics female age, BMI and duration of subfertility was performed. When a particular variable was missing in an individual database, data were not imputed.

Random intercept logistic regression prediction models were then created with 'Lme4' library in R (version 2.9.0. (http://www.r-project.org/), using the Laplace approximation to the likelihood. These models were created to quantitatively estimate the added value of the ovarian reserve tests on the patient characteristics in predicting poor response or ongoing pregnancy. By using a random intercept, the heterogeneity in prevalence of poor response or ongoing pregnancy between the original studies could be corrected for.

Three different models for the prediction of poor response or ongoing pregnancy were used. The first model included the patient characteristics female age, BMI and duration of subfertility. In the second set of models the predictive capacity of individual ovarian reserve tests FSH, AFC and AMH in combination with significant patient characteristics was estimated. In the third set of models the added value of combinations of ovarian reserve tests on patient characteristics was evaluated.

We then constructed Receiver Operating Characteristic (ROC) curves. Using the random intercept logistic regression models, probabilities of poor response or ongoing pregnancy could be calculated. Based on these, we plotted stratified ROC curves, with the ROC regression model as proposed by Janes and Pepe (116;117). This model assumes that studies share a common ROC for each ORT, but allow the positivity threshold corresponding to each sensitivity-specificity pair to vary between studies. With this model the improvement in predictive accuracy of adding an ORT to other variables can be studied, while correcting for the heterogeneity between studies. This way we could compare the ROC and AUCs of the models described above and evaluate them for statistically significant difference.

Because not all studies in this meta-analysis would report data for all three ORT, we constructed the prediction models using those databases from the total dataset that included the three ovarian reserve tests (FSH, AFC and AMH) and age to allow for direct comparison The results of our analyses in the three-test study groups were then checked against the effects in the total study group. Data were analyzed using SPSS 17.0 (SPSS Inc., Chicago, Il, USA), SAS 9.1 (SAS Institute Inc., Cary, NC, USA) and R version 2.9.0. (http://www.r-project.org/).

Results

Data acquisition

We identified 115 eligible manuscripts, of which we obtained contact information from 93 authors. From these 93 authors, 67 replied to our (repeated) email or phone contact. Ultimately, we received 28 study-databases which had been used for the publication of 55 manuscripts, provided by 24 collaborating authors (73;80;81;84;85;87;89;96;101;118-136). These 28 databases contained data on 5,705 subfertile women (*Figure 1*). Data from 4,170 women were suitable for poor response analysis, of which 893 (21%) had a poor response. Data from 5,367 women could be used for the analysis of ongoing pregnancy prediction, of which 1,231 women (23%) obtained an ongoing pregnancy.

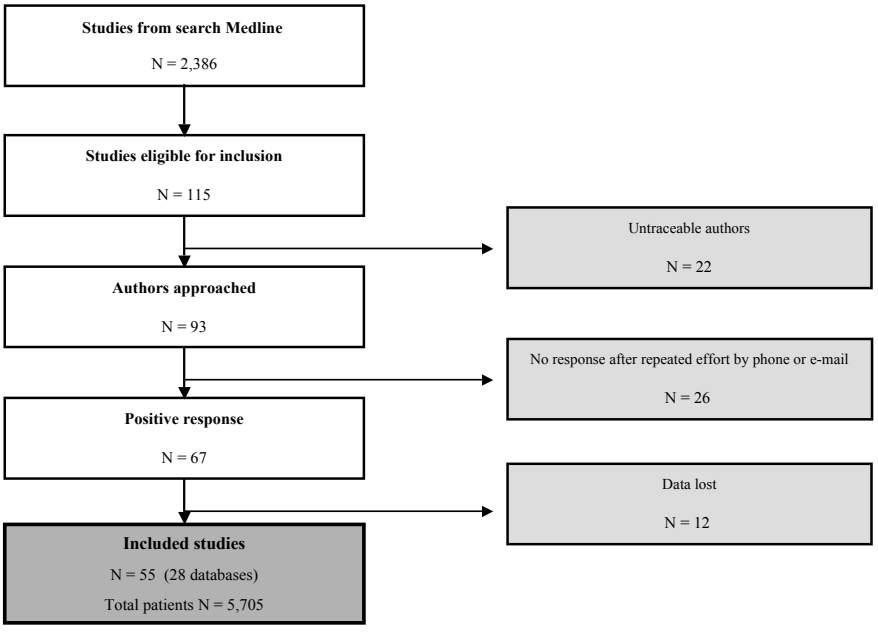


Figure 1. Flowchart of included studies

Baseline characteristics of the study group are summarized in *Table 1*. Baseline characteristics of the original studies are summarized, showing a high degree of variation in poor response and pregnancy incidences, as well as in ovarian reserve test averages between the original studies (*Table A-I, Addendum*). Study characteristics in terms of sampling, data collection and study design are shown in *Table A-II Addendum*. With the original data we were able to replicate

the primary findings of the original study in 10 databases. In 11 databases, the study database we received contained a number of patients that differed from the publication, whereas in seven other databases there were slight inconsistencies in the baseline data previously published. The level of consistency between the individual data and the data reported in the published manuscript was considered sufficient for all included studies. The comparison of the Spearman correlations of the included and not included studies for each ORT and outcome showed that for none of the ORT in both outcome measures a significant difference was found.

Table 1. Baseline characteristics of included studies

	Mean *(5th–95th percentile)*
Patient characteristics	
Female age (years)	34.3 (26.7 - 41.9)
FSH (IU/L)	7.8 (3.8 -14.0)
AFC (number)	11.6 (3.0 - 25.0)
AMH (ng/ml)	2.1 (0.1 - 6.0)
BMI (kg/m^2)	23.2 (18.5 - 30.1)
Duration of subfertility (years)	4.01 (1.0 - 9.1)
Prevalences	
Poor Response	21.4%
Ongoing Pregnancy	22.9%

Poor Response: ≤ 4 oocytes retrieved. Ongoing pregnancy: positive heartbeat at AD >9wk. Duration of subfertility: the period from the cessation of oral contraceptive use or start of unprotected intercourse until the first IVF attempt. N = 5,705

Prediction of a poor response and pregnancy from patient characteristics

For the model building exercises we could use the data from 617 women for poor response analysis and from 420 women for ongoing pregnancy analysis. Of all patient characteristics, age was the strongest single predictor of poor response (OR 1.12: 95% CI 1.08 to 1.17) (*Addendum Table A-III*). BMI and duration of subfertility were not significantly predictive of poor response.

In pregnancy prediction, age was the strongest single predictor of pregnancy, compared to other patient characteristics (OR 0.93: 95%CI 0.92 to 0.95) (*Addendum Table A-III*). Duration of subfertility was found not to be significantly associated with ongoing pregnancy, but BMI was. In a multivariable model only BMI added any predictive value to age (*Addendum Table A-III*). Since age was the

single constant and strongest predictor of poor response and ongoing pregnancy, all further multivariable analyses studied the added predictive effect of the ovarian reserve tests FSH, AFC and/or AMH on the predictive value of age.

Prediction of a poor response or ongoing pregnancy from ovarian reserve tests

We compared the ORT using the random intercept logistic regression model in predicting poor response (see *Table 2* and *Addendum Table A-IV*). The ROC regression analysis showed a high accuracy for AMH (AUC 0.78: 95% CI 0.72 to 0.84) and for AFC (AUC 0.76: 95% CI 0.70 to 0.82), but only a moderate accuracy for FSH (AUC 0.68: 95% CI 0.61 to 0.74) (*Table 3*). In predicting pregnancy after IVF all three ORT only had very small or no predictive effect (*Table 2* and *Addendum Table A-IV*). The AUC were 0.53, 0.50 and 0.55 for FSH, AFC and AMH, respectively) (*Table 3*). Age was the strongest single predictor of pregnancy after IVF, with moderate accuracy (AUC 0.57).

Table 2. Univariable and multivariable models of age and ORTs in the prediction of a poor response and ongoing pregnancy

	Poor Response Prediction			Ongoing Pregnancy Prediction		
	OR	*95% CI*	*P - value*	*OR*	*95% CI*	*P - value*
Univariable models						
Age (per year)	1.12	1.08 - 1.17	< 0.001	0.94	0.89 - 0.99	0.011
FSH (per IU/L)	1.27	1.19 - 1.35	< 0.001	0.98	0.92 - 1.04	0.477
AFC (per N)	0.77	0.73 - 0.82	< 0.001	1.00	0.97 - 1.03	0.951
AMH (per ng/ml)	0.50	0.41 - 0.60	< 0.001	1.09	0.96 - 1.24	0.197
Multivariable models						
Age and FSH						
Age (per year)	1.12	1.07 - 1.17	< 0.001	0.94	0.89 - 0.99	0.013
FSH (per IU/L)	1.26	1.18 - 1.34	< 0.001	0.99	0.93 - 1.05	0.632
Age and AFC						
Age (per year)	1.07	1.02 - 1.11	0.007	0.93	0.89 - 0.98	0.020
AFC (per N)	0.78	0.74 - 0.83	< 0.001	0.99	0.96 - 1.02	0.625
Age and AMH						
Age (per year)	1.08	1.03 - 1.13	0.001	0.94	0.89 - 0.99	0.017
AMH (per ng/ml)	0.54	0.44 - 0.66	< 0.001	1.06	0.93 - 1.21	0.373

Results of random intercept logistic regression model in the prediction of poor response or ongoing pregnancy. For the prediction of a poor response the multivariable analyses showed that all three ORTs add predictive information to female age alone.

Female age is the strongest predictor of ongoing pregnancy. All three ORTs show a very small or absent predictive effect in the prediction of an ongoing pregnancy. Multivariable analyses show that all three ORTs do not add predictive information to female age alone in the prediction of an ongoing pregnancy. P values reflect whether the variable plays a significant role in the model.

NB in three-test study group N = 617 for poor response prediction and N = 420 for ongoing pregnancy prediction. OR (Odds Ratio), 95%CI (95% Confidence Interval).

Table 3. AUCs of prediction models of age and ovarian reserve tests for the prediction of a poor response

	Three-test study group				Total study group			
	AUC	*95% CI*	*P value*	*N*	*AUC*	*95% CI*	*P value*	*N*
Poor Response Prediction								
Univariable models								
Age	0.61	0.54 - 0.68	NA	617	0.60	0.57 - 0.64	NA	4034
FSH	0.68	0.61 - 0.74	0.051	617	0.66	0.62 - 0.69	0.004	3652
AFC	0.76	0.70 - 0.82	< 0.001	617	0.73	0.69 - 0.77	< 0.001	2118
AMH	0.78	0.72 - 0.84	< 0.001	617	0.81	0.77 - 0.84	< 0.001	1274
Multivariable Models								
Age & FSH	0.71	0.65 - 0.78	< 0.001	617	0.69	0.66 - 0.72	< 0.001	3652
Age & AFC	0.79	0.73 - 0.85	< 0.001	617	0.76	0.72 - 0.80	< 0.001	2118
Age & AMH	0.77	0.70 - 0.83	< 0.001	617	0.80	0.76 - 0.84	< 0.001	1274
Age & AMH & AFC	0.80	0.74 - 0.86	< 0.001	617	0.80	0.74 - 0.86	< 0.001	618
Age & AMH & AFC & FSH	0.81	0.75 - 0.86	< 0.001	617	0.81	0.75 - 0.86	< 0.001	617
Ongoing Pregnancy Prediction								
Univariable models								
Age	0.57	0.47 - 0.66	NA	420	0.56	0.54 - 0.59	NA	5207
FSH	0.53	0.43 - 0.62	0.348	420	0.54	0.51 - 0.58	0.084	3521
AFC	0.50	0.40 - 0.59	0.100	420	0.52	0.48 - 0.57	0.612	1977
AMH	0.55	0.45 - 0.64	0.630	420	0.58	0.51 - 0.64	0.495	1008
Multivariable models								
Age & FSH	0.58	0.48 - 0.67	0.195	420	0.60	0.57 - 0.64	0.116	3521
Age & AFC	0.58	0.48 - 0.67	0.247	420	0.57	0.52 - 0.61	0.709	1977
Age & AMH	0.57	0.48 - 0.67	0.753	420	0.59	0.53 - 0.65	0.415	1008
Age & AMH & AFC	0.59	0.49 - 0.68	0.371	420	0.59	0.49 - 0.68	0.341	421
Age & AMH & AFC & FSH	0.58	0.49 - 0.68	0.414	420	0.58	0.49 - 0.68	0.414	420

Poor Response Prediction. In the univariable analysis it is shown that both AMH and AFC have a high accuracy, while FSH only has a moderate accuracy. In the multivariable models the added value to the AUC of an ORT on female age is shown, the p value indicates whether this added value is significant in comparison to age alone. All ORTs show a significant rise in the AUC. Moreover, the added value of adding several ORTs to female age is shown. The model including age, AFC and AMH reached the maximum predictive power. Addition of FSH to this model did not improve the predictive accuracy (P = 0.449). This level of accuracy is however also obtained when using a two factor model in the total study group (Age and AMH), or even with just AMH alone and the addition of age or AFC to AMH alone is not significant (p = 0.167 or p = 0.187, respectively).

Ongoing Pregnancy. In the univariable analysis it is shown that age is the strongest predictor compared to the single ovarian reserve tests. The multivariable analysis shows that no single or combined ORT adds substantional predictive power to age alone. This is shown in the three-tests study group, as well as in the total study group.

AUC = Area Under the Curve, ORT = Ovarian Reserve Test, AMH = Anti-Müllerian Hormone, AFC = Antral Follicle Count, FSH = Follicle Stimulating Hormone.

Multivariable prediction models for poor response and ongoing pregnancy

The multivariable analyses for poor response prediction showed that a model including age, AFC and AMH (AUC 0.80) had a significantly higher predictive accuracy than a model based on age alone (AUC 0.61). Addition of FSH to this model did not significantly improve the predictive accuracy (P = 0.45) (*Table 3*). Yet the predictive value of the multivariable model, including age and the two ORT AMH and AFC, was not significantly better than that of a single ORT, when used in isolation (p = 0.17 for AMH; p = 0.99 for AFC). AMH as a single predictor has an accuracy comparable to all multivariable models with AMH and age or any other ORT. The ROC curves corresponding to the multivariable analysis are shown in *Figure 2A*.

Multivariable analysis for prediction of ongoing pregnancy indicated that no single or combined ORT added substantial predictive power to age alone (*Table 3*).The AUC for the combination of Age, AMH and AFC was 0.59. The ROC curves corresponding to the multivariable analyses are shown in *Figure 2B*.

Discussion

The results of this IPD meta-analysis demonstrate that both AFC and AMH clearly add value to female age in the prediction of poor ovarian response in IVF, and that good predictions can be made with either AMH and AFC alone, even without using female age, with areas under the ROC curve reaching 0.80. Combining ORT does not significantly improve prediction. For prediction of ongoing pregnancy after IVF, ORT do not add to the limited capacity of female age.

The findings from our analysis, which was able to deal with the substantial heterogeneity across the contributing studies and data, are in line with conclusion suggested by previous systematic reviews and meta-analysis of both single ovarian reserve tests and multivariable prediction models for poor response to ovarian hyperstimulation (58;111;114). Both AMH and AFC strongly represent the size of the cohort of FSH sensitive follicles continuously present in the ovaries, often referred to as the quantitative ovarian reserve. Response to ovarian hyperstimulation has been shown to be directly linked to this cohort size (96). The role for AMH in marking the ovarian ageing process has been demonstrated in several studies showing that AMH decreases gradually with age and is predictive of the timing of menopause (36;37;71;137-139). From these data the capacity of AMH as a marker of the quantitative ovarian reserve has become established.

For ongoing pregnancy prediction, age is the single most important predictor, although accuracy of pregnancy prediction is far from optimal. In contrast with their performance in predicting poor response, the present data demonstrate

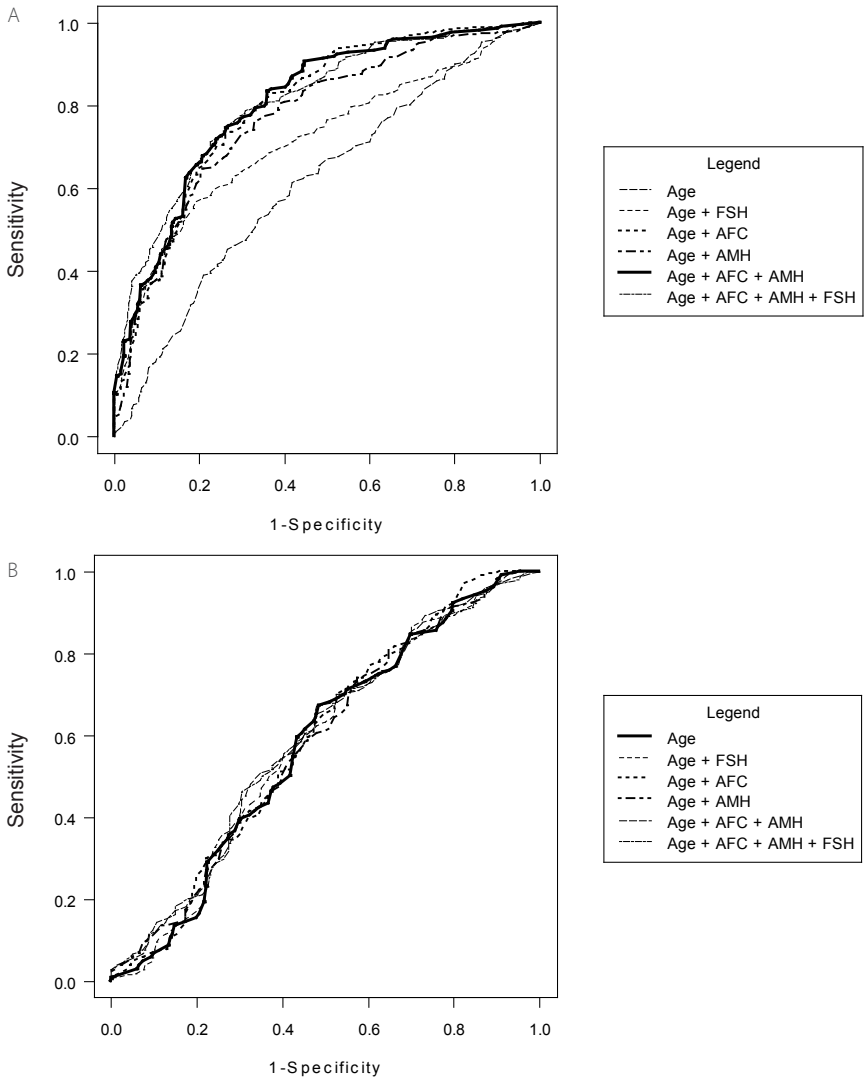


Figure 2. ROC curves of age and ORTs in the prediction of poor response and ongoing pregnancy
(A) Poor response prediction based on age and ORTs. In the upper panel the ROC curves of age and age combined with a single or more ORTs are depicted. The ROC curves for 'Age + AMH', 'Age + AMH + AFC' and 'Age + AMH + AFC + FSH' run toward the upper left corner, indicating a good capacity to discriminate between normal and poor responders at certain cut-off levels.
(B) Ongoing pregnancy prediction based on age and ORTs. The ROC curves age and age combined with one or more ORT run almost parallel to or even cross the X=Y line, indicating that the test will not perform better than flipping a coin in predicting who will become pregnant. The test is then considered useless for pregnancy prediction.
AFC, Antral Follicle Count; AMH, Anti-Müllerian Hormone; FSH, Follicle Stimulating Hormone; ORT, Ovarian Reserve Test; ROC, receiver-operating characteristic.

that ovarian reserve tests perform poorly in predicting pregnancy. The large body of data in the present analyses finally clarifies the lack of added value of currently known ORTs to knowing female age. Since ovarian response to controlled hyperstimulation expresses quantitative ovarian reserve and the occurrence of an ongoing pregnancy after IVF is mainly related to qualitative ovarian reserve, it can be emphasized that ORT reflect the quantitative aspect of the ovarian reserve status only. Qualitative ovarian reserve appears much harder to evaluate. In addition, ovarian reserve may not be the only factor affecting pregnancy chances in IVF/ICSI. Several factors, such as embryo quality, transfer technique or endometrial receptivity may act here (140). It is likely that only by studying several consecutive treatment cycles, a true representation of a woman's remaining reproductive capacity may be obtained. Over the past decades, only one study evaluated the predictive role for ovarian reserve tests in a series of subsequent IVF cycles, demonstrating that female age was the only factor predicting ongoing pregnancy after 3 treatment cycles, with no apparent role for ORT (106).

The performance of assisted reproduction technology (ART) in infertile couples is far from optimal. Out of every 100 couples initiating IVF, only 50-60 will achieve their goal, even after having undergone several treatment cycles. This high failure rate could be attributed to several factors, of which drop out rates and reduced ovarian reserve are the most popular ones. The urge to improve ART performance put a high focus on identifying adequate ovarian reserve tests. The limited accuracy of current tests has led to the situation that unfavorable test outcomes only lead to counseling and treatment adaptations that lack a solid scientific basis, in stead of a refusal to offer ART treatment in the first place. Recent studies have suggested a role for the use of patient characteristics combined with AMH for identification of poor prognosis categories (141;142). The question how these predictions could alter patient management or aid in upgrading ART performance has remained unanswered yet. This may be also explained by the fact that very poor prognosis categories are very difficult to identify with sufficient precision.

Recent publications have suggested to calculate age specific decline curves in order to maximize the ORT accuracy (143-145). One study calculated age specific FSH levels and live birth chances and demonstrated that variation in the chances of live birth is primarily determined by age, and only to a much lesser degree by basal FSH (144). The analysis also demonstrated that FSH decline curves for five age groups yielded different cut off values in the prediction of delivery rates (145). Since there was a very low rate of abnormal tests, the authors' conclusion that basal FSH could serve as a reliable prognostic tool, remains to be demonstrated, even if age-specific cut-off values of the ORT would be applied.

In the long lasting debate on the true value of ovarian reserve testing prior to IVF, a systematic review of the literature with meta-analysis can be of help as an objective and systematic approach in summarizing the available evidence. At present, the conventional meta-analysis of multivariable models still poses considerable challenges that are difficult to tackle, especially where the added value of several factors of interest can not be analyzed. In addition to the conventional hurdles, in terms of heterogeneity in study population and variable measurement, a strong prerequisite therefore is that the models are sufficiently comparable in form and structure. The current level of reporting does not easily allow such comparison, making conventional meta-analysis not an easy task. The strength of the present collaborative effort obviously is the ability to analyze the independent added value of several relevant predictors in a large body of data. With the generous help of a large group of contributors, we have been able to collect data on a number of patients that, though not covering the entirely evidence base, far surpasses that of the largest study performed so far. Thereby we have achieved consistency in variable coding and a form of statistical analysis that accommodates the remaining heterogeneity between studies. It is to be expected that similar issues in the evaluation of tests and markers can be resolved with the meta-analysis of individual patient data. More and more funding agencies are inviting investigators to have a data sharing policy, and to allow others to benefit from the resources invested in the research. Inspired by the major successes achieved by the multicenter genetic consortia, those of us more interested in clinical research could develop similar initiatives for patient centered research. We strongly believe that joining efforts in multicenter collaborations, possible even fine-tuning and coordinating study protocols through prospective meta-analysis, is an inevitable next step for clinical science in the 21st century, not just for randomized trials of interventions, but also in the evaluation of medical tests and biomarkers.

Some weaknesses of the current approach may be acknowledged, for the current study the databases of 55 of the eligible 115 manuscripts could be obtained. We were unable to reach a number of authors, primarily because of inaccurate contact information or because authors did not reply to e-mail addresses provided. Furthermore, older data were often lost or kept in a format that could no longer be read or converted. The Spearman correlations of the included and not included studies were calculated and compared in order to study whether these groups were comparable. For none of the ORT in both outcome measures a significant difference in the Spearman correlations was found. We therefore believe, that the included and not included studies are comparable and that the current number of participants and level of detail allowed us to analyze a representative selection of the collected data.

The clinical use of markers like AMH, basal FSH and the AFC is mostly based on cut off levels. From the IPD based dataset, cut off levels for poor response prediction could be derived that have a general applicability. Unfortunately, the methods used for assessment of follicle numbers and AMH and FSH serum levels varied across the studies, thereby prohibiting the calculation of relevant cut off levels. To some extent, correction factors to standardize the results from various studies could be applied. Currently, however, this approach has not yielded final data for one of the three tests of interest. Therefore, centers for ART, applying tests for poor response prediction should rely on their own data analyses for cut off level assignment. Indeed, development of centre based prediction models for patient management or counseling is now gaining rapid attention.

The clinical implications of the present findings will necessarily remain limited to the use of ORTs in predicting poor response to controlled ovarian hyperstimulation. The real clinical value of the prediction of a poor response will depend on the consequences of the prediction result. So far clinicians do not agree on what alterations in treatment regimen may be of help in predicted poor responders (103;146;147). Various (pseudo)randomized controlled trials have investigated whether individualization of the FSH treatment dose results in higher pregnancy chances in poor responders (101;102;148-150). Only one study reports a dosing algorithm that would increase pregnancy chances in poor responders, while others were unable to reproduce these effects (101). For the optimization of the ovarian response, two studies have shown that with an individual dose the response could be optimized and fewer patients would have a poor response (101;150). This could have consequences for the treatment efficacy and costs. Future large, well designed randomized controlled trials are necessary to identify the best treatment option for poor responders. At present, the accuracy of pregnancy prediction is such that exclusion of patients other than on the basis of female age is not to be supported.

In conclusion, this IPD meta-analysis demonstrates that the ovarian reserve tests AFC and AMH add predictive accuracy to age in the prediction of poor response to ovarian hyperstimulation for IVF. The accuracy of AFC and AMH when used in isolation, is similar to that of multivariable models with age and ORT. A single test of AFC and AMH can be considered sufficient. The clinical applicability of ORT based dose adaptation on efficacy and costs remains to be demonstrated. More importantly, the correct identification of patients with a very poor prognosis for pregnancy after ART, will not be improved by any currently known ovarian reserve test. In the field of patient selection prior to ART, female age therefore remains the most important, though modestly effective, tool.

Addendum

Table A-I. Study Characteristics of the included studies

Study	Female age (years)	BMI (kg/m²)	Duration of subfertility (years)	FSH (IU/l)
	Mean (5th–95th percentile)	*Mean (5th–95th percentile)*	*Mean (5th–95th percentile)*	*Mean (5th–95th percentile)*
Anderson	34.3 (26.6-42.2)	23.9 (17.9-34.9)	5.1 (2.0-10.3)	7.5 (3.7-12.3)
Ashrafi	30.0 (22.6-39.5)	NA	6.4 (1.0-17.4)	6.2 (1.6-15.1)
Bancsi	34.6 (27.0-40.7)	NA	4.8 (1.9-11.1)	8.4 (4.1-15.0)
Caroppo	38.0 (35.0-43.0)	NA	NA	11.4 (4.9-21.2)
Copperman	35.5 (26.9-42.9)	NA	NA	7.4 (3.4-13.6)
Ebner	32.7 (24.0-39.2)	NA	4.1 (1.0-11.7)	8.1 (4.4-13.8)
Eldar-Geva	30.0 (22.3-37.0)	23.8 (17.7-37.3)	4.2 (1.5-10.3)	6.7 (3.7-11.1)
Erdem	35.2 (27.6-44.4)	NA	9.6 (1.3-20.8)	8.1 (3.9-14.7)
Greenblatt	33.5 (27.0-39.0)	NA	NA	6.6 (4.1-9.6)
Jayaprakasan	33.5 (25.1-39.0)	NA	NA	7.2 (4.0-10.7)
Klinkert	41.1 (38.2-44.7)	NA	NA	9.6 (3.7-20.0)
Kwee	34.0 (27.6-40.0)	NA	3.8 (1.3-7.0)	8.1 (4.2-14.1)
La Marca	35.5 (27.0-42.0)	NA	2.9 (1.0-6.3)	NA
McIlveen	37.3 (29.3-42.8)	NA	4.6 (1.0-13.9)	8.3 (4.7-12.0)
Merce	34.4 (27.3-39.0)	20.6 (17.2-24.4)	2.7 (1.0-6.0)	NA
Muttukrishna 2004	37.6 (28.4-45.0)	NA	NA	7.9 (3.2-16.7)
Muttukrishna 2005	35.4 (28.0-43.0)	NA	NA	6.9 (3.8-12.4)
Nelson	33.9 (26.0-40.0)	24.5 (19.7-30.1)	3.5 (3.0-4.0)	8.7 (3.9-16.5)
Ng 2000	34.3 (27.0-39.0)	22.2 (18.3-28.4)	4.9 (2.0-10.0)	6.5 (3.8-10.8)
Ng 2005	32.8 (28.0-37.0)	20.7 (17.5-26.3)	4.9 (2.0-10.0)	6.5 (4.0-9.0)
Popovic-Todorovic 2003a	32.3 (26.0-38.9)	22.8 (18.8-29.3)	NA	7.0 (4.5-10.0)
Popovic-Todorovic 2003b	32.6 (26.3-37.0)	23.3 (18.6-31.3)	NA	6.3 (3.8-9.0)
Smeenk 2000	34.5 (28.4-41.4)	23.8 (18.5-30.6)	NA	6.8 (3.4-11.4)
Smeenk 2007	32.9 (26.0-40.0)	NA	3.7 (1.0-8.0)	NA
Tomás	33.3 (26.0-39.0)	23.9 (19.1-30.0)	NA	NA
van Rooij	36.3 (28.4-43.9)	23.7 (18.6-31.2)	2.9 (1.0-6.9)	8.5 (3.7-18.2)
van der Linden	NA	NA	NA	8.5 (4.1-14.8)
Vladimirov	34.3 (26.0-44.0)	21.6 (18.9-26.3)	6.5 (3.0-18.0)	7.3 (2.4-14.1)

For each individual study the mean, 5th and 95th percentile of the patient characteristics female age, BMI and duration of subfertility and ovarian reserve tests FSH, AFC and AMH are shown. The percentage poor responders and women that achieved an ongoing pregnancy are also shown. A = AFC2-10mm, B = AFC2-5mm, C = AFC2-8mm, D = DSL assay, E = Beckman Coulter assay. NA = not available.

Table A-I. Continued

AFC (number)	AMH (ng/ml)	Prevalence Poor Response	Prevalence Ongoing Pregnancy	Number of patients
Mean (5th–95th percentile)	*Mean (5th – 95th percentile)*	*%*	*%*	*N*
12.9 (4.8-26.6)[A]	NA	22	14	58
NA	NA	40	NA	50
NA	NA	31	14	505
NA	NA	40	17	76
NA	NA	3	36	701
NA	3.4 (0.6-7.9)[E]	17	39	135
22.6 (5.0-50.4)[A]	3.1 (0.6-8.6)[E]	7	35	54
7.0 (2.8-16.0)[C]	NA	13	34	32
13.8 (5.0-28.5)[C]	NA	11	27	297
16.3 (6.1-29.0)[A]	NA	7	43	100
7.7 (2.0-17.0)[B]	NA	36	15	221
10 (2.6-20.0)[A]	3.0 (0.3-8.5)[D]	22	NA	110
NA	2.1 (0.4-6.1)[E]	26	20	118
7.4 (2.0-13.0)[A]	1.6 (0.5-3.7)[E]	49	11	84
9.2 (1.0-21.0)[B]	NA	19	35	65
NA	0.9 (0.1-4.4)[E]	44	NA	66
9.0 (2.6-16.5)	2.1 (0.1-6.0)[E]	15	NA	70
NA	1.8 (0.1-5.0)[D]	32	22	340
11.9 (4.0-20.0)	NA	25	17	131
8.9 (4.0-16.0)	NA	16	10	127
14.0 (5.0-27.0)[B]	NA	6	31	262
16.2 (5.3-29.7)[B]	NA	7	28	145
15.9 (5.0-30.0)[A]	3.0 (0.5-8.9)[E]	16	50	80
NA	NA	NA	17	1292
10.9 (2.0-23.0)[B]	NA	28	15	166
8.4 (1.0-20.9)[B]	1.1 (0.0-3.9)[E]	44	19	222
NA	NA	16	26	159
8.9 (3.0-17.0)[A]	2.8 (0.5-8.4)[E]	36	NA	39

Table A-II. Baseline characteristics of the included studies

Author	Consecutive	Cohort / Case control	Pro-/Retro spective	Blinding	Selection bias	Verification bias	One cycle per couple	Data per cycle
Anderson	no	cohort	prospective	no	yes	no	yes	yes
Ashrafi	no	cohort	retrospective	no	yes	yes	yes	yes
Bancsi	yes	cohort	prospective	no	no	no	yes	yes
Caroppo	no	cohort	retrospective	no	yes	no	yes	yes
Copperman	no	cohort	retrospective	no	no	no	no	yes
Ebner	yes	cohort	prospective	no	yes	yes	yes	yes
Eldar-Geva	yes	cohort	prospective	no	yes	no	yes	yes
Erdem	yes	cohort	retrospective	no	yes	yes	no	yes
Greenblatt	yes	cohort	retrospective	no	yes	no	yes	yes
Jayaprakasan	yes	cohort	prospective	no	yes	no	yes	yes
Klinkert	yes	cohort	prospective	no	yes	yes	yes	yes
Kwee	yes	cohort	prospective	no	yes	yes	yes	yes
La Marca	yes	cohort	prospective	no	yes	yes	yes	yes
McIlveen	yes	cohort	prospective	no	yes	no	yes	yes
Mercé	yes	cohort	prospective	no	yes	yes	yes	yes
Muttukrishna 2004	No	cohort	prospective	no	yes	no	yes	yes
Muttukrishna 2005	yes	cohort	retrospective	no	no	no	yes	yes
Nelson	yes	cohort	prospective	no	yes	no	yes	yes
Ng 2000	yes	cohort	prospective	no	yes	yes	yes	yes
Ng 2005	yes	cohort	prospective	no	yes	yes	yes	yes

Table A-II. Continued

Author	Consecutive	Cohort / Case control	Pro-/Retro spective	Blinding	Selection bias	Verification bias	One cycle per couple	Data per cycle
Popovic-Todorovic 2003a	yes	cohort	prospective	no	yes	yes	yes	yes
Popovic-Todorovic 2003b	yes	cohort	prospective	no	yes	yes	yes	yes
Smeenk 2000	yes	cohort	retrospective	no	yes	no	yes	yes
Smeenk 2007	no	cohort	prospective	no	no	no	yes	yes
Tomas	yes	cohort	prospective	no	no	yes	yes	yes
van Rooij	yes	cohort	prospective	no	yes	yes	yes	yes
van der Linden	yes	cohort	prospective	no	no	no	yes	yes
Vladimirov	yes	cohort	prospective	no	yes	no	yes	yes

Characteristics of al included studies were assessed and summarized in this table. All studies were cohort studies, with the majority prospectively set up. All studies analyzed the results per cycle, some studies analyzed more cycles per couple, in which case only the first cycle was analyzed.

Table A-III. Univariable and multivariable analyses of patient characteristics in the prediction of poor response or ongoing pregnancy

	Three tests study group			Total study group		
	OR	*95% CI*	*P - value*	*OR*	*95% CI*	*P - value*
Poor Response Prediction						
Univariable models						
Age (per year)	1.12	1.08 - 1.17	< 0.001	1.13	1.10 - 1.15	< 0.001
BMI (per kg/m^2)	1.05	0.99 - 1.11	0.114	1.03	0.99 - 1.06	0.154
Duration (per year)	1.01	0.93 - 1.09	0.854	1.04	1.00 - 1.08	0.038
Multivariable models						
Age and BMI						
Age (per year)	1.11	1.06 - 1.17	< 0.001	1.13	1.10 - 1.17	< 0.001
BMI (per kg/m^2)	1.03	0.97 - 1.09	0.356	1.02	0.98 - 1.06	0.284
Age and duration						
Age (per year)	1.12	1.07 - 1.17	< 0.001	1.12	1.09 - 1.14	< 0.001
Duration (per year)	1.01	0.94 - 1.09	0.796	1.03	0.99 - 1.07	0.199
Ongoing Pregnancy Prediction						
Univariable models						
Age (per year)	0.94	0.89 - 0.99	0.011	0.93	0.92 - 0.95	< 0.001
BMI (per kg/m^2)	0.91	0.85 - 0.97	0.005	0.94	0.90 - 0.97	< 0.001
Duration (per year)	0.90	0.79 - 1.04	0.145	0.93	0.89 - 0.97	< 0.001
Multivariable models						
Age and BMI						
Age (per year)	0.93	0.89 - 0.99	0.017	0.95	0.92 - 0.98	< 0.001
BMI (per kg/m^2)	0.91	0.85 - 0.98	0.009	0.94	0.91 - 0.97	< 0.001
Age and duration						
Age (per year)	0.92	0.87 - 0.97	0.002	0.95	0.93 - 0.97	< 0.001
Duration (per year)	0.88	0.77 - 1.02	0.085	0.93	0.89 - 0.97	0.002

Results of the random intercept logistic regression model in the prediction of a poor response are shown in the upper part of the table. Age is the strongest predictor of a poor response, both in the three-test study group as in the total study group. BMI and duration of infertility showed no predictive information for a poor response. Results of random intercept logistic regression model in the prediction of an ongoing pregnancy are shown in the lower part of the table. Age is the strongest predictor of pregnancy in the three-test population as in the total population. BMI is predictive of pregnancy in both the subpopulation and the total population. Duration of infertility is only predictive of ongoing pregnancy in the total database.
Dataset three-study group N = 617 for poor response prediction and N= 420 for ongoing pregnancy prediction. OR (Odds Ratio), 95% CI (95% Confidence Interval).

Table A-IV. Univariable and multivariable models of age and ORTs in the prediction of a poor response or ongoing pregnancy

	Poor Response Prediction			**Ongoing Pregnancy Prediction**		
	OR	95% CI	P - value	OR	95% CI	P - value
Univariable models						
Age (per year)	1.13	1.10 - 1.15	< 0.001	0.93	0.92 - 0.95	< 0.001
FSH (per IU/L)	1.19	1.16 - 1.23	< 0.001	0.93	0.91 - 0.96	< 0.001
AFC (per N)	0.80	0.78 - 0.83	< 0.001	1.02	1.00 - 1.03	0.031
AMH (per ng/ml)	0.40	0.34 - 0.47	< 0.001	1.14	1.06 - 1.22	0.001
Multivariable models						
Age and FSH						
Age (per year)	1.11	1.09 - 1.14	< 0.001	0.92	0.91 - 0.95	< 0.001
FSH (per IU/L)	1.18	1.14 - 1.21	< 0.001	0.95	0.92 - 0.98	< 0.001
Age and AFC						
Age (per year)	1.07	1.04 - 1.11	< 0.001	0.95	0.92 - 0.98	< 0.001
AFC (per N)	0.82	0.79 - 0.84	< 0.001	1.01	0.99 - 1.03	0.238
Age and AMH						
Age (per year)	1.06	1.03 - 1.10	< 0.001	0.95	0.92 - 0.98	0.002
AMH (per ng/ml)	0.44	0.37 - 0.51	< 0.001	1.09	1.01 - 1.18	0.027

In the prediction of a poor response the effects of AMH and AFC are stronger than that of the FSH. Multivariable analyses showed that all three ORTs add predictive information to female age alone in the prediction of a poor response. Female age is the strongest predictor of ongoing pregnancy. All three ORTs show a very small or absent predictive effect in the prediction of an ongoing pregnancy in the univariable and multivariable analysis.
NB in total study group. OR (Odds Ratio), 95% CI (95% Confidence Interval).

Chapter 4

Performance of ovarian reserve tests in clinical subgroups of patients undergoing ART: an Individual Patient Data meta-analysis

the IMPORT* studygroup

S.L. Broer, J. van Disseldorp, K.A. Broeze, B.C. Opmeer, R.A. Anderson, M. Ashrafi, L. Bancsi, E. Caroppo, A.B. Copperman, T. Ebner, T. Eldar-Geva, M. Erdem, E.M. Greenblatt, K. Jayaprakasan, N. Raine-Fenning, E. Klinkert, J. Kwee, A. La Marca, M. McIlveen, L.T. Merce, S. Muttukrishna, S.M. Nelson, H.Y. Ng, B. Popovic-Todorovic, J.M.J. Smeenk, C. Tomás, P.J.Q. Van der Linden, I.K. Vladimirov, P. Bossuyt, M.J.C. Eijkemans, B.W. Mol and F.J.M. Broekmans

** Individual Meta-analysis of Patient data for Ovarian Reserve Testing*

Submitted

Abstract

Introduction

A recent systematic review has pointed out that ovarian reserve tests (ORTs) can predict poor response to ovarian hyperstimulation for IVF/ICSI, but fail to correctly identify women with a very poor prognosis of pregnancy. Despite this overall result, it is very well possible that the accuracy of ovarian reserve tests actually varies across patient subgroups defined by age, duration of subfertility or other patient characteristics. We evaluated the discriminatory capacity of ORTs for IVF/ICSI outcome in clinically relevant subgroups, using the individual patient data from published studies.

Methods

Authors of primary published studies on ovarian response to controlled ovarian stimulation and ongoing pregnancy in IVF/ICSI cycles and at least one ovarian reserve test (FSH, AFC or AMH) were invited to share their databases with individual patient data. After databases had been merged according to a standard procedure, receiver operating characteristic (ROC) regression analyses were performed to study the effect of specific patient characteristics on the discriminatory capacity of the ORTs.

Results

We obtained data from 28 studies reporting on 5,705 subfertile women undergoing IVF/ICSI. For the prediction of a poor response, ROC regression analysis showed that the accuracy of all ORTs was lower with increasing age and the effect was significant for AMH ($P = 0.004$). In women with a longer duration of subfertility, the accuracy of AFC was significantly lower ($P = 0.002$). For the prediction of ongoing pregnancy, the predictive capacity of the ORTs improved with increasing age and the effect was significant for FSH ($P = 0.004$). Despite this effect, the accuracy of ongoing pregnancy prediction remained low in all subgroups.

Conclusion

This IPD meta-analysis demonstrated that AMH is less accurate in older women and AFC less accurate when the duration of subfertility is longer. Obvious improvement of poor response prediction in one of the specific subgroups could not be found. Ovarian reserve tests for predicting ongoing pregnancy continued to perform poorly across the various clinical subgroups. This implies that ovarian reserve tests remain applicable solely for response prediction, in unselected populations indicated for IVF.

Introduction

Since the introduction of IVF more then three decades ago, clinicians have searched for tests that can predict the outcome of IVF/ICSI treatment, in addition to female age. At present, various screening tests are available, of which Follicle Stimulating Hormone (FSH), Antral Follicle Count (AFC), and Anti-Müllerian Hormone (AMH) are most frequently used. Each of these quantitatively expresses features of ovarian reserve and the responsiveness to gonadotrophin stimulation for IVF. The test results are used for informed counseling of the patient, to adjust the treatment approach or the stimulation dosage, or to discuss entrance into the IVF program. So far, these tests have shown to be useful in discriminating between poor, normal and excessive responders, but they fail to inform correctly on pregnancy prospects (42;114;151).

Information from medical history (such as female age, duration of subfertility) and physical examination (like bodyweight and height) is commonly used in addition to the results of ovarian reserve testing. It may be possible that the accuracy of ovarian reserve tests actually varies across specific patient subgroups. As ovarian reserve decreases with increasing age, it could be questioned whether the accuracy of the prediction of poor response or ongoing pregnancy alters in age categories. For BMI it could be postulated that response prediction may be more difficult due to for example technical difficulties with the AFC or because of altered biologic availability of recombinant FSH. Longer duration of subfertility could reflect the severity of ovarian ageing and therefore constitute a subgroup, where accuracy of these tests will alter in conjunction with higher prevalence of poor ovarian reserve.

In previously published literature on ovarian reserve testing, these patient characteristics were usually not taken into account in the analysis. This may be explained by a limited sample size and therefore a lack of power to evaluate patient characteristics as determinants of the accuracy in specific subgroups. Systematic reviews and meta-analyses of diagnostic testing aim to summarize the results of multiple studies in order to offer more precise and reliable estimates of the accuracy of such tests (152-155). However, conventional meta-analysis, in which data are aggregated on the level of the individual study, cannot evaluate determinants of accuracy. In contrast, meta-analysis on the level of the individual patient data (IPD) is able to do so (115). Such an approach allows an analysis not at the study level but at the level of the individual patient in each study, having access to additional patient characteristics, while taking study differences into account (113;115).

In the present study we compared the discriminatory capacity of three ovar-

ian reserve tests: basal FSH, Antral Follicle Count (AFC) and Anti-Müllerian Hormone (AMH) in clinical subgroups defined by female age, BMI and duration of subfertility

Methods

Method of Individual Patient Data systematic review and meta-analysis

Existing systematic searches of studies of AMH, AFC and FSH as prognostic indicators of ovarian response to hyperstimulation and/or clinical or ongoing pregnancy were updated, and used to identify papers published up to December 2009 (58;114) (*Figure 1*). A systematic search in Medline was carried out using synonyms for in vitro fertilization (IVF, controlled ovarian stimulation, in vitro fertilisation) and synonyms for the various tests (FSH, Follicle Stimulating Hormone, AFC, Antral Follicle Count or number, AMH, Anti-Müllerian Hormone, Müllerian inhibiting substance) as keywords.

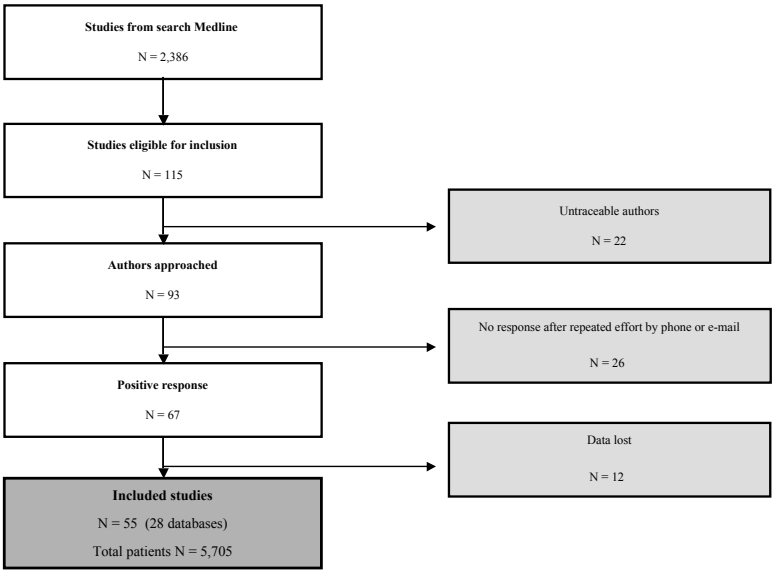


Figure 1. Flowchart of the search and selection strategy

Potentially eligible were studies that reported on the association of one or more of these ovarian reserve tests and the outcome measures poor ovarian response and/or pregnancy after an IVF/ICSI treatment and that had registered one or more patient characteristics. Studies including patients with ovulation disorders as the cause of subfertility were excluded.

All retrieved titles and abstracts were evaluated by two authors (SB, JvD) for eligibility and if necessary the opinion of a third author was decisive (FB). The authors of identified primary studies that met our eligibility criteria were approached and informed about this collaborative IPD meta-analysis project, and invited to share their data. If they were inclined to participate, they were provided with an author's agreement form and a data request form, informing them on the format of the data requested.
After data acquisition, all data were carefully examined and when possible converted into a single format. Any issues or inconsistencies were checked with the original author. For more detailed description of IPD meta-analysis methodology the reader is referred to previous papers (113;115).
A comparison was made between the studies that were and were not included. If possible, sensitivity and specificity of the ORTs in the prediction of a poor response or ongoing pregnancy were calculated for the included and not included studies. For these two groups a Spearman correlation was calculated for every ORT and outcome measure, to test whether the differences in sensitivity and specificity were the result of different threshold levels and therefore to study the association between sensitivity and specificity. The Spearman correlations of for each ORT and outcome were then compared between these groups, to see whether the included and not included studies were comparable.

Statistical analysis

All analyses were performed both for poor response as well as for ongoing pregnancy after IVF/ICSI treatment. A poor response was defined as the yield of 4 or less oocytes at follicle aspiration or a cancelled cycle due to poor ovarian response (less than 3-4 dominant follicles (>12 mm diameter) growing), since this is a common used definition for poor response (114). Ongoing pregnancy was defined as a visible gestational sac on ultrasound with heartbeat at a gestational age of at least 9 weeks. Duration of subfertility was defined as the period from the cessation of oral contraceptive use or start of unprotected intercourse until the first IVF attempt.
Within each study, we calculated the poor response and ongoing pregnancy rates. We then obtained Receiver Operator Characteristics (ROC) curves for each test in each study and estimated the Areas Under the Curve (AUC). These curves were unconditional, as they did not take patient characteristics into account. After this, we performed a meta-analysis of the accuracy data for each test, based on the data of all studies. We obtained summary estimates of the AUC while adjusting for the individual studies, using the model proposed by Janes and Pepe (116;117). In this model studies are assumed to share a common ROC

for each ORT, but the positivity threshold corresponding to each sensitivity-specificity pair is allowed to vary between studies.
To study whether the ORTs perform differently in subgroups defined by age, BMI or duration of subfertility we used the ROC regression model proposed by Pepe and Janes (116;117). In this model the ROC curves of the ORTs are modeled as a function of the covariates (age, BMI or duration of subfertility), since these patient characteristics can impact the inherent discriminatory accuracy of the ovarian reserve tests. This regression analysis resulted in an estimate that reflects the impact of the patient characteristics on the ROC curve and corresponding AUC of the ovarian reserve tests. In this meta-analysis, we assumed the effect of the covariate to be identical across studies, but, as in the previous analysis, the positivity threshold corresponding to each sensitivity-specificity pair was allowed to vary between studies, and therefore heterogeneity between studies is corrected for.
As a visual illustration of the results of the subgroup differences, we drew the ROC curves in subgroups defined by female age, BMI and duration of subfertility categories. The corresponding AUCs were calculated for each group in turn in order to express the overall discriminatory capacity (accuracy) of the ORT in women in the respective subgroups.
Data were analyzed using SPSS 17.0 (SPSS Inc., Chicago, Il, USA), SAS 9.1 (SAS Institute Inc., Cary, NC, USA) and R version 2.9.0. (http://www.r-project.org/).

Results

Data acquisition

In total, 115 eligible papers were identified in our literature search. The contact information of 93 authors could be retrieved and these authors were invited to join the collaborative project. The authors of 67 manuscripts replied to our repeated email messages or phone contact. Eventually, 24 authors were willing to collaborate (73;80;81;84;85;87;89;96;101;118-136). These authors provided us with 28 study-databases with the original individual patient data, which had been used for preparing 55 manuscripts (*Figure 1*).
Study characteristics, such as a clear description of sampling, data collection and study, were assessed (*Addendum Table A-I*). For 10 of the 28 databases baseline data tables were fully consistent, while 11 databases had a different number of patients in the database compared to the data tables described in the manuscript. In another 7 databases there were slight inconsistencies in the baseline data previously published. These differences were checked with the authors and resolved if possible. The level of consistency between the individual data and the data reported in the published manuscript was considered sufficient for all included studies.

The comparison of the Spearman correlations of the included and not included studies for each ORT and outcome showed that for none of the ORTs in both outcome measures a significant difference was found. Therefore, it can be assumed that the included and not included studies are comparable.

Statistical analysis

These 28 databases reported on a total of 5,705 subfertile women. Summary baseline statistics are shown in *Table 1*. These results are also shown per study in *Table A-II in the Addendum*. Data from 4,170 women were suitable for ovarian response analysis, of which 893 women (21%) had a poor response. For ongoing pregnancy analysis data from 5,367 women were available, of which 1,231 women (23%) obtained an ongoing pregnancy.

Table 1. Baseline characteristics of included studies

	Mean *(5th–95th percentile)*
Patient characteristics	
Female age (years)	34.3 (26.7 - 41.9)
FSH (IU/L)	7.8 (3.8 -14.0)
AFC (number)	11.6 (3.0 - 25.0)
AMH (ng/ml)	2.1 (0.1 - 6.0)
BMI (kg/m^2)	23.2 (18.5 - 30.1)
Duration of subfertility (years)	4.01 (1.0 - 9.1)
Prevalences	
Poor Response	21.4%
Ongoing Pregnancy	22.9%

Poor Response: ≤ 4 oocytes retrieved. Ongoing pregnancy: positive heartbeat at AD >9wk. Duration of subfertility: the period from the cessation of oral contraceptive use or start of unprotected intercourse until the first IVF attempt. N = 5,705

We obtained ROC curves for the different ORTs in the prediction of a poor response to controlled ovarian stimulation and in the prediction of an ongoing pregnancy, adjusting for the between study heterogeneity, using the model of Pepe and Janes (116;117). For each ORT, we estimated the area under the curve (AUC). The results of are shown in *Table 2*. The estimated AUCs of the ORTs in each individual study are shown in *Table A-III Addendum*. The overall AUC for

Table 2. Areas Under the Curve of the ORTs in the prediction of Poor Response or Ongoing Pregnancy

	Poor Response Prediction		
	FSH	*AFC*	*AMH*
AUC (95% CI)	0.66 (0.62 – 0.69)	0.73 (0.69 – 0.77)	0.81 (0.77 – 0.85)
N	3777	2118	1275
	Ongoing Pregnancy Prediction		
AUC (95% CI)	0.54 (0.51 – 0.58)	0.52 (0.48 – 0.57)	0.58 (0.48 – 0.64)
N	3666	1977	1009

Area Under the Curve, (95% Confidence Interval). AUCs were calculated using the ROC regression model as proposed by Janes and Pepe (116;117).

AMH was 0.81, versus 0.73 for the AFC and 0.66 for basal FSH. For ongoing pregnancy prediction, all the ORTs had more limited accuracy with AUCs of 0.54, 0.52 and 0.58, for FSH, AFC and AMH, respectively.

We then fitted ROC regression models to study the effect of the patient characteristics on the ROC curve of the ORTs in identifying women with a poor response and with an ongoing pregnancy (*Table 3*).

For the prediction of a poor response, the accuracy of all ORTs was lower with increasing age and the effect was significant for AMH (P = 0.004). Accuracy was also significantly lower for AFC in couples with a longer duration of subfertility (P = 0.002). BMI had no significant effect on the ROC for any of the ORTs in the prediction of a poor response. In contrast, the predictive capacity of the ORTs in predicting an ongoing pregnancy was higher with increasing age and the effect was significant for FSH (P = 0.004). Neither BMI nor duration of subfertility had any significant effect on the accuracy of the three ORTs to predict ongoing pregnancy.

To illustrate the influence of patient characteristics on the predictive capacity, we performed an ROC curve analysis in three sets of clinical subgroups, while taking the heterogeneity between studies into account (116;117). The summary subgroup ROC curves of AFC and AMH in the prediction of a poor response in clinical subgroups, defined by female age, BMI and duration of subfertility are shown in *Figure 2*. The AUCs of the three ORTs in the prediction of a poor response in subgroups defined by age, BMI and duration of subfertility are presented in *Table 4*. For the prediction of ongoing pregnancy, the ROC curves of FSH in subgroups defined by female age are shown in the addendum in *Figure A-I Addendum* and the ROC-AUCs of all ORTs are shown in *Table A-IV Addendum*.

Table 3. Results of the ROC regression analysis. ROC regression analysis showing the effect of the patient characteristics on the ROC curve of the ovarian reserve tests in the prediction of a poor response or ongoing pregnancy

	Coefficient	*95% CI*	*P-value*
		Age	
Poor Response Prediction			
FSH	-0.006	-0.025 to 0.012	0.555
AFC	-0.021	-0.048 to 0.006	0.150
AMH	-0.069	-0.115 to -0.023	**0.004**
Ongoing Pregnancy Prediction			
FSH	0.033	0.009 to 0.055	**0.004**
AFC	0.020	-0.006 to 0.047	0.142
AMH	0.027	-0.006 to 0.064	0.144
		BMI	
Poor Response Prediction			
FSH	-0.022	-0.059 to 0.014	0.23
AFC	-0.014	-0.057 to 0.037	0.562
AMH	0.002	-0.063 to 0.089	0.952
Ongoing Pregnancy Prediction			
FSH	0.008	-0.040 to 0.055	0.749
AFC	0.002	-0.045 to 0.044	0.92
AMH	-0.030	-0.098 to 0.044	0.424
		Duration	
Poor Response Prediction			
FSH	0.004	-0.034 to 0.044	0.857
AFC	-0.100	-0.168 to -0.042	**0.002**
AMH	-0.047	-0.131 to 0.023	0.242
Ongoing Pregnancy Prediction			
FSH	0.000	-0.061 to 0.053	0.992
AFC	-0.065	-0.132 to 0.005	0.056
AMH	-0.049	-0.146 to 0.019	0.242

Bold: Significant influence of the patient characteristics on the discriminatory capacity of the ovarian reserve test in the prediction of a poor response or ongoing pregnancy.

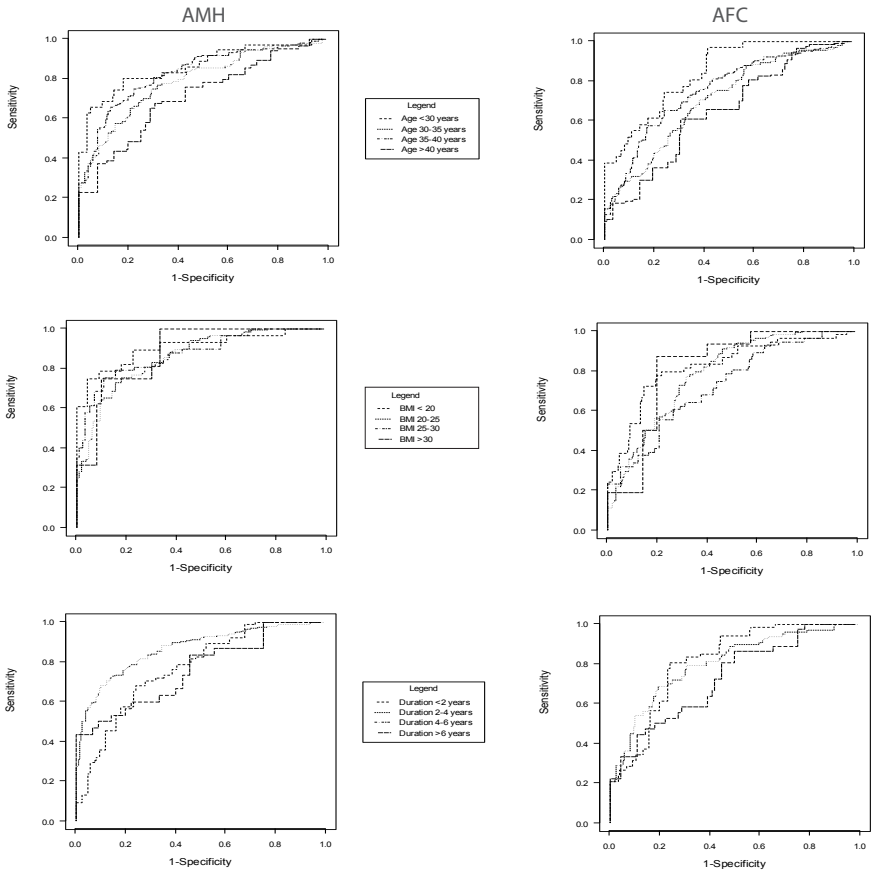


Figure 2. ROC curves for AMH and AFC in the prediction of a Poor Response in several clinical subgroups
In the left panel the ROC curves of AMH in the prediction of a poor response in several clinical subgroups, defined by female age, BMI or duration of subfertility are shown. Only age significantly influences the ROC curves of AMH. In the right panel the ROC curves of AFC in the prediction of a poor response in several clinical subgroups, defined by female age, BMI or duration of subfertility are shown. Only duration of subfertility significantly influences the ROC curves of AFC. The AUC and number of patients of each subgroup are mentioned in *Table 4*.

Table 4. AUCs of the ORT's in Poor Response Prediction in different subgroups

	FSH			AFC			AMH		
	AUC	*95%CI*	*N*	*AUC*	*95%CI*	*N*	*AUC*	*95%CI*	*N*
Age category									
< 30 yrs	0.69	0.57-0.80	527	0.81	0.67-0.96	281	**0.82**	0.69-0.96	188
30 – 35 yrs	0.64	0.58-0.70	1285	0.69	0.61-0.77	777	**0.77**	0.69-0.86	431
35 – 40 yrs	0.66	0.61-0.71	1369	0.74	0.68-0.81	778	**0.81**	0.75-0.87	458
> 40 yrs	0.60	0.52-0.67	401	0.61	0.51-0.70	230	**0.67**	0.57-0.78	144
BMI category									
< 20 (kg/m^2)	0.72	0.59-0.84	244	0.80	0.69-0.91	269	0.86	0.72-0.99	79
20-25 (kg/m^2)	0.74	0.67-0.81	784	0.78	0.71-0.85	728	0.84	0.77-0.91	366
25-30 (kg/m^2)	0.69	0.58-0.80	291	0.69	0.57-0.82	211	0.84	0.74-0.94	173
> 30 (kg/m^2)	0.77	0.57-0.97	58	0.68	0.43-0.93	39	0.77	0.54-0.99	42
Duration of subfertility category									
<2 yrs	0.62	0.52-0.73	144	**0.79**	0.69-0.89	106	0.74	0.64-0.84	150
2 - 4 yrs	0.71	0.65-0.76	634	**0.78**	0.70-0.87	278	0.86	0.81-0.91	415
4 - 6 yrs	0.65	0.54-0.75	490	**0.73**	0.60-0.87	183	0.87	0.73-1.00	264
> 6 yrs	0.64	0.54-0.75	178	**0.69**	0.53-0.85	180	0.69	0.52-0.87	84

AUC of the ORTs in several clinical subgroups defined by female age, BMI or duration of subfertility are shown. Bold: significant trend in differences between the AUC as calculated with the ROC regression model as showed in *Table 3*. There is significant trend in the difference of the AUCs of AMH in subgroups defined by female age and in the difference of the AUCs of the AFC in subgroups defined by duration of subfertility. The AUCs of FSH do not differ significantly in these subgroups.

Discussion

The results of the present IPD-meta-analysis confirm that AFC and AMH have substantial accuracy in the prediction of a poor response. Our results suggest that increasing age negatively influences the accuracy of ORT, with the discriminatory capacity diminishing in older women. This effect was significant for AMH. Likewise, longer duration of subfertility negatively affects the performance for the AFC. From the present data, BMI appeared not to influence the capacity of the ORTs in the prediction of a poor response.

These findings implicate that age influences the accuracy of AMH and that duration of subfertility influences the accuracy of AFC. Although ovarian reserve decreases with age, and that it could be postulated that ovarian reserve is also decreased in case of a longer duration of subfertility, AMH and AFC are believed to reflect the true level of the quantitative ovarian reserve directly. Indeed in older women or women with a longer duration of subfertility the prevalence of a poor response may differ, but this should not alter the accuracy of the tests. It could also implicate that within subgroups there is a different distribution of the sensitivity-specificity pairs. This means that a different cut-off could be used in these subgroups to maintain the desired level of sensitivity (consequently with a decrease in specificity), or the other way around with maintenance of a certain specificity level. Unfortunately, even in this IPD meta-analysis with 5,705 patients, it appeared not possible to determine these cut-offs because of the different assays and ultrasound methods used. These changes in the accuracy between subgroups may be significant only from the statistical point of view, without a true implication of clinical practice, and without an obvious explanatory mechanism.

In contrast, the predictive capacity of the ORTs in predicting an ongoing pregnancy was higher with increasing age and the effect was significant for FSH. This could possibly be explained by the unbalanced relationship between the quantitative and qualitative aspect of the ovarian reserve. With increasing age both the quantitative and qualitative aspect of the ovarian reserve decreases. For young women with a low quantitative ovarian reserve, the quality of the remaining oocyte may still be maintained and therefore they will still have favorable pregnancy prospects. In contrast, for an older woman with a low quantitative ovarian reserve the quality of the remaining oocyte is likely to be poor and therefore the pregnancy prospect is less favorable. Therefore, it is likely that with increasing age the prediction of the qualitative aspect (pregnancy) through the quantitative measures is more accurate. However, the accuracy of all three ORTs in the prediction of an ongoing pregnancy is poor, even in sub-

groups with increasing age. Therefore, accurate prediction of an ongoing pregnancy with the use of ORTs should still be considered impossible without the substantial risk of misclassification.

Some studies have evaluated the accuracy of ORTs in the prediction of live birth, in different subgroups for age. Scott et al. (156), studied the discriminatory capacity in subgroups defined by female age, but using an age specific cut-off value based on the 95%CI for that age group. In all age groups the discriminatory capacity remained poor. Lee et al. (157), did show a significant improvement of AMH in the prediction of live birth in women older then 35 years. Moreover, they also demonstrated an improvement of the accuracy of AMH in women when a male factor for fertility was absent. However, in both groups of women the area under the curve was at best 0.65, indicating that even in specific subgroups, the performance of AMH as predictor of live birth is insufficient. No previous studies regarding ORTs in the prediction of response in different subgroups have been performed.

Since prediction of pregnancy seems impossible, the clinical value of ovarian reserve tests will depend on the consequences of poor response prediction on clinical management. Can we prevent a poor response or alter clinical management based on a predicted poor response? Several studies have been performed regarding the adjustment or the starting dose or changing the treatment protocol. Several studies showed no effect of altered treatment or increasing the dosage of recombinant FSH (102;148;149). In contrast, two studies showed a positive effect of an individual starting dosage on the oocyte yield (101;150). Randomized trials in a large population are needed to confirm these results. Alternatively, can we do nothing but counsel and cancel the patient? As we usher in a new era of genomic medicine, we look forward to more personalized and precise markers of reproductive competence. In conjunction, it is our hope that patient-tailored protocols could be used to optimize stimulation even in those who are exhibiting diminished ovarian reserve.

Ovarian reserve testing for prediction of response and pregnancy has been an active area of research, and we were able to identify 83 eligible manuscripts. We could not reach many of the authors of these eligible manuscripts, primarily because of inaccurate contact information or because authors did not reply to e-mail addresses provided. Also, older data were often lost or kept in a format that could not be read anymore. Moreover, these eligible manuscripts also include studies regarding older ovarian reserve test like the clomiphene citrate challenge test (CCCT) and exogenous follicle stimulating hormone ovarian reserve test (EFORT). Unfortunately, this is a common problem in IPD meta-

analysis. Currently, studies are conducted to study the possibility to combine IPD data with aggregated data, to overcome this limitation (158). The Spearman correlations of the included and not included studies were calculated and compared to study whether these groups are comparable. For none of the ORTs in both outcome measures a significant difference in the Spearman correlations of these groups was found. Therefore, we believe, that the included and not included studies are comparable and that with the current number of participants and amount of data, we were able to analyze a representative selection of data available.

Using original data of different studies comes with heterogeneity between studies. The incorporation of ovarian reserve tests and restrictions based on test results in everyday IVF practice has led to selection bias in some study populations. Heterogeneity found in the included studies pertained to differences in IVF indications or access to IVF resources, differing treatment protocols and embryo laws and discordant definitions of ongoing pregnancy. There is also a variation in hormone assays and AFC sizes measured, for which no international consensus exist to correct for these differences. Consequently, no cut-off values for these tests could be used or mentioned. We have used the model by Janes and Pepe et al. (116;117) in which the heterogeneity between studies is corrected for.

In conclusion, this IPD meta-analysis shows that ORTs may be less accurate in poor response and non pregnancy prediction in certain subgroups. However, the observed effects will not disqualify the tests as predictors of poor response in IVF, nor qualify any test for non pregnancy prediction in a certain clinical subgroup. Therefor, ORTs remain applicable for poor response prediction in unselected populations indicated for IVF.

Addendum

Table A-I. Baseline characteristics of the included studies

Author	Consecutive	Cohort / Case control	Pro-/Retro spective	Blinding	Selection bias	Verification bias	One cycle per couple	Data per cycle
Anderson	no	cohort	prospective	no	yes	no	yes	yes
Ashrafi	no	cohort	retrospective	no	yes	yes	yes	yes
Bancsi	yes	cohort	prospective	no	no	no	yes	yes
Caroppo	no	cohort	retrospective	no	yes	no	yes	yes
Copperman	no	cohort	retrospective	no	no	no	no	yes
Ebner	yes	cohort	prospective	no	yes	yes	yes	yes
Eldar-Geva	yes	cohort	prospective	no	yes	no	yes	yes
Erdem	yes	cohort	retrospective	no	yes	yes	no	yes
Greenblatt	yes	cohort	retrospective	no	yes	no	yes	yes
Jayaprakasan	yes	cohort	prospective	no	yes	no	yes	yes
Klinkert	yes	cohort	prospective	no	yes	yes	yes	yes
Kwee	yes	cohort	prospective	no	yes	yes	yes	yes
La Marca	yes	cohort	prospective	no	yes	yes	yes	yes
McIlveen	yes	cohort	prospective	no	yes	no	yes	yes
Mercé	yes	cohort	prospective	no	yes	yes	yes	yes
Muttukrishna 2004	No	cohort	prospective	no	yes	no	yes	yes
Muttukrishna 2005	yes	cohort	retrospective	no	no	no	yes	yes
Nelson	yes	cohort	prospective	no	yes	no	yes	yes
Ng 2000	yes	cohort	prospective	no	yes	yes	yes	yes
Ng 2005	yes	cohort	prospective	no	yes	yes	yes	yes

Table A-I. Continued

Author	Consecutive	Cohort / Case control	Pro-/Retro spective	Blinding	Selection bias	Verification bias	One cycle per couple	Data per cycle
Popovic-Todorovic 2003a	yes	cohort	prospective	no	yes	yes	yes	yes
Popovic-Todorovic 2003b	yes	cohort	prospective	no	yes	yes	yes	yes
Smeenk 2000	yes	cohort	retrospective	no	yes	no	yes	yes
Smeenk 2007	no	cohort	prospective	no	no	no	yes	yes
Tomas	yes	cohort	prospective	no	no	yes	yes	yes
van Rooij	yes	cohort	prospective	no	yes	yes	yes	yes
van der Linden	yes	cohort	prospective	no	no	no	yes	yes
Vladimirov	yes	cohort	prospective	no	yes	no	yes	yes

Characteristics of al included studies were assessed and summarized in this table. All studies were cohort studies, with the majority prospectively set up. All studies analyzed the results per cycle, some studies analyzed more cycles per couple, in which case only the first cycle was analyzed.

Table A-II. Study Characteristics of the included studies

Study	Female age (years)	BMI (kg/m²)	Duration of subfertility (years)	FSH (IU/l)
	Mean (5th–95th percentile)	*Mean (5th–95th percentile)*	*Mean (5th–95th percentile)*	*Mean (5th–95th percentile)*
Anderson	34.3 (26.6-42.2)	23.9 (17.9-34.9)	5.1 (2.0-10.3)	7.5 (3.7-12.3)
Ashrafi	30.0 (22.6-39.5)	NA	6.4 (1.0-17.4)	6.2 (1.6-15.1)
Bancsi	34.6 (27.0-40.7)	NA	4.8 (1.9-11.1)	8.4 (4.1-15.0)
Caroppo	38.0 (35.0-43.0)	NA	NA	11.4 (4.9-21.2)
Copperman	35.5 (26.9-42.9)	NA	NA	7.4 (3.4-13.6)
Ebner	32.7 (24.0-39.2)	NA	4.1 (1.0-11.7)	8.1 (4.4-13.8)
Eldar-Geva	30.0 (22.3-37.0)	23.8 (17.7-37.3)	4.2 (1.5-10.3)	6.7 (3.7-11.1)
Erdem	35.2 (27.6-44.4)	NA	9.6 (1.3-20.8)	8.1 (3.9-14.7)
Greenblatt	33.5 (27.0-39.0)	NA	NA	6.6 (4.1-9.6)
Jayaprakasan	33.5 (25.1-39.0)	NA	NA	7.2 (4.0-10.7)
Klinkert	41.1 (38.2-44.7)	NA	NA	9.6 (3.7-20.0)
Kwee	34.0 (27.6-40.0)	NA	3.8 (1.3-7.0)	8.1 (4.2-14.1)
La Marca	35.5 (27.0-42.0)	NA	2.9 (1.0-6.3)	NA
McIlveen	37.3 (29.3-42.8)	NA	4.6 (1.0-13.9)	8.3 (4.7-12.0)
Merce	34.4 (27.3-39.0)	20.6 (17.2-24.4)	2.7 (1.0-6.0)	NA
Muttukrishna 2004	37.6 (28.4-45.0)	NA	NA	7.9 (3.2-16.7)
Muttukrishna 2005	35.4 (28.0-43.0)	NA	NA	6.9 (3.8-12.4)
Nelson	33.9 (26.0-40.0)	24.5 (19.7-30.1)	3.5 (3.0-4.0)	8.7 (3.9-16.5)
Ng 2000	34.3 (27.0-39.0)	22.2 (18.3-28.4)	4.9 (2.0-10.0)	6.5 (3.8-10.8)
Ng 2005	32.8 (28.0-37.0)	20.7 (17.5-26.3)	4.9 (2.0-10.0)	6.5 (4.0-9.0)
Popovic-Todorovic 2003a	32.3 (26.0-38.9)	22.8 (18.8-29.3)	NA	7.0 (4.5-10.0)
Popovic-Todorovic 2003b	32.6 (26.3-37.0)	23.3 (18.6-31.3)	NA	6.3 (3.8-9.0)
Smeenk 2000	34.5 (28.4-41.4)	23.8 (18.5-30.6)	NA	6.8 (3.4-11.4)
Smeenk 2007	32.9 (26.0-40.0)	NA	3.7 (1.0-8.0)	NA
Tomás	33.3 (26.0-39.0)	23.9 (19.1-30.0)	NA	NA
van Rooij	36.3 (28.4-43.9)	23.7 (18.6-31.2)	2.9 (1.0-6.9)	8.5 (3.7-18.2)
van der Linden	NA	NA	NA	8.5 (4.1-14.8)
Vladimirov	34.3 (26.0-44.0)	21.6 (18.9-26.3)	6.5 (3.0-18.0)	7.3 (2.4-14.1)

For each individual study the mean, 5th and 95th percentile of the patient characteristics female age, BMI and duration of subfertility and ovarian reserve tests FSH, AFC and AMH are shown. The percentage poor responders and women that achieved an ongoing pregnancy are also shown. A = AFC2-10mm, B = AFC2-5mm, C = AFC2-8mm, D = DSL assay, E = Beckman Coulter assay. NA = not available.

Table A-II. Continued

AFC (number)	AMH (ng/ml)	Prevalence Poor Response	Prevalence Ongoing Pregnancy	Number of patients
Mean (5th–95th percentile)	*Mean (5th – 95th percentile)*	*%*	*%*	*N*
12.9 (4.8-26.6)[A]	NA	22	14	58
NA	NA	40	NA	50
NA	NA	31	14	505
NA	NA	40	17	76
NA	NA	3	36	701
NA	3.4 (0.6-7.9)[E]	17	39	135
22.6 (5.0-50.4)[A]	3.1 (0.6-8.6)[E]	7	35	54
7.0 (2.8-16.0)[C]	NA	13	34	32
13.8 (5.0-28.5)[C]	NA	11	27	297
16.3 (6.1-29.0)[A]	NA	7	43	100
7.7 (2.0-17.0)[B]	NA	36	15	221
10 (2.6-20.0)[A]	3.0 (0.3-8.5)[D]	22	NA	110
NA	2.1 (0.4-6.1)[E]	26	20	118
7.4 (2.0-13.0)[A]	1.6 (0.5-3.7)[E]	49	11	84
9.2 (1.0-21.0)[B]	NA	19	35	65
NA	0.9 (0.1-4.4)[E]	44	NA	66
9.0 (2.6-16.5)	2.1 (0.1-6.0)[E]	15	NA	70
NA	1.8 (0.1-5.0)[D]	32	22	340
11.9 (4.0-20.0)	NA	25	17	131
8.9 (4.0-16.0)	NA	16	10	127
14.0 (5.0-27.0)[B]	NA	6	31	262
16.2 (5.3-29.7)[B]	NA	7	28	145
15.9 (5.0-30.0)[A]	3.0 (0.5-8.9)[E]	16	50	80
NA	NA	NA	17	1292
10.9 (2.0-23.0)[B]	NA	28	15	166
8.4 (1.0-20.9)[B]	1.1 (0.0-3.9)[E]	44	19	222
NA	NA	16	26	159
8.9 (3.0-17.0)[A]	2.8 (0.5-8.4)[E]	36	NA	39

Table A-III. AUCs of the individual studies in the prediction of a Poor Response or Ongoing Pregnancy

Study	Poor Response Prediction					
	FSH		AFC		AMH	
	AUC	*N*	*AUC*	*N*	*AUC*	*N*
Anderson	0.76	46	0.74	46	NA	NA
Ashrafi	0.57	50	NA	NA	NA	NA
Bancsi	0.57	505	NA	NA	NA	NA
Caroppo	0.48	76	NA	NA	NA	NA
Copperman	0.67	570	NA	NA	NA	NA
Ebner	0.70	127	NA	NA	0.82	135
Eldar-Geva	0.96	52	0.97	36	0.98	54
Erdem	0.84	24	0.72	24	NA	NA
Greenblatt	0.71	261	0.64	223	NA	NA
Jayaprakasan	0.56	100	0.93	100	NA	NA
Klinkert	0.62	212	0.73	221	NA	NA
Kwee	0.84	109	0.81	109	0.81	105
La Marca	NA	NA	NA	NA	0.69	118
McIlveen	0.54	71	0.72	70	0.70	71
Merce	NA	NA	0.65	88	NA	NA
Muttukrishna 2004	0.75	66	NA	NA	0.87	66
Muttukrishna 2005	0.56	68	0.7	68	0.6	68
Nelson	0.78	338	NA	NA	0.87	319
Ng 2000	0.74	131	0.75	131	NA	NA
Ng 2005	0.58	109	0.70	127	NA	NA
Popovic-Todorovic 2003a	0.69	256	0.70	256	NA	NA
Popovic-Todorovic 2003b	0.60	143	0.88	143	NA	NA
Smeenk 2000	0.67	80	0.67	80	0.75	80
Smeenk 2007	NA	NA	NA	NA	NA	NA
Tomas	NA	NA	0.69	160	NA	NA
van Rooij	0.71	220	0.82	220	0.83	220
van der Linden	0.72	124	NA	NA	NA	NA
Vladimirov	0.66	39	0.89	39	0.85	39

The AUCs of the ORTs in the prediction of a poor response and ongoing pregnancy are shown per individual study. NA = not available.

Table A-III. Continued

Ongoing Pregnancy Prediction					
FSH		**AFC**		**AMH**	
AUC	*N*	*AUC*	*N*	*AUC*	*N*
0.54	52	0.67	57	NA	NA
NA	NA	NA	NA	NA	NA
0.56	505	NA	NA	NA	NA
0.65	76	NA	NA	NA	NA
0.61	701	NA	NA	NA	NA
0.56	124	NA	NA	0.59	132
0.64	52	0.63	36	0.6	54
0.65	32	0.84	32	NA	NA
0.50	297	0.57	250	NA	NA
0.58	100	0.59	100	NA	NA
0.48	212	0.64	221	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	0.69	118
0.68	84	0.65	83	0.58	84
NA	NA	0.62	65	NA	NA
NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA
0.58	338	NA	NA	0.58	319
0.49	131	0.53	131	NA	NA
0.67	109	0.59	127	NA	NA
0.53	262	0.49	262	NA	NA
0.50	145	0.58	145	NA	NA
0.58	80	0.60	80	0.55	40
NA	NA	NA	NA	NA	NA
NA	NA	0.58	166	NA	NA
0.50	222	0.47	222	0.59	222
0.58	144	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA

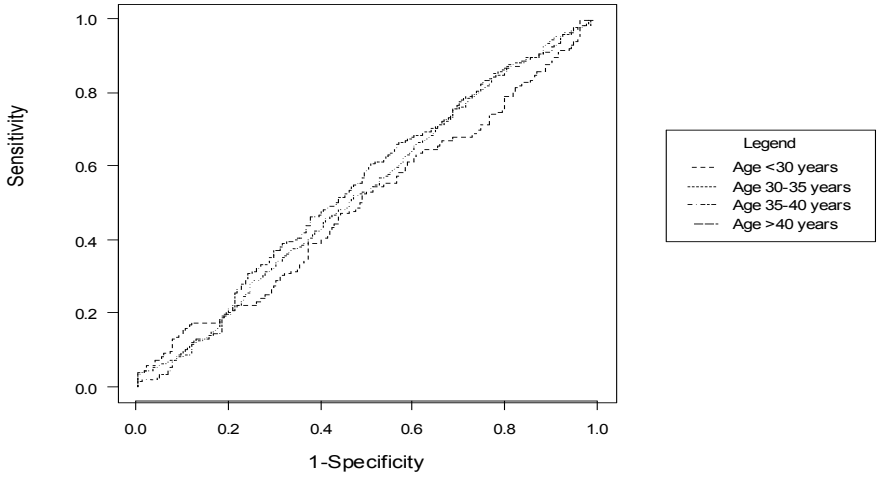


Figure A-I. ROC curves of FSH in the prediction of an ongoing pregnancy in subgroups defined by female age

The ROC curves of FSH in subgroups defined by female age in the prediction of an ongoing pregnancy are shown. ROC regression analysis showed a significant influence of female age on the discriminatory capacity of FSH for an ongoing pregnancy. However, in all subgroups the discriminatory capacity of FSH is low and therefore prediction of an ongoing pregnancy with FSH levels remains inaccurate.

Table A-IV. AUCs of the ORTs in Ongoing Pregnancy prediction in different subgroups

	FSH			AFC			AMH		
	AUC	95%CI	N	AUC	95%CI	N	AUC	95%CI	N
Age category									
< 30 yrs	**0.49**	0.42-0.56	470	0.44	0.35-0.53	242	0.52	0.39-0.65	158
30 – 35 yrs	**0.53**	0.48-0.58	1226	0.51	0.44-0.58	720	0.56	0.46-0.65	336
35 – 40 yrs	**0.54**	0.48-0.60	1332	0.54	0.45-0.62	721	0.52	0.41-0.63	349
> 40 yrs	**0.57**	0.42-0.71	436	0.56	0.36-0.75	240	0.67	0.34-1.00	128
BMI category									
< 20 (kg/m^2)	0.55	0.42-0.69	237	0.54	0.42-0.66	264	0.60	0.38-0.81	74
20-25 (kg/m^2)	0.52	0.45-0.59	767	0.48	0.40-0.55	722	0.56	0.46-0.66	356
25-30 (kg/m^2)	0.51	0.38-0.63	291	0.56	0.39-0.72	213	0.64	0.48-0.81	181
> 35 (kg/m^2)	0.53	0.23-0.83	79	0.34	0.00-0.73	65	0.45	0.06-0.83	48
Duration of subfertility category									
<2 yrs	0.45	0.30-0.61	133	0.52	0.34-0.70	96	0.54	0.41-0.67	150
2 - 4 yrs	0.56	0.49-0.64	634	0.49	0.37-0.62	277	0.56	0.48-0.65	415
4 - 6 yrs	0.54	0.37-0.70	250	0.42	0.15-0.70	158	0.60	0.23-0.98	264
> 6 yrs	0.56	0.39-0.73	284	0.57	0.29-0.84	144	0.75	0.51-0.99	84

AUC of the ORTs in several clinical subgroups defined by female age, BMI or duration of subfertility are shown. Bold: significant trend in differences between the AUC as calculated with the ROC regression model as showed in Table 3. The AUCs of FSH differ significantly in subgroups defined by female age. The AUCs of AFC and AMH do not differ significantly in these subgroups

Chapter 5

AMH and AFC as predictors of excessive response in controlled ovarian hyperstimulation: a meta-analysis

S.L. Broer, M. Dólleman, B.C. Opmeer, B.C. Fauser, B.W. Mol and F.J.M. Broekmans

Hum Reprod Update. 2011;17(1):46-54.

Abstract

Background

Anti-Müllerian Hormone (AMH) is a marker of ovarian reserve status and represents a good predictor of ovarian response to ovarian hyperstimulation. The aim of this study was to assess the accuracy of AMH and Antral Follicle Count (AFC) as predictors of an excessive response in IVF/ICSI treatment.

Methods

A systematic review and meta-analysis of the existing literature was performed. Studies were included if 2x2 tables for the outcome excessive response in IVF patients in relation to AMH/AFC could be constructed. Using a bivariate meta-analytic model, both summary point estimates for sensitivity and specificity were calculated, as well as summary ROC curves. Clinical value was analyzed by calculating post-test probabilities of excessive response at optimal cut-off levels, as well as the corresponding abnormal test rates.

Results

Nine studies reporting on AMH and five on AFC could be detected. Summary estimates of sensitivity and specificity for AMH were 82% and 76 %, respectively and 82% and 80 %, respectively for AFC. Comparison of the summary estimates and ROC curves for AMH and AFC showed no statistical difference. Abnormal test rates for AMH and AFC amounted to ~ 14 and 16%, respectively, at cut off levels where test performance is optimal (LR+ > 8), with a post test probability of +/- 70%.

Conclusions

Both AMH and AFC are accurate predictors of excessive response to ovarian hyperstimulation. Moreover, both tests appear to have clinical value. This opens ways to explore the potential of individualized FSH dose regimens based on ovarian reserve testing.

Introduction

In in vitro fertilization (IVF) treatment, excessive response to FSH stimulation introduces the risk for abdominal discomfort, painful follicle aspirations and cycle cancellations (159). An excessive response will typically introduce the risk of ovarian hyperstimulation syndrome (OHSS), a potentially life threatening condition (160). Excessive response to ovarian stimulation will generate many oocytes for the laboratory, that will not unequivocally lead to a full range of good quality embryos (161-163). In addition, chances for pregnancy may decrease (99). In view of these drawbacks, elimination of exaggerated ovarian response in stimulation protocols will improve safety, success and cost factors of assisted reproduction technology (ART) programs.

For primary preventive management to be developed, the reliability of tools for prediction of ovarian response needs to be assessed first. Ovarian response prediction is mainly based on ovarian reserve tests like the Antral Follicle Count (AFC) and Anti-Müllerian Hormone (AMH) (58;114). The AFC comprises the number of 2-5 or 2-10 mm diameter follicles measured in the ovaries at the start of the menstrual cycle (164) and is highly correlated to the number of oocytes retrieved at pick up (91;114). AMH has been implicated as the most valuable marker of ovarian reserve as serum concentrations correlate highly with baseline AFC and the number of oocytes retrieved at aspiration (73;83;84;165-168). The aim of the present systematic literature review was to asses the true accuracy of AMH and AFC as prognosticators for the prediction of an excessive response after IVF/ICSI treatment.

Methods

Search and selection strategy

The literature was searched for studies that addressed the capacity of AFC or AMH as prognosticators of excessive ovarian response after controlled ovarian hyperstimulation in an IVF or ICSI treatment. No preset definition of excessive ovarian response was used. Excessive ovarian response definition included oocytes at retrieval above a certain threshold, estrogen-level above a certain threshold, the development of ovarian hyperstimulation syndrome (OHSS) or cycle cancellation due to a high response, or combinations of these. Also, any cut-off or set of cut-offs for an abnormal AMH or AFC were included in this review.

A systematic search in Medline was carried out using the keywords 'in vitro fertilization', 'in vitro fertilisation', 'assisted', 'intracytoplasmic', 'intracytoplasmatic' in combination with 'Anti-Müllerian Hormone', 'mullerian inhibiting factor', 'mul-

lerian inhibiting substance' or 'Antral Follicle Count'. A period of all the years through November 2009 was covered by the search. The abstracts of all studies identified were read by one researcher (M.D.). Any article that could possibly be of value for the association between AMH and AFC and the IVF outcome excessive ovarian response was preselected. In the next step, two researchers (M.D. and S.B.) carefully read and judged all preselected articles independently. If it was judged possible to construct 2x2 tables from the data presented in the paper, the study was selected for final inclusion and analysis in this review. In a 2x2 or contingency table, the true positive, true negative, false positive and false negative test results at a certain cut-off are displayed. In the event of any disagreement between the two authors, the opinion of a third researcher (F.B.) was final.

The authors of studies that reported on the ovarian reserve test result in relation to IVF outcome without the possibility of constructing 2x2 tables were contacted by email and asked to provide the necessary data for the construction of such a table. If adequate data were obtained in this way, the study was added to the selection. In every selected study, the reference list was scanned to identify studies that could possibly be included in the selection and then processed as described.

Each selected study was further scored by the researchers (M.D and S.B) regarding the following study quality characteristics: 1) patient sampling (consecutive vs. other); 2) data collection method (prospective vs. retrospective); 3) study design (cohort vs. case control); 4) blinding (present vs. absent); 5) selection bias, i.e. exclusion of cases based on criteria that affect the ability to generalize the findings of the study, for instance women with elevated basal FSH or women over 38 years of age (present or absent); 6) verification bias, ie. the use of results of the test under study in adapting the treatment protocol in order to prevent the predicted outcome, for instance poor ovarian response (present or absent); 7) analysis upon one or multiple cycles per couple; and 8) stimulation protocol (GnRH-agonist or GnRH-antagonist). Also, data on the cut-off levels used were recorded, as was the assay used for AMH measurement and whether AFC was measured in 2-5 or 2-10 mm follicles. Because this review used only published data form the literature, no approval from our institutional review board was required.

Analysis

First, 2x2 tables were constructed from which sensitivity and specificity were calculated. Sensitivity-specificity points were displayed in the Receiver Operating Characteristics (ROC) space (1-specificity versus sensitivity). Combinations

of sensitivity and 1-specificity are indicative of the test accuracy, with studies reporting high accuracy for both sensitivity and specificity are located in the upper-left corner of the ROC space, and poor test results are located close to the x=y line.

A meta-analysis was performed using a bivariate regression model (155). In short, this bivariate model preserves the two dimensional nature of prognostic data in a single model, rather than using a single outcome measure for each study such as the diagnostic odds ratio. The bivariate model simultaneously estimates sensitivity and specificity, and incorporates the negative correlation that may exist between sensitivity and specificity within studies, owing to possible implicit differences in the applied threshold between studies. When necessary, the bivariate model uses a random approach for both sensitivity and specificity, allowing for heterogeneity beyond chance due to clinical or methodological differences between studies. In addition, the model acknowledges the difference in precision by which sensitivity and specificity have been measured in each study. This means that studies with a larger number of women with an excessive response received more weight in the calculation of the pooled estimate of sensitivity, whereas studies with a high number of women without an excessive response were more influential in the pooling of specificity.

Sensitivity was plotted against 1-specificity (false positive rate) and pooled estimates for sensitivity and specificity were calculated and also plotted, together with the 95% Confidence Interval (CI) ellipse. As different studies have reported results for different thresholds to define a positive test (cut off), we did not limit our analysis to a single threshold value, but took advantage of the fact that the model incorporates opposite effects on sensitivity and specificity when using different cut-offs. In order to account for dependent observations (observations on different cut-offs from the same study are likely to be correlated), we estimated the model in 250 stratified bootstrap samples, in which only one threshold value from each study was randomly selected. The overall estimates of sensitivity and specificity were based on the average from 250 bootstrap samples.

The results of the model were used to estimate summary ROC-curves, where the increase in sensitivity and decrease in specificity reflect the shift in threshold value of the ovarian reserve in the model. We thereby had to convert parameter estimates from the bivariate model to those in the Summary ROC model, as these are basically different statistical approaches for the same underlying model (169). The difference between AMH and AFC in pooled sensitivity and specificity was tested by fitting the bivariate model on data for both tests, with test included as a covariate in the model.

To asses the clinical value of both tests, post-test probabilities for the prediction of an excessive response were calculated, by using the estimated summary ROC curve and assuming an arbitrary prevalence (or pre-test probability) of 20% for an excessive response. A series of likelihood ratio ranges for an abnormal test result was then derived from several points of the estimated summary ROC curve, and at these various ranges of likelihood ratios, the post-test probabilities for both tests were computed, as well as the corresponding abnormal test rates.
All statistical analyses were done using SAS 9.1 for Windows (Proc NlMixed in the bivariate model).

Results

Systematic review

The systematic Medline search produced 170 hits. Of these, 126 articles were excluded on the basis of title and abstract. Another 30 studies were excluded on the basis of the fully read article. Finally, 14 studies were selected to be appropriate for the current meta-analysis. From those 14 studies, in four studies 2x2 tables could be constructed from the article itself. The remaining 10 authors were contacted and asked for the necessary data. Three studies could not be included as the authors did not reply to the email request (74;83;170). Seven authors did respond with the appropriate data to construct 2x2 tables. Thus, a final number of 11 studies could be included for data extraction and meta-analysis (73;84;87;88;91;128;130;167;168;171;172). Six studies reported on the capacity of AMH to predict excessive response after IVF, two studies reported on the capacity of AFC, and three studies studied both AMH and AFC (*Figure 1*). The characteristics of the included studies are listed in *Table 1*. From this table it becomes clear that all studies but one presented data for one cycle per couple and that the majority used a prospective cohort design. However, selection bias was judged to be present in all studies. This concerned the exclusion of older women or women with signs of decreased ovarian reserve, or exclusion of cases with the PCO syndrome. The definition of excessive response was not uniform. It ranged from number of oocytes retrieved over 14 up to over 21 or the development of OHSS.

Accuracy of AMH in excessive response prediction

Sensitivities and specificities for the prediction of excessive ovarian response, as calculated from each study reporting on AMH, are summarized in *Table 2a*. A plot of sensitivity-specificity combinations in an ROC space is shown in *Figure 2*. For AMH, the sensitivity varied between 40% and 95% and the specificity between 31% and 96%.

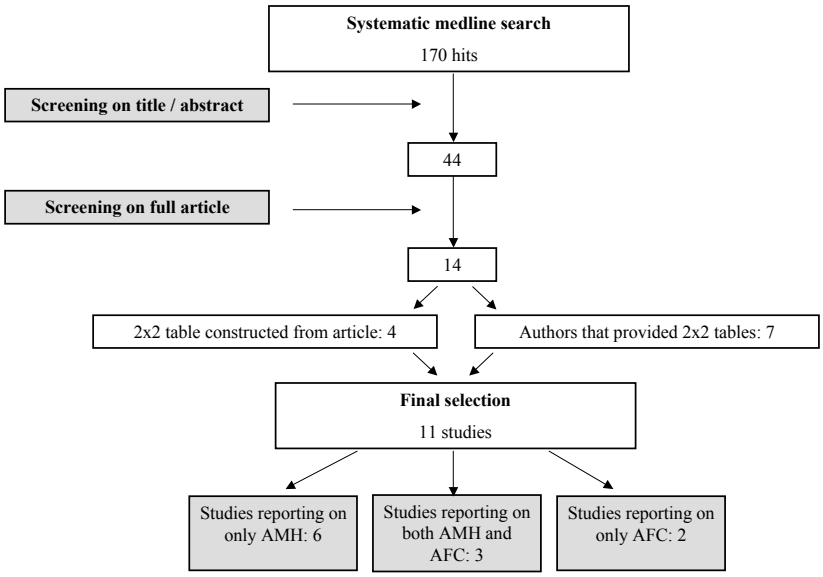


Figure 1. Search and selection strategy

Using the bivariate model that accounts for the heterogeneity of the studies the summary estimates for sensitivity and specificity were calculated. The summary estimates were 82% (95%CI 52% to 95%) for sensitivity and 76 % (95%CI 43% to 93%) for specificity.

Figure 2 shows the summary estimate for the overall test accuracy as calculated from the bivariate model and its 95%CI ellipse, as well as the summary ROC curve.

Accuracy of AFC in excessive response prediction

Sensitivities and specificities for the prediction of an excessive ovarian response, as calculated from each study reporting on AFC are summarized in *Table 2b*. A plot of sensitivity-specificity combinations in an ROC space is shown in *Figure 2*. For the AFC, the sensitivity varied between 20% and 94% and specificity between 33% and 98%. Using the bivariate model that accounts for the heterogeneity of the studies the summary estimates for sensitivity and specificity were calculated. The summary estimate of sensitivity was 82% (95%CI 30% to 98%) and the summary estimate of specificity was 80 % (95%CI 31 % to 97 %).

Figure 2 shows the summary estimates as calculated by the bivariate model and its 95%CI ellipse, as well as the summary ROC curve for the AFC in the prediction of an excessive response.

Table 1. Characteristics of the included studies

Author	Test	Consecutive	Cohort / Case control	Pro-/Retro spective	Blinding	Selection-bias	Verification-bias	One cycle per couple	Data per cycle	AMH assay / AFC count	Definition of excessive response
Nelson 2007	AMH	Yes	Cohort	Prospective	No	Yes	No	Yes	Yes	DSL	≥21 oocytes
La Marca 2007	AMH	Yes	Cohort	Prospective	No	Yes	Yes	Yes	Yes	BC	>16 oocytes
Ebner 2006	AMH	Yes	Cohort	Prospective	No	Yes	Yes	Yes	Yes	BC	≥15 oocytes
Lee 2008	AMH	Yes	Cohort	Prospective	No	Yes	No	Yes	Yes	DSL	OHSS
Riggs 2008	AMH	Yes	Cohort	Retrospective	Yes	Yes	Yes	No	Yes	DSL	≥15 oocytes
Nardo 2009	AMH	Yes	Cohort	Prospective	No	Yes	Yes	Yes	Yes	DSL	≥20 oocytes, E2 >17,000
Aflatoonian 2009	AMH & AFC	Yes	Cohort	Prospective	No	Yes	No	Yes	Yes	BC, AFC 2-6mm	≥15 oocytes, E2>3000 pg/ml
van Rooij 2002	AMH & AFC	Yes	Cohort	Prospective	No	Yes	Yes	Yes	Yes	BC, AFC 2-5mm	≥14 oocytes
Eldar Geva 2005	AMH & AFC	Yes	Cohort	Prospective	No	Yes	No	Yes	Yes	BC, AFC 2-10mm	≥14 oocytes
Kwee 2008	AFC	Yes	Cohort	Prospective	No	Yes	Yes	Yes	Yes	AFC 2-10mm	>20 oocytes
Ng 2000	AFC	Yes	Cohort	Prospective	No	Yes	No	Yes	Yes	NS	≥15 oocytes

DSL = Diagnostic System Laboratories, BC = Beckman-Coulter, NS = not stated

Table 2a. Performance of AMH in the prediction of excessive response

Author	Cycles (n)	AMH (ng/ml) cut-off	Abnormal test result (%)	Sensitivity	Specificity	LR+	Pre-test probability	Post-test Probability
van Rooij 2002	114	3.50	0.08	0.40	0.95	8.32	0.09	0.44
Eldar-Geva 2005	53	3.50	0.32	0.72	0.89	6.32	0.34	0.76
Ebner 2006	135	1.66	0.75	0.95	0.31	1.38	0.16	0.21
	135	4.52	0.25	0.55	0.81	2.80	0.16	0.35
La Marca 2007	48	2.60	0.50	0.86	0.56	1.95	0.15	0.25
	48	7,00	0.23	0.57	0.83	3.35	0.15	0.36
Nelson 2007	314	2.10	0.27	0.88	0.79	4.10	0.08	0.26
	314	3.50	0.08	0.57	0.96	13.8	0.07	0.52
Lee 2008	262	1.99	0.50	0.90	0.62	2.38	0.23	0.42
	262	3.36	0.25	0.62	0.87	4.64	0.23	0.58
Riggs 2008	123	1.59	0.49	0.84	0.67	2.56	0.31	0.53
Nardo 2009	165	3.50	0.36	0.88	0.70	2.90	0.10	0.24
Aflatoonian 2009	159	4.83	0.42	0.93	0.78	4.26	0.28	0.63

Table 2b. Performance of AFC in the prediction of excessive response

Author	Cycles (n)	AMH (ng/ml) cut-off	Abnormal test result (%)	Sensitivity	Specificity	LR+	Pre-test probability	Post-test Probability
Ng 2000	128	9	0.39	0.60	0.71	2.09	0.31	0.49
	128	14	0.10	0.20	0.94	3.48	0.31	0.62
van Rooij 2002	114	14	0.42	0.92	0.63	2.49	0.1	0.22
Eldar-Geva 2005	56	14	0.78	0.94	0.33	1.14	0.4	0.48
Kwee 2008	110	10	0.38	0.94	0.71	3.26	0.15	0.36
	110	12	0.30	0.88	0.80	4.33	0.15	0.42
	110	14	0.21	0.81	0.89	7.64	0.15	0.57
	110	16	0.11	0.50	0.96	11.75	0.15	0.67
	110	18	0.06	0.31	0.98	14.69	0.15	0.71
Aflatoonian 2009	159	16	0.31	0.89	0.92	11.26	0.28	0.82

If a study reported on multiple cut-off values, data for all cut-off values are shown. LR+ = likelihood ratio for a positive test result.

Clinical value

Based on the summary ROC curves depicted in *Figure 2*, a range of positive likelihood ratios was calculated corresponding to various sensitivity-specificity points on this ROC curves. For each of these likelihood ratio values, the pre AMH or AFC test probabilities of an excessive response were converted into post-test probabilities of an excessive response. *Table 3* depicts a series of likelihood ratios ranges and the probability of obtaining an abnormal test result for AMH or AFC corresponding to this likelihood ratio range, as well as the post test probability of an excessive response. At a positive likelihood ratio of at least ~8, the post test probability of having an excessive response is close to 70%, if the pre test probability is assumed to be ~ 20%. The probability of obtaining a test result for AMH or AFC with a likelihood ratio of at least ~8 is 14% and 16%, respectively.

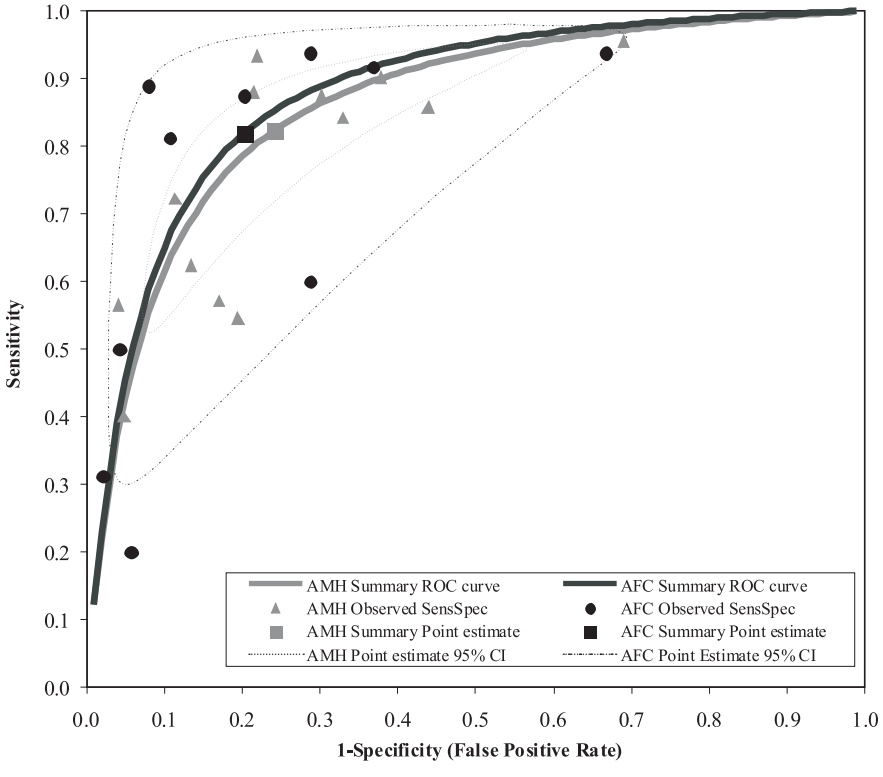


Figure 2. AMH and AFC in the prediction of an excessive response
Regardless of the number of cut-off mentioned per study, only one cut-off was taken into analysis. For the observed values of sensitivity-specificity points, all cut-offs are displayed.

Table 3. Clinical value of AMH and AFC in the prediction of an excessive response

Prediction of an excessive response (Pre-test probability 20%)			
	Occurrence of abnormal test result in LR range (%)		Post-test probability of excessive response (%)
LR range	*AMH*	*AFC*	
<2	44	43	20-33
2-3	16	16	33-43
3-4	9	8	43-50
4-5	5	5	50-56
5-6	5	5	56-60
6-7	4	4	60-64
7-8	3	3	64-67
>8	14	16	> 67

Shown is the occurrence of both Anti-Müllerian Hormone (AMH) and Antral Follicle Count (AFC) results within a specified likelihood ratio (LR) range and the concomitant post-test probabilities of an excessive response, given a prevalence of an excessive response of 20%. For example, at a positive likelihood ratio of at least 8, the post test probability is 70% if the prior chance of having an excessive response is 20%. With the cut off levels for the test corresponding to these LR+ levels the proportion of abnormal tests is 14% for the AMH and 16% for the AFC.

Comparison of AMH with AFC

Comparison of summary point estimates for accuracy of the prediction of excessive response showed no statistically significant difference in the performance for AMH compared with the AFC, when sensitivity ($p = 0.87$) and specificity ($p = 0.80$) at the estimated summary cut-off point were considered. In the comparison of the estimated summary ROC-curves, AFC seemed to perform slightly better than AMH, although the curves did not differ statistically. It should be noted that, the summary curve for the AFC was based on fewer studies (*Figure 2*).

Clinical value as outlined in *Table 3* indicated a similar performance for AMH compared to the AFC. This is in line with the course of the ROC curves along the y-axis suggesting that many cases of an excessive response can be identified with only a limited number of false positives. For both AMH and AFC, sensitivity can amount up to 70% with a false positive rate of 15%, and this performance level will imply a realistic number of abnormal tests (~25%).

Discussion

Main Findings

The current meta-analysis summarizes the available evidence concerning the accuracy of AMH and the AFC in the prediction of excessive ovarian response to stimulation for IVF. It appears that both tests have a good discriminatory capacity to separate normal and excessive responders, with a definition that varies across the studies from more than 14-21 oocytes yielded. From the ROC curves (*Figure 2*) it becomes clear that, currently, AMH and AFC have an equal level of accuracy in the prediction of an excessive response and that there is no statistical difference between both tests. Moreover, both AMH and AFC have clinical value, with an abnormal test rate of 14% and 16%, respectively, at cut off levels where test performance is optimal (LR+>8). At these cut-off levels the post test probability of an excessive response appeared to be close to 70%.

The comparison between the AFC and AMH for their use as predictive tests for ovarian response may imply other factors than accuracy alone. For AMH as a laboratory test, measurement stability will be dealt with according to routine procedures, but routine assays may not yet be easily available. To date, two commercially available immuno-assays, the Beckmann Coulter and DSL ELISA, exist. These assays have demonstrated a very good correlation, making it possible to translate results from on to the other within the same dataset. However, there is no obvious match in absolute levels between studies; with the Beckmann-Coulter measurements reported as being approximately 4-5 times higher than the DSL measurements (90;173). Therefore, standardization of these assays is urgently needed. This situation also hampers the efforts to extract a generally applicable cut off level for deciding who will be a predicted excessive responder. But, AMH is a cycle independent test (54-56;92) any measurement in the period before starting the ART cycle will be at the disposition of the clinician, making the test an ideal tool. For the AFC, standardization needs to be dealt with by the physician (41), implying choices on ultrasound equipment, dedicated personnel and a systematic visualization and counting process. As the intra- and between cycle stability for the AFC may be comparable to that for AMH (174), the unlimited availability of this test makes it the preferable one for the short term. Currently, the identification of patients at risk for excessive response is based on a variety of factors, such as age, body weight and the presence of polycystic ovaries (91;168;175). However, the predictive value of these factors is quite poor. Whether their addition to tests like the AFC and AMH will improve the predictive capacity in identifying excessive responders remains to be established. In fact, patients with the polycystic ovary syndrome have clearly elevated AFC's

and AMH levels (42;176). As most studies have not excluded PCOS cases, these cases will add to the current analysis. In studies on PCOS cases only, AMH levels do indeed predict ovarian response to controlled hyperstimulation (177). A limited number of studies exist on the use of multifactor prediction of the number of oocytes retrieved, using female age, basal FSH, ovarian ultrasound and smoking behavior as predictors (131;178). Validation of these prediction models in external populations has not been carried out so far. Individual patient data analyses of published literature may enable the assessment of the true value of such multivariable approach, combining patient characteristics and test results (113).

Implications for clinical practice

Excessive response to ovarian stimulation induces the risk of the ovarian hyperstimulation syndrome, especially in cases where exaggerated response is followed by a pregnancy. It may also cause increased patient discomfort and even reduced prospects for pregnancy. Up to 30% of IVF cycles are accompanied by complaints of mild or moderate OHSS and in 3-8% the severe form of OHSS may develop (159). Once an excessive response has occurred, hCG administration could be withheld in an effort to eliminate this risk. Protective measures have also been reported for conditions in which oocyte retrieval has been allowed. Albumin administered at the time of oocyte retrieval, elective cryopreservation of all embryos to prevent the occurrence of pregnancy in the fresh cycle, GnRH agonist use for endogenous LH triggered ovulation in gonadotrophins / GnRH antagonist cycles and of the use of a single-dose recombinant LH to trigger ovulation have all been proposed (179-182). These measures may limit collateral damage linked to excessive response, but they certainly do not offer absolute protection. The prevention of excessive ovarian response may be considered the corner stone of preventive management for OHSS, as such responses add heavily to the risk of developing the syndrome (183).

Reduction of pregnancy chances in excessive responders is most likely caused by detrimental effects of concomitant supraphysiological hormone levels on oocytes and embryo quality (184-186). Moreover, exaggerated and untimely estrogen and progesterone concentrations will affect the orderly proliferation and subsequent luteinisation of the endometrium and thereby its receptivity (187-189). Moreover, an excessive ovarian response results in the yield of additional immature oocytes that are likely to be of insufficient quality to result in conception (99;190).

Prior information on the expected ovarian response may allow the application of individualized stimulation protocols that will mitigate the number of

follicles growing. The ideal test for excessive response prediction would identify all women with an excessive response and exclude all those women with a normal or poor response to standard dose stimulation. These women could be given individualized, milder treatment regimens, ensuring a yield of oocytes between 5 and 12 oocytes (191). In reality, tests like the AFC and AMH will never be absolutely accurate in their prediction. Assuming that a cut off can be used at which 75% of excessive responders will be identified, a considerable number of excessive responders will be turned into normal responders by using for instance a lower than standard dose of FSH. At the same time, the abnormal test will be falsely positive in some 15% of cases, and a lower dose may turn these cases into poor responders. Whether this "poor" response may alter the prospects for pregnancy may be disputed, as mild stimulation protocols have demonstrated that in normal profile cases a mild response does not affect outcome (150;161;192;193).

Currently, only few studies have addressed the use of reduced dosages of FSH based on prior prediction of the ovarian response level. In the study by Popovic et al., individualized FSH dosing appeared not to reduce the proportion of excessive responders, although the reduced dose group produced on average 2 oocytes less than the standard stimulated group (101). In one other study, Olivennes et al., demonstrated that in predicted excessive responders the use of FSH dosages lower than 150 IU produced mild ovarian responses without compromising pregnancy rates. A randomized comparison of standard versus individualized treatment based on the CONSORT prediction algorithm has recently been finalized and results are awaited (150). Such studies should not only focus on the achievement of a more homogenous ovarian response. Also, cost-economic effects regarding prevention of severe OHSS and a reduction of FSH consumption will aid in rationalizing ovarian stimulation protocols for IVF.

Limitations

Although the process of systematic literature review and meta-analysis is a practical way to generate a more powerful estimate of true effect-size with less random error than individual studies, it does come with some limitations. First of all, the heterogeneity of studies must be addressed, as it may affect the justification for pooling the data into one analysis. In the case of the present meta-analysis, heterogeneity was caused by both different study quality characteristics and slight differences in study populations. Additionally, the definition of excessive response was not uniform across studies (*Table 1*) and varied from the use of a threshold for number of oocytes aspirated to the development of OHSS. Another limitation was the allocation, by the authors, of different cut-off

values for AMH and AFC. This is problematic as it interferes with the identification of a single threshold for AMH of AFC that could be predictive of an excessive response. The solution for this problem is the construction of a ROC curve, by which the effect of different cut offs on the sensitivity/specificity combinations will become clear and overall accuracy becomes apparent.
Many of these methodological problems may be overcome by using individual patient data meta-analysis. From such data sets, population and patient characteristics, test results, stimulation data and outcome variables can be uniformed as much as possible before applying meta-analysis. Currently, initiatives in this field have been employed (113).
Lastly, there are some limitations that apply specifically to the method used to assess AMH levels and the AFC. The studies in this meta-analysis did not all use the same AMH assay. There is a noteworthy difference between the Beckman-Coulter ELISA and the Diagnostic System Laboratories (DSL) ELISA leading to a wide dispersion of AMH concentrations (90). This compatibility problem can only be overcome by the development of an internationally standardized AMH assay (90). Similar problems arise with the use of AFC results, where either follicle sizes of 2-5 or 2-10 mm are included into the counts. Although both methods of measurement may deliver the same level of accuracy for the test, it certainly will hamper the identification of a generally applicable cut off.

Future Research

The role of ovarian reserve tests in excessive response prediction combined with simple patient characteristics could be further analyzed by using large individual patient data sets. The EXPORT (individual meta-analysis of patient data for Excessive Response Prediction with Ovarian Reserve Tests) initiative may offer the opportunity to start such effort. Moreover, stimulation protocols tailored on the basis of ovarian response prediction should be analyzed as to their effects on pregnancy rates, costs for medication and patient satisfaction. Only large randomized comparisons of standard treatment strategies versus individualized treatment approaches will provide the correct answers, and will enforce previous undertakings in this area (101).

Summary

The current systematic review and meta-analysis demonstrates that both the AFC and AMH are capable of identifying excessive responders to ovarian stimulation for IVF. Test optimization for clinical application may be more promising for AMH**.**

Chapter 6

Prediction of an excessive response from patient characteristics and ovarian reserve tests and comparison in subgroups: an Individual Patient Data meta-analysis

The EXPORT* studygroup

S.L. Broer, M. Dólleman, J. van Disseldorp, K.A. Broeze, B.C. Opmeer, A. Aflatoonian, R.A. Anderson, M. Ashrafi, L. Bancsi, E. Caroppo, A.B. Copperman, T. Ebner, T. Eldar-Geva, M. Erdem, T. Fréour, C. Gnoth, E.M. Greenblatt, K. Jayaprakasan, N. Raine-Fenning, E. Klinkert, J. Kwee, A. La Marca, M. McIlveen, L.T. Merce, S. Muttukrishna, L. Nardo, S.M. Nelson, H.Y. Ng, B. Popovic-Todorovic, J.M.J. Smeenk, C. Tomás, P.J.Q. Van der Linden, I.K. Vladimirov, P. Bossuyt, M.J.C. Eijkemans, B.W. Mol and F.J.M. Broekmans

** Excessive Response Prediction using Ovarian Reserve Tests*

Submitted

Abstract

Introduction

An excessive response to ovarian hyperstimulation is highly related to increased patient discomfort and complications like OHSS. Ovarian Reserve Tests (ORTs) are capable of prior identification of excessive responders. However, it is unclear whether they add prognostic value to readily available patient characteristic, like female age. The performance in different clinical subgroups has also not been assessed properly. This study evaluates the added value of ORTs to patient characteristics and studies the predictive capacity of ORTs in different clinical relevant subgroups.

Methods

Studies published until December 2009 regarding basal FSH, AMH or AFC in relation to ovarian response to hyperstimulation for IVF were searched. The authors of these studies were asked to join this Individual Patient Data (IPD) meta-analysis and were invited to share their data. Random intercept logistic regression models were used to correct for heterogeneity between studies and to quantitatively estimate the added value of the ORTs on basic patient characteristics. ROC regression analyses were performed to study the effect of specific patient characteristics on the discriminatory capacity of the ORTs.

Results

33 databases were included, regarding 6,852 women undergoing IVF. Age had an area under the ROC curve (AUC) of 0.61 for excessive response prediction. Both AFC and AMH clearly and significantly added prognostic value to age (P-value for each <0.001). A model with age, AFC and AMH had an AUC of 0.85. Similar accuracy was also reached by the combination of AMH and AFC, without the addition of age (P=0.98). The subgroup analysis showed that age was the only characteristic which significantly influenced the accuracy of AFC and FSH (P=0.010). The accuracy of AMH was not influenced by age.

Conclusion

This IPD meta-analysis demonstrates that the ORTs AFC and AMH add value to age alone in the prediction of an excessive response and that these tests in conjunction have a similar accuracy. An obvious improvement or decline of the performance of ORTs in one of the specific subgroups could not be found. This implies that ORTs remain applicable for excessive response prediction in unselected populations indicated for IVF.

Introduction

In women undergoing in vitro fertilization (IVF) treatment, the development of a large number of oocytes complicates up to thirty percent of IVF cycles (159). Such an excessive response may lead to poorer quality embryos, decreased chances of pregnancy or cycle cancellation (99;161;163;194). Additionally, the patient is at risk of developing ovarian hyperstimulation syndrome (OHSS), a potentially life threatening condition (160). In order to maximize safety and efficacy of assisted reproductive technology (ART) programs, there is a need to identify patients at risk of an excessive response at the start of IVF/ICSI treatment, with the possible application of measures to prevent an excessive response.

Several patient characteristics such as a lean habitus, young age and the presence of polycystic ovary syndrome (PCOS) have been identified as conditions that predispose patients to OHSS (195). Unfortunately, systematic studies on the predictive accuracy of these characteristics are lacking. In contrast, ovarian reserve tests (ORTs), such as Anti-Müllerian Hormone (AMH), Antral Follicle Count (AFC) and Follicle Stimulation Hormone (FSH) have been assessed for their value in the prediction of an excessive response (73;84;151;166-168). However, the added value of the ORTs on patient characteristics has not been evaluated.

Moreover, no previous studies have been performed in which these tests are evaluated in clinical subgroups. It is conceivable that age influences these tests since ovarian reserve also decreases with age. The accuracy of the AFC measurements could be complicated by a higher BMI. BMI could also influence the accuracy by possibly reducing the biologic availability of recombinant FSH for ovarian stimulation. Most studies, however, have a limited sample size and either lack the power to evaluate patient characteristics as determinants of accuracy in specific subgroups, or they fail to analyze the added value of the tests on patient characteristics.

To overcome the problem of small studies with restricted power, the current study used an individual patient database (IPD) meta-analysis approach. By aggregating data on the level of the individual patient, reliable estimates of accuracy could be made while being able to explore and correct for heterogeneity. In the present IPD meta-analysis we were able to study the added value of ORTs on the predictive capacity of basic patient characteristics in excessive response prediction. Furthermore, to assess whether the predictive capacity of ORTs is influenced by patient characteristics, we assessed the discriminatory capacity of ORTs in clinical subgroups defined by female age, BMI and duration of subfertility.

Methods

Data acquisition

The current database used is an expanded version of a database used for a previous IPD study, namely IPD-IMPORT, focusing on poor response and ongoing pregnancy prediction (196;197). First, all collaborators of the IPD-IMPORT consortium were asked for permission to use their data for the current IPD meta-analysis focusing on excessive response prediction. Concurrently, the search used for the IPD-IMPORT study was updated until the end of 2009. A systematic search in Medline was carried out using synonyms for In Vitro Fertilization (IVF, controlled ovarian stimulation, in vitro fertilisation) and synonyms for the various tests (FSH, Follicle Stimulating Hormone, AFC, Antral Follicle Count or number, AMH, Anti-Müllerian Hormone, Müllerian inhibiting substance) as keywords.

All additional titles and abstracts were evaluated by two authors (MD and SB) for eligibility. If necessary, the opinion of a third author was decisive (FB). Studies presenting data on ovarian response, at least one ovarian reserve test (ORT) and at least one patient characteristic were eligible for the current review. The authors of identified primary studies that met our eligibility criteria were approached and informed about this collaborative IPD meta-analysis project, and invited to share their data. If they were inclined to participate, they were provided with an author's agreement form to participate and a data request form, informing them on the format of the requested data.

After data acquisition, all data were scrutinized on quality and consistency and when possible converted into a single format. Any issues or inconsistencies were checked with the original author. For a more detailed description of IPD meta-analysis methodology the reader is referred to previous papers (113;115).

A comparison was made between the studies that could be included and could not be included for analysis. If possible, sensitivity and specificity of the ORTs in the prediction of an excessive response were calculated for all studies in these two groups. A Spearman correlation was calculated for every ORT, to study the association between sensitivity and specificity and to test whether the differences in sensitivity and specificity were the result of different threshold levels. The Spearman correlations of each ORT and outcome were then compared between these groups, to see whether these groups were comparable.

Statistical Analysis

An excessive response was defined as the retrieval of more than 15 oocytes. Duration of subfertility was defined as the period from the cessation of oral contraceptives or start of unprotected intercourse until the first IVF attempt. When a particular variable was missing in an individual database, data were not imputed. Baseline characteristics were analyzed for the total data and for each of the individual studies. The AUCs of the ORTs in the prediction of an excessive response in each individual study were calculated.

Added value of ORTs on patient characteristics

To study the added value of the ORTs on patient characteristics the following analyses were performed. Random intercept logistic regression prediction models were created with the 'Lme4' library in R (version 2.9.0. (http://www.r-project.org/), using the Laplace approximation to the likelihood. These models were created to quantitatively estimate the added value that ORTs have on patient characteristics in predicting an excessive response. By using a random intercept, the heterogeneity in prevalence of excessive response could be corrected for.

The three different sets of models that were used for the prediction of an excessive response are described below. The first model included the patient characteristics female age, BMI and duration of subfertility. In the second set of models, the predictive capacity of individual ovarian reserve tests (FSH, AFC and AMH) in combination with significant patient characteristics was estimated. In the third set of models, the added value of combinations of ovarian reserve tests on patient characteristics was evaluated.

The next step was to construct receiver operating characteristic (ROC) curves. Using the random intercept logistic regression models, probabilities of excessive response could be calculated. Based on these, we plotted stratified ROC curves with the ROC regression model as proposed by Janes and Pepe (116;117). This model assumes that studies share a common ROC for each ORT, but allows the positivity threshold corresponding to each sensitivity-specificity pair to vary between studies. With this model the improvement in predictive accuracy of adding an ORT to other variables can be studied, while correcting for the heterogeneity between studies. This way, we could compare the ROC curves and Area Under the Curves (AUCs) of the models described above and evaluate them for statistically significant differences.

Because not all studies in this meta-analysis would report data of all three ORT, we constructed the prediction models using those databases from the total dataset that included the three ovarian reserve tests (FSH, AFC and AMH) and

age to allow for a direct comparison. The results of all analyses in the three-test study group were checked in the total study group.

Accuracy of ORTs in different subgroups

To study the accuracy of the ORTs in the prediction of an excessive response in different clinical subgroups, the following analyses were performed. The ROC regression model proposed by Pepe and Janes was applied (116;117). In this model the ROC curves of the ORTs are modeled as a function of the covariates (age, BMI or duration of subfertility), since these patient characteristics can impact the inherent discriminatory accuracy of the ovarian reserve tests. This regression analysis resulted in an estimate that reflects the impact of the patient characteristics on the ROC curve and corresponding AUC of the ovarian reserve tests. In this meta-analysis, we assumed the effect of the covariate to be identical across studies, but, as in the previous analysis, the positivity threshold corresponding to each sensitivity-specificity pair was allowed to vary between studies, and therefore heterogeneity between studies is corrected for.

As a visual illustration of the results of the subgroup differences, we drew the ROC curves in subgroups defined by female age, BMI and duration of subfertility categories. The corresponding AUCs were calculated in order to express the overall discriminatory capacity (accuracy) of the ORT in women in the respective subgroups.

Data were analyzed using SPSS 17.0 (SPSS Inc., Chicago, Il, USA) and R version 2.9.0. (http://www.r-project.org/).

Results

Data acquisition

A total of 33 databases, that were used for the preparation of 60 manuscripts were included for analysis in this IPD-study. Twenty-eight databases were used from the IPD-IMPORT study (196;197). Ten additional studies were identified from the systematic MEDLINE search. These authors were approached for permission to use their databases in the present study on excessive response prediction. Of these authors only four sent their data (90;167;172;198) and one of them submitted two separate databases (167). Thus, a total of 33 datasets contributed to the overall study database, comprising a number of 6,852 study cases (*Figure 1*).

Study characteristics in terms of sampling, data collection and study design are shown in *Table A-I Addendum*. With the original data we were able to replicate the primary findings of the original study in 14 databases. In 12 cases, the study database received contained a number of patients that differed from the pub-

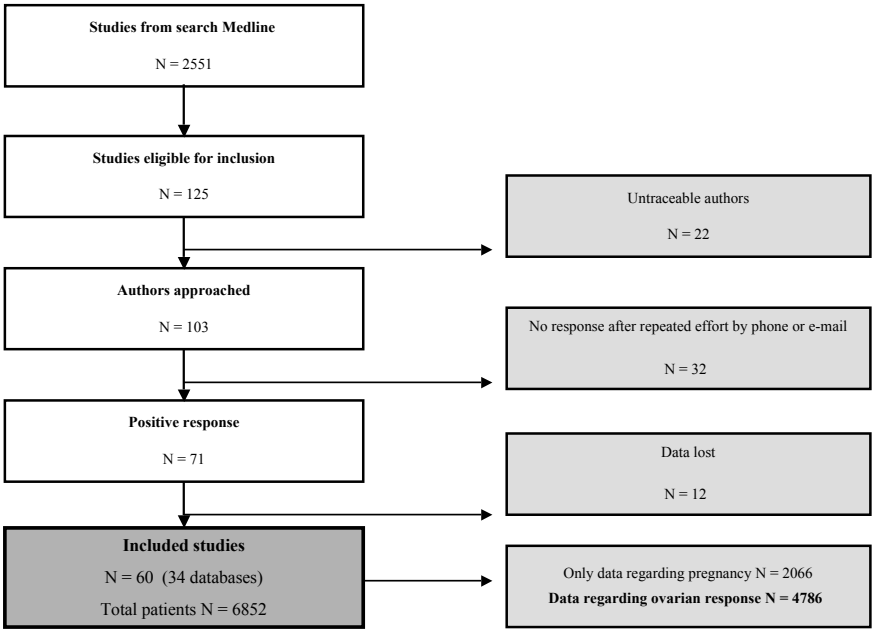


Figure 1. Flowchart of included studies

lication, whereas in seven other databases there were slight inconsistencies in the baseline data previously published. These inconsistencies were checked and discussed with the corresponding author and could be resolved in a vast majority of cases. Finally the level of consistency between the individual data and the data reported in the published manuscript became sufficient for all included studies.

Data from 4,786 out of the 6,582 women were suitable for the analysis of prediction of excessive response. The remaining women were originally included in the IMPORT study to asses the accuracy of the ORTs in the prediction of an ongoing pregnancy (196;197). Unfortunately, no information about the ovarian response, in terms of number of oocytes, was provided for these women (*Figure 1*). Of the 4,786 women, 894 (18.7%) had an excessive response. Baseline characteristics of the total study group are summarized in *Table 1*. Baseline characteristics of the original studies are summarized in *Table A-II Addendum*. The AUCs of all individual studies in the prediction of an excessive response are shown in *Table A-II Addendum*.

Table 1. Baseline characteristics

	Total population	Excessive Responders	Non-excessive responder	P-value
	Mean (5th–95th percentile)	*Mean (5th–95th percentile)*	*Mean (5th–95th percentile)*	
Female age (years)	34.1 (26.0-41.5)	32.3 (25.0-39.3)	34.7 (27.0-42.1)	< 0.001
FSH (IU/L)	7.7 (3.8-14.0)	6.2 (3.4-9.3)	7.9 (3.8-14.0)	< 0.001
AFC (number)	12.1 (3.0-25.0)	17.7 (7.0-32.0)	11.0 (3.0-22.0)	< 0.001
AMH (ng/ml)	2.5 (0.1-7.5)	4.8 (1.3-10.2)	2.0 (0.1-5.7)	< 0.001
BMI (kg/m^2)	23.6 (18.6-30.1)	23.4 (18.5-29.4)	23.4 (18.6-30.1)	0.943
Duration of subfertility (years)	4.1 (1.0-10.0)	4.3 (1.5-10.0)	4.3 (1.8-10.0)	0.937

Excessive Response definition: > 15 oocytes retrieved. Duration of subfertility: the period from the cessation of contraceptive methods or start of unprotected intercourse until the first IVF attempt.
Excessive responders N = 894 (18.7%). Non excessive responders = 3,892.
AFC, Antral Follicle Count; AMH, Anti-Müllerian Hormone; FSH, Follicle Stimulating Hormone.

For the comparison of the included and not included studies, we aimed to calculate sensitivity and specificity of the ORTs in the prediction of an excessive response. Of the not included studies only one reported sensitivity and specificity values for AFC in the prediction of an excessive response. Therefore, Spearman correlation could not be calculated. For the majority of the studies this was performed in the IMPORT study (196;197) which showed that there was no difference. Since there is no difference in poor response prediction, it is reasonable to assume that there is no difference for excessive response prediction.

Statistical analyses

Added value of ORTs on patient characteristics

For the model building exercises we could use data of 1,023 women for excessive response analysis. Of all patient characteristics, age was the strongest single predictor of an excessive response (OR 0.89: 95%CI 0.85 to 0.93). BMI and duration of subfertility were not significantly predictive of an excessive response (*Addendum Table A-IV*).

We compared the ORTs using the random intercept logistic regression model in predicting excessive response (*Table 2*). The ROC regression analysis showed a high accuracy for AMH (AUC 0.81: 95%CI 0.76 to 0.87) and for AFC (AUC 0.79: 95%CI 0.74 to 0.84), but only a moderate accuracy for FSH (AUC 0.66: 95%CI 0.60 to 0.73) (*Table 3*).

Table 2. Univariable and multivariable models of age and ORTs in the prediction of an excessive response

Excessive Response Prediction						
	Three-test study group			**Total study group**		
	OR	*95% CI*	*P - value*	*OR*	*95% CI*	*P - value*
Univariable models						
Age (per year)	0.89	0.85 - 0.93	< 0.001	0.90	0.88 - 0.91	< 0.001
FSH (per IU/L)	0.76	0.70 - 0.84	< 0.001	0.83	0.80 - 0.86	< 0.001
AFC (per N)	1.18	1.15 - 1.22	< 0.001	1.14	1.12 - 1.16	< 0.001
AMH (per ng/ml)	1.61	1.48 - 1.76	< 0.001	1.59	1.49 - 1.70	< 0.001
Multivariable models						
Age and FSH						
Age (per year)	0.91	0.87 - 0.94	< 0.001	0.91	0.89 - 0.93	< 0.001
FSH (per IU/L)	0.79	0.72 - 0.87	< 0.001	0.85	0.82 - 088	< 0.001
Age and AFC						
Age (per year)	0.93	0.89 - 0.98	0.003	0.95	0.92 - 0.98	0.001
AFC (per N)	1.17	1.13 - 1.21	< 0.001	1.13	1.11 - 1.15	< 0.001
Age and AMH						
Age (per year)	0.92	0.88 - 0.97	< 0.001	0.92	0.89 - 0.95	< 0.001
AMH (per ng/ml)	1.57	1.43 - 1.71	< 0.001	1.54	1.44 - 1.64	< 0.001

Results of random intercept logistic regression model in the prediction of an excessive response. Multivariable analyses showed that all three ORTs add predictive information to female age alone. P- values reflect whether the variable plays a significant role in the model. NB three-test study group N = 1,023, total study group N=4,786. OR (Odds Ratio), 95% CI (95% Confidence Interval).

The multivariable analyses showed that a model including age, AFC and AMH (AUC 0.85) had a significantly higher predictive accuracy than a model based on age alone (AUC 0.61). Addition of FSH to this model did not further improve the predictive accuracy (P = 0.73) (*Table 3*). Interestingly, a single test of AMH or AFC already yielded a comparable accuracy (0.81 and 0.79, respectively). Addition of AMH to AFC or the other way around is significant (P = <0.001 or P = 0.003, respectively). Consequently, a model combining these two tests resulted in an AUC of 0.85, and age did not add to this model (P = 0.98). The ROC curves corresponding to the multivariable analyses are shown in *Figure 2*.

Accuracy of ORTs in different subgroups defined by age, BMI or duration of subfertility

The results of the ROC regression model which studied the effect of several patient characteristics on the ROC curve of the ORTs in the prediction of an excessive response are shown in *Table 4*. The accuracy of AFC was significantly higher

Table 3. AUCs of prediction models of age and ovarian reserve tests for the prediction of an excessive response

	Three-test study group				Total study group			
	AUC	*95% CI*	*P-value*	*N*	*AUC*	*95% CI*	*P-value*	*N*
Univariable analysis								
Age	0.61	0.54 - 0.68	NA	1023	0.61	0.58 - 0.64	NA	4650
FSH	0.66	0.60 - 0.73	0.071	1023	0.64	0.61 - 0.67	0.026	4254
AFC	0.79	0.74 - 0.85	< 0.001	1023	0.73	0.69 - 0.77	< 0.001	2524
AMH	0.81	0.76 - 0.87	< 0.001	1023	0.82	0.77 - 0.86	< 0.001	1890
Multivariable analysis								
Age & FSH	0.68	0.62 - 0.75	< 0.001	1023	0.67	0.64 - 0.71	< 0.001	4254
Age & AFC	0.81	0.76 - 0.87	< 0.001	1023	0.75	0.71 - 0.79	< 0.001	2524
Age & AMH	0.81	0.76 - 0.87	< 0.001	1023	0.81	0.77 - 0.85	< 0.001	1890
Age & AMH & AFC	0.85	0.80 - 0.90	< 0.001	1023	0.85	0.80 - 0.90	< 0.001	1024
Age & AMH & AFC & FSH	0.85	0.80 - 0.90	< 0.001	1023	0.85	0.80 - 0.90	< 0.001	1023
AMH & AFC	0.85	0.80 - 0.90	< 0.001	1023	0.85	0.80 - 0.90	< 0.001	1024

The Area Under the Curve (AUC) of the univariable and multivariable models of age or ORTs in the prediction of an excessive response are shown. In the univariable analysis it is shown that both AMH and AFC have a high accuracy, while FSH only has a moderate accuracy. In the multivariable models the added value to the AUC of an ORT on female age is shown, the P-value indicates whether this added value is significant in comparison to the model based on age alone. Adding any of the ORTs shows a significant rise in the AUC. Moreover, the added value of adding several ORTs to female age is shown. The model including age, AFC and AMH reached the maximum predictive power. Addition of FSH to this model did not improve the predictive accuracy (P = 0.725). However, a model with AMH and AFC alone has a comparable AUC.

with increasing age (P = 0.010), while the accuracy of FSH was significantly lower with increasing age (P = 0.010). The discriminatory capacity of AMH in response prediction was not influenced by age. BMI and duration of subfertility had no significant effect on the ROC curves for any of the ORTs.

To illustrate the influence of patient characteristics on the predictive capacity, we performed an ROC curve analysis in three sets of clinical subgroups, while taking the heterogeneity between studies into account (116;117). The AUCs of the three ORTs in the prediction of an excessive response in subgroups defined by age, BMI and duration of subfertility are presented in *Table 5*.

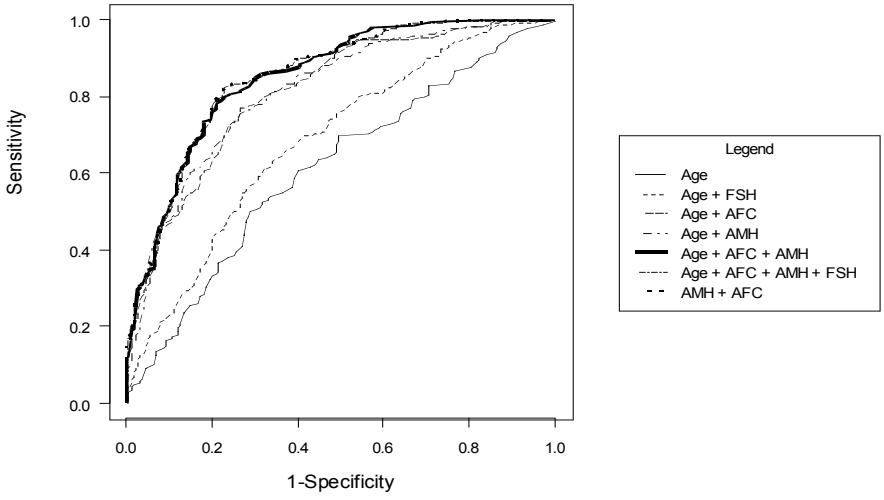


Figure 2. ROC curves of age and ORTs in the prediction of an excessive response
The ROC curves of age and age combined with a single or more ORTs are depicted. The ROC curves for 'Age + AMH', 'Age + AFC', 'Age + AMH + AFC' and 'Age + AMH + AFC + FSH' run toward the upper left corner of the ROC space, indicating a good capacity to discriminate between normal and excessive responders at certain cut-off levels. NB ROC curves in the three-test study group (N = 1023). AFC, Antral Follicle Count; AMH, Anti-Müllerian Hormone; FSH, Follicle Stimulating Hormone; ORT, Ovarian Reserve Test; ROC, receiver-operating characteristic.

Table 4. Results of the ROC regression analysis

	Coefficient	*95% CI*	*P-value*
		Age	
FSH	-0.029	-0.051 - -0.006	**0.010**
AFC	0.032	0.006 - 0.056	**0.010**
AMH	-0.021	-0.049 - 0.005	0.139
		BMI	
FSH	0.026	-0.024 - 0.070	0.267
AFC	-0.009	-0.048 - 0.033	0.674
AMH	0.019	-0.024 - 0.056	0.363
		Duration	
FSH	0.018	-0.044 - 0.078	0.569
AFC	0.047	-0.022 - 0.112	0.177
AMH	-0.041	-0.113 - 0.026	0.246

ROC regression analysis showing the effect of the patient characteristics on the ROC curve of the ovarian reserve tests in the prediction of an excessive ovarian response. Bold: significant influence of the patient characteristics on the discriminatory capacity of the ovarian reserve test in the prediction of an excessive response. AFC = Antral Follicle Count; AMH = Anti-Müllerian Hormone; FSH = Follicle Stimulating Hormone.

Table 5. Areas under the ROC curve of the ORTs in Excessive Response Prediction in different subgroups

	FSH			AFC			AMH		
	AUC	*95%CI*	*N*	*AUC*	*95%CI*	*N*	*AUC*	*95%CI*	*N*
Age category									
< 30 yrs	**0.62**	0.56 - 0.68	682	**0.75**	0.68 - 0.82	440	0.80	0.73 - 0.87	379
30 – 35 yrs	**0.63**	0.58 - 0.68	1508	**0.71**	0.65 - 0.77	943	0.78	0.70 - 0.85	658
35 – 40 yrs	**0.63**	0.56 - 0.70	1523	**0.72**	0.62 - 0.81	860	0.82	0.72 - 0.93	616
> 40 yrs	**0.54**	0.37 - 0.71	507	**NA**	NA	281*	NA	NA	237*
BMI category									
< 20 (kg/m^2)	0.65	0.52 - 0.79	292	0.81	0.69 - 0.92	299	0.86	0.73 - 0.99	133
20-25 (kg/m^2)	0.64	0.59 - 0.70	1033	0.74	0.68 - 0.81	948	0.79	0.71 - 0.86	630
25-30 (kg/m^2)	0.64	0.53 - 0.74	443	0.72	0.61 - 0.83	350	0.82	0.73 - 0.91	335
> 30 (kg/m^2)	NA	NA	96*	NA	NA	76*	NA	NA	76*
Duration of subfertility category									
<2 yrs	0.69	0.55 - 0.83	169	0.80	0.66 - 0.95	123	0.82	0.71 - 0.94	171
2 - 4 yrs	0.67	0.60 - 0.74	817	0.79	0.70 - 0.88	412	0.84	0.77 - 0.90	588
4 - 6 yrs	0.64	0.50 - 0.77	636	0.84	0.71 - 0.96	297	0.77	0.60 - 0.94	404
> 6 yrs	0.64	0.51 - 0.77	346	0.78	0.65 - 0.92	204	0.90	0.78 - 1.00	162

AUC of the ORTs in several clinical subgroups defined by female age, BMI or duration of subfertility are shown. Bold: significant trend in differences between the AUC as calculated with the ROC regression model as showed in *Table 4*. There is significant trend in the difference of the AUCs of AFC and FSH in subgroups defined by female age. The AUCs of AMH do not differ significantly in these subgroups. * = Too few patients with an excessive response to perform ROC analysis. NA = Not available

Discussion

The results of the present IPD meta-analysis demonstrate that in the prediction of an excessive response both AFC and AMH clearly add value to the use of female age alone. Based on the predictive accuracy measures, AMH and the AFC in concert provide the vast majority of the predictive information, and female age has no additional value. The results also indicate that the performance of the ORTs may slightly depend on the specific patient profile, although the clinical significance of this finding may be marginal.

It was also demonstrated that there is a high degree of variation in excessive response incidences as well as in ovarian reserve test averages between different studies. This implicates that individual study results can not be automatically extrapolated to other IVF populations. Using the IPD approach, this meta-analysis allows for correction of heterogeneity, thereby making the findings generally applicable. The results of this IPD-study are in line with a conventional systematic review of ovarian reserve tests and excessive response (151). However, in the present analysis the added value of ORTs to knowing female age and the accuracy in subgroups has been studied for the first time. We could thereby demonstrate that in the prediction of an excessive response, the role for female age is negligible.

AMH and AFC are thought to represent the size of the cohort of FSH sensitive follicles continuously present in the ovaries. Response to ovarian hyperstimulation was shown to be directly linked to this cohort size (96). The size of this cohort decreases with increasing age. AMH levels also decline gradually with age and become undetectable a few years before the occurrence of menopause. Therefore it is not surprising that these ORTs are accurate single predictors of a poor and excessive response.

Moreover, the results of this IPD meta-analysis suggest that age influences the accuracy of AFC and FSH. Although ovarian reserve decreases with age and is therefore influenced by age, the AFC is believed to reflect the true level of the quantitative ovarian reserve directly, in contrast to basal FSH which constitutes an indirect marker of follicle numbers. Indeed, in older women the prevalence of excessive response may become too low for any test to gain sufficient accuracy, and this may be especially true for FSH. For the AFC, the change in accuracy may be significant only from the statistical point of view, without a true implication of clinical practice, and without an obvious explanatory mechanism.

A challenge with the IPD approach is collecting sufficient data for the database. For the current study databases of 60 of the eligible 125 manuscripts were obtained. We were unable to reach a number of authors, primarily because of inac-

curate contact information or because authors did not reply to e-mail addresses provided. Furthermore, older data were often lost or kept in a format that could no longer be read. Currently, studies are being conducted to investigate the possibility of combining IPD data with aggregated data (158). In order to make an appropriate comparison between studies that were included and the studies that were not included we aimed to calculate a Spearman correlation for the included and non included studies. Unfortunately, of the non-included studies only one reported sensitivity and specificity values for AFC in the prediction of an excessive response. Therefore, Spearman correlation could not be calculated. For the majority of the studies this was performed in the IMPORT study (196;197) which showed that there was no difference. Since there is no difference in the poor response prediction, it is reasonable to assume that there is no difference for excessive response prediction. Therefore, we believe that with the current number of participants and amount of data, we were able to analyze a representative selection of available data.

Using original data of different studies comes with heterogeneity between studies. The incorporation of ovarian reserve tests and restrictions based on test results in everyday IVF practice has led to selection bias in some study populations. Heterogeneity found in the included studies pertained to differences in IVF indications or access to IVF resources, different treatment protocols and embryo laws and discordant definitions of ongoing pregnancy. There is also a variation in hormone assays and AFC sizes measured, for which no international consensus exist to correct for these differences. Consequently, no cut-off values for these tests could be used or mentioned. We have used the model by Janes and Pepe et al. (116;117) in which the heterogeneity between studies is corrected for.

The clinical value of excessive response prediction will depend on the consequences for clinical management. Several studies have looked at the effect of individualized treatment protocols. By providing women with personally tailor-made stimulation protocols, ie with a lower dosage of FSH, it is attempted to keep the oocyte yield between 5-12 oocytes. At present, the evidence is inconclusive upon the result of such personalized treatment regimens based on a priori prediction of ovarian response (101;150). In the study of Popovic-Todorovic the use of individualized protocol had effect for poor responders but not for predicted excessive responders (101). In contract, Olivennes et al., do demonstrate that lower individualized dosage protocols allow for a similar oocyte yield, implantation rate and pregnancy as for higher dosage protocols (150). More evidence, in the form of large scale randomized control trials, need

to be performed to demonstrate whether an individualized treatment protocol is an effective strategy in the prevention of an excessive response.
In conclusion, this IPD meta-analysis shows that AFC and AMH add predictive accuracy to age in the prediction of an excessive response. A model combining these ORTs provides almost optimal prediction. Moreover, the performance in several clinical subgroups seemed not to be sufficient altered to be recognized as clinically relevant. The high predictive accuracy for both AMH and the AFC or a combination of both urges the need for studies that examine the effect of ORT based dose adaptations in which efficacy of treatment, costs and response normalization is analyzed.

Addendum

Table A-I. Characteristics of the included studies

Author	Consecutive	Cohort / Case control	Pro-/Retro-spective	Blinding	Selection bias	Verification bias	One cycle per couple	Data per cycle
Aflatoonian	yes	cohort	prospective	no	yes	no	yes	yes
Anderson	no	cohort	prospective	no	yes	no	yes	yes
Ashrafi	no	cohort	retrospective	no	yes	yes	yes	yes
Bancsi	yes	cohort	prospective	no	no	no	yes	yes
Caroppo	no	cohort	retrospective	no	yes	no	yes	yes
Copperman	no	cohort	retrospective	no	no	no	no	yes
Ebner	yes	cohort	prospective	no	yes	yes	yes	yes
Eldar-Geva	yes	cohort	prospective	no	yes	no	yes	yes
Erdem	yes	cohort	retrospective	no	yes	yes	no	yes
Freour	yes	cohort	prospective	no	yes	no	yes	yes
Gnoth	yes	cohort	prospective	no	yes	no	yes	yes
Greenblatt	yes	cohort	retrospective	no	yes	no	yes	yes
Jayaprakasan	yes	cohort	prospective	no	yes	no	yes	yes
Klinkert	yes	cohort	prospective	no	yes	yes	yes	yes
Kwee	yes	cohort	prospective	no	yes	yes	yes	yes
La Marca	yes	cohort	prospective	no	yes	yes	yes	yes
McIlveen	yes	cohort	prospective	no	yes	no	yes	yes
Mercé	yes	cohort	prospective	no	yes	yes	yes	yes
Muttukrishna 2004	no	cohort	prospective	no	yes	no	yes	yes
Muttukrishna 2005	yes	cohort	retrospective	no	no	no	yes	yes

Table A-I. Continued

Author	Consecutive	Cohort / Case control	Pro-/Retro spective	Blinding	Selection bias	Verification bias	One cycle per couple	Data per cycle
Nardo*	yes	cohort	prospective	no	yes	no	yes	yes
Nardo	yes	cohort	prospective	no	yes	no	yes	yes
Nelson	yes	cohort	prospective	no	yes	no	yes	yes
Ng 2000	yes	cohort	prospective	no	yes	yes	yes	yes
Ng 2005	yes	cohort	prospective	no	yes	yes	yes	yes
Popovic-Todorovic 2003a	yes	cohort	prospective	no	yes	yes	yes	yes
Popovic-Todorovic 2003b	yes	cohort	prospective	no	yes	yes	yes	yes
Smeenk 2000	yes	cohort	retrospective	no	yes	no	yes	yes
Smeenk 2007	no	cohort	prospective	no	no	no	yes	yes
Tomas	yes	cohort	prospective	no	no	yes	yes	yes
van Rooij	yes	cohort	prospective	no	yes	yes	yes	yes
van der Linden	yes	cohort	prospective	no	no	no	yes	yes
Vladimirov	yes	cohort	prospective	no	yes	no	yes	yes

*= unpublished data.

Table A-II. Baseline characteristics of the included studies.

Study	Female age (years)	BMI (kg/m²)	Duration of subfertility (years)	FSH (IU/l)
	Mean (5th–95th percentile)	*Mean (5th–95th percentile)*	*Mean (5th–95th percentile)*	*Mean (5th–95th percentile)*
Aflatoonian	28.3 (21.0-34.0)	25.0 (20.5-30.0)	6.1 (2.0-12.0)	5.3 (3.0-7.1)
Anderson	34.3 (26.6-42.2)	23.9 (17.9-34.9)	5.1 (2.0-10.3)	7.5 (3.7-12.3)
Ashrafi	30.0 (22.6-39.5)	NA	6.4 (1.0-17.4)	6.2 (1.6-15.1)
Bancsi	34.6 (27.0-40.7)	NA	4.8 (1.9-11.1)	8.4 (4.1-15.0)
Caroppo	38.0 (35.0-43.0)	NA	NA	11.4 (4.9-21.2)
Copperman	35.5 (26.9-42.9)	NA	NA	7.4 (3.4-13.6)
Ebner	32.7 (24.0-39.2)	NA	4.1 (1.0-11.7)	8.1 (4.4-13.8)
Eldar-Geva	30.0 (22.3-37.0)	23.8 (17.7-37.3)	4.2 (1.5-10.3)	6.7 (3.7-11.1)
Erdem	35.2 (27.6-44.4)	NA	9.6 (1.3-20.8)	8.1 (3.9-14.7)
Freour	30.2 (24.0-37.5)	22.9 (17.7-31.9)	3.7 (1.8-8.0)	6.1 (3.8-8.5)
Gnoth	36.4 (29.0-43.0)	NA	NA	9.3 (3.4-24.5)
Greenblatt	33.5 (27.0-39.0)	NA	NA	6.6 (4.1-9.6)
Jayaprakasan	33.5 (25.1-39.0)	NA	NA	7.2 (4.0-10.7)
Klinkert	41.1 (38.2-44.7)	NA	NA	9.6 (3.7-20.0)
Kwee	34.0 (27.6-40.0)	NA	3.8 (1.3-7.0)	8.1 (4.2-14.1)
La Marca	35.5 (27.0-42.0)	NA	2.9 (1.0-6.3)	NA
McIlveen	37.3 (29.3-42.8)	NA	4.6 (1.0-13.9)	8.3 (4.7-12.0)
Merce	34.4 (27.3-39.0)	20.6 (17.2-24.4)	2.7 (1.0-6.0)	NA
Muttukrishna 2004	37.6 (28.4-45.0)	NA	NA	7.9 (3.2-16.7)
Muttukrishna 2005	35.4 (28.0-43.0)	NA	NA	6.9 (3.8-12.4)
Nardo*	32.8 (25.0-38.8)	24.3 (19.1-30.1)	4.1 (1.5-10.0)	7.2 (3.5-13.2)
Nardo	32.6 (25.8-38.5)	24.4 (19.0-30.0)	NA	7.8 (4.5-12.7)
Nelson	33.9 (26.0-40.0)	24.5 (19.7-30.1)	3.5 (3.0-4.0)	8.7 (3.9-16.5)
Ng 2000	34.3 (27.0-39.0)	22.2 (18.3-28.4)	4.9 (2.0-10.0)	6.5 (3.8-10.8)
Ng 2005	32.8 (28.0-37.0)	20.7 (17.5-26.3)	4.9 (2.0-10.0)	6.5 (4.0-9.0)
Popovic-Todorovic 2003a	32.3 (26.0-38.9)	22.8 (18.8-29.3)	NA	7.0 (4.5-10.0)
Popovic-Todorovic 2003b	32.6 (26.3-37.0)	23.3 (18.6-31.3)	NA	6.3 (3.8-9.0)
Smeenk 2000	34.5 (28.4-41.4)	23.8 (18.5-30.6)	NA	6.8 (3.4-11.4)
Smeenk 2007	32.9 (26.0-40.0)	NA	3.7 (1.0-8.0)	NA
Tomás	33.3 (26.0-39.0)	23.9 (19.1-30.0)	NA	NA
van Rooij	36.3 (28.4-43.9)	23.7 (18.6-31.2)	2.9 (1.0-6.9)	8.5 (3.7-18.2)
van der Linden	NA	NA	NA	8.5 (4.1-14.8)
Vladimirov	34.3 (26.0-44.0)	21.6 (18.9-26.3)	6.5 (3.0-18.0)	7.3 (2.4-14.1)

*= Unpublished data. FSH = Follicle Stimulating Hormone, AFC = Antral Follicle Count, AMH = Anti-Müllerian Hormone. A = AFC2-10mm, B = AFC2-5mm, C = AFC2-8mm, D = DSL assay, E = Beckman Coulter assay. NA = not available.

Table A-II. Continued

AFC (number)	AMH (ng/ml)	Prevalence Excessive Response	Number of patients
Mean (5th–95th percentile)	*Mean (5th – 95th percentile)*	*%*	*N*
15.7 (8.0-28.8)[A]	4.8 (1.5-11.3)[D]	24.5	143
12.9 (4.8-26.6)[A]	NA	5.2	58
NA	NA	10	50
NA	NA	12.5	505
NA	NA	1.3	76
NA	NA	32.8	701
NA	3.4 (0.6-7.9)[E]	14.1	135
22.6 (5.0-50.4)[A]	3.1 (0.6-8.6)[E]	51.9	54
7.0 (2.8-16.0)[C]	NA	18.8	32
NA	4.1 (1.0-10.9)[E]	17.4	69
NA	2.0 (0.0-7.9)[D]	4.4	316
13.8 (5.0-28.5)[C]	NA	23.6	297
16.3 (6.1-29.0)[A]	NA	9	100
7.7 (2.0-17.0)[B]	NA	14.5	221
10 (2.6-20.0)[A]	3.0 (0.3-8.5)[D]	26.4	110
NA	2.1 (0.4-6.1)[E]	3.4	118
7.4 (2.0-13.0)[A]	1.6 (0.5-3.7)[E]	15.5	84
9.2 (1.0-21.0)[B]	NA	12.3	65
NA	0.9 (0.1-4.4)[E]	6.1	66
9.0 (2.6-16.5)	2.1 (0.1-6.0)[E]	15.7	70
14.6 (7-27.1)[B]	3.0 (0.4-8.5)[D]	7.5	334
12.1 (3.0-26.0)[B]	3.1 (0.3-7.3)[D]	21.9	233
NA	1.8 (0.1-5.0)[D]	16.5	340
11.9 (4.0-20.0)	NA	11.5	131
8.9 (4.0-16.0)	NA	26.8	127
14.0 (5.0-27.0)[B]	NA	15.6	262
16.2 (5.3-29.7)[B]	NA	18.6	145
15.9 (5.0-30.0)[A]	3.0 (0.5-8.9)[E]	17.5	80
NA	NA	NA	1292
10.9 (2.0-23.0)[B]	NA	7.2	166
8.4 (1.0-20.9)[B]	1.1 (0.0-3.9)[E]	5	222
NA	NA	13.2	159
8.9 (3.0-17.0)[A]	2.8 (0.5-8.4)[E]	10.3	39

Table A-III. AUCs of the included studies in the prediction of an excessive response

Study	FSH		AFC		AMH	
	AUC	*N*	*AUC*	*N*	*AUC*	*N*
Aflatoonian	0.60 (0.50-0.69)	143	0.96 (0.93-0.99)	143	0.94 (0.90-0.98)	143
Anderson	0.92 (0.99-1.00)	46	0.61(0.67-0.85)	46	NA	
Ashrafi	0.59 (0.31-0.87)	50	NA		NA	
Bancsi	0.61(0.54-0.68)	505	NA		NA	
Caroppo	0.81(0.72-0.90)	76	NA		NA	
Copperman	0.65 (0.60-0.69)	570	NA		NA	
Ebner	0.61 (0.46-0.75)	127	NA		0.82 (0.74-0.90)	135
Eldar-Geva	0.71(0.57-0.85)	52	0.88 (0.75-1.00)	36	0.75 (0.62-0.88)	54
Erdem	0.77 (0.57-0.97)	24	0.85 (0.70-1.00)	24	NA	
Freour	0.58 (0.41-0.73)	62	NA		0.70 (0.55-0.86)	64
Gnoth	0.64 (0.51-0.78)	122	NA		0.87 (0.79-0.95)	134
Greenblatt	0.67(0.59-0.74)	261	0.69 (0.61-0.77)	223	NA	
Jayaprakasan	0.74(0.57-0.91)	100	0.82 (0.70-0.95)	100	NA	
Klinkert	0.42 (0.30-0.55)	212	0.45 (0.33-0.57)	221	NA	
Kwee	0.79 (0.70-0.88)	109	0.87 (0.82-0.96)	109	0.84 (0.76-0.92)	105
La Marca	NA		NA		0.90 (0.76-1.00)	118
McIlveen	No >15	71	No >15	71	No >15	
Merce	NA		0.62 (0.42-0.83)	65	NA	
Muttukrishna 2004	0.81 (0.59-1.00)	66	NA		0.92 (0.83-1.00)	66
Muttukrishna 2005	0.67 (0.52-0.82)	68	0.84 (0.73-0.94)	68	0.73 (0.56-0.91)	68
Nardo*	0.65 (0.53-0.77)	135	0.71(0.59-0.83)	123	0.74 (0.64-0.83)	135
Nardo	0.68 (0.59-0.77)	145	0.71(0.63-0.80)	145	0.79 (0.72-0.87)	145
Nelson	0.64 (0.58-0.71)	338	NA		0.88 (0.82-0.91)	319
Ng 2000	0.70 (0.56-0.83)	131	0.80 (0.70-0.90)	131	NA	
Ng 2005	0.72 (0.56-0.83)	109	0.77 (0.68-0.85)	127	NA	
Popovic-Todorovic 2003a	0.62 (0.54-0.71)	256	0.71(0.63-0.80)	256	NA	
Popovic-Todorovic 2003b	0.62 (0.50-0.73)	143	0.76 (0.67-0.86)	143	NA	
Smeenk 2000	0.54 (0.40-0.68)	80	0.66 (0.5300.79)	80	0.71 (0.57-0.84)	80
Smeenk 2007	NA		NA		NA	
Tomas	NA		0.82 (0.72-0.91)	160	NA	
Van Rooij	0.68 (0.58-0.79)	215	0.86 (0.79-0.93)	215	0.87 (0.77-0.97)	215
Van der Linden	0.82 (0.72-0.92)	124	NA		NA	
Vladimirov 2	0.67 (0.48-0.87)	39	0.74 (0.52-0.97)	39	0.80 (0.67-0.93)	39

Definition excessive response: > 15 oocytes retrieved. AUC = Area Under the Curve. FSH = Follicle Stimulating Hormone, AFC = Antral Follicle Count, AMH = Anti-Müllerian Hormone. *= Unpublished data.

Table A-IV. Univariable and multivariable models of patient characteristics in the prediction of an excessive response

	Three tests study group			Total study group		
	OR	*95% CI*	*P - value*	*OR*	*95% CI*	*P - value*
Univariable models						
Age (per year)	0.89	0.85 - 0.93	< 0.001	0.90	0.88 - 0.91	< 0.001
BMI (per kg/m^2)	0.98	0.93 - 1.03	0.405	1.00	0.97 - 1.03	0.954
Duration (per year)	0.98	0.90 - 1.06	0.555	0.97	0.92 - 1.01	0.156
Multivariable models						
Age and BMI						
Age (per year)	0.91	0.87 - 0.95	< 0.001	0.9	0.87 - 0.93	< 0.001
BMI (per kg/m^2)	0.99	0.93 - 1.04	0.616	1.00	0.97 - 1.04	0.976
Age and duration						
Age (per year)	0.90	0.85 - 0.94	< 0.001	0.89	0.86 - 0.91	< 0.001
Duration (per kg/m^2)	1.01	0.93 - 1.10	0.750	1.00	0.95 - 1.05	0.956

OR = Odds Ratio, 95%CI = 95% Confidence Interval. Duration = duration of subfertility.

Chapter 7

Anti-Müllerian Hormone predicts Menopause: a long term follow-up study in normo-ovulatory women

S.L.Broer, M.J.C. Eijkemans, G.J. Scheffer, I.A.J. van Rooij, A. de Vet, A.P.N. Themmen, J.S.E. Laven, F.H. de Jong, E.R. te Velde, B.C. Fauser, and F.J.M. Broekmans

JCEM in press

Abstract

Context

It has been hypothesized that a fixed interval exists between age at natural sterility and age at menopause. Both events show considerable individual variability, with a range of 20 years. Correct prediction of age at menopause could open avenues for individualized prevention of age related infertility and other menopause related conditions, like cardiovascular disease and breast carcinoma.

Objective

To explore the ability of ovarian reserve tests (ORTs) to predict age at menopause.

Design and Setting

Long term follow-up study in an academic hospital.

Participants

257 normo-ovulatory women (age 21-46 years), derived from 3 cohorts with highly comparable selection criteria.

Interventions

Anti-Müllerian Hormone (AMH), Antral Follicle Count (AFC) and Follicle Stimulating Hormone (FSH) were assessed at T1. At T2 ~11 years later, cycle status (strictly regular, menopausal transition or postmenopause) and age at menopause were inventoried.

Main Outcome Measures

Accuracy of the ORTs in predicting time to menopause was assessed by Cox-regression and a nomogram was constructed for the relationship between age-specific AMH concentrations at T1 and age at menopause.

Results

A total of 48 (19%) women had reached postmenopause at T2. Age, AMH and AFC at T1 were significantly related with time to menopause ($P<0.001$) and showed a good percentage of correct predictions (C-statistic 0.87, 0.86 and 0.84, respectively). After adjusting for age, only AMH added to this prediction (C-statistic 0.90). From the constructed nomogram it appeared that the normal distribution of age at menopause will shift considerably, depending on the individual age-specific AMH level.

Conclusions

AMH is highly predictive for timing of menopause. Using age and AMH, the age range in which menopause will occur, can be individually calculated.

Introduction

Menopause, defined as the final menstrual period, marks the end of the female reproductive life span. This event occurs at a median age of about 51 years, but age at menopause varies between 40 and 60 years (7). The definitive loss of natural fertility is experienced at a median age of 41 years, with a distribution and age variation range highly similar to that for age at menopause (20;60;199). These reproductive events are dictated by the decline in number of follicles in the ovaries (the ovarian reserve) with increasing age. When follicle numbers fall below a critical threshold of a few thousand, the menstrual cycle pattern becomes irregular (8). At menopause, fewer than 1,000 follicles are left (9-11). For human fertility, optimal conditions are present until on average of 31 years of age, followed by gradually decline until natural sterility (200;201). It has been postulated that these events follow a time sequence with a more or less fixed interval, with the end of natural fertility occurring some 10 years before menopause (7).

As the rate of decline of the ovarian reserve varies considerably between individual women, the development of tests that correctly forecast an individual's reproductive lifespan would represent a major step forward (7;20). It has been shown that the number of antral follicles in the ovaries is proportionally related to the size of the primordial follicle pool from which they were recruited (39). A marker correctly reflecting the number of antral follicles is therefore potentially suitable for the prediction of ovarian senescence. Current candidate markers for such purpose are Anti-Müllerian Hormone (AMH) levels (42), the Antral Follicle Count (AFC) as measured by transvaginal ultrasound (41), and early follicular Follicle Stimulating Hormone (FSH) concentration (40).

In case correct individual prediction of menopause would be feasible, several options emerge for the preventive management of age related female infertility, and other female health conditions influenced by timing of menopause (31). Predicted early menopause could emphasize the need for timely prevention of bone demineralization, and cardiovascular and neurological disease (32-34), while the prediction of late menopause would open options for preventive management of breast and intestinal cancer (35).

In the current long term follow up study we therefore aim to explore the ability of endocrine and ultrasound markers to predict the timing of the occurrence of menopause and age at menopause in a group of normo-ovulatory female volunteers.

Methods

Participants

This study group is comprised of 3 cohorts of participants. The first cohort of women was derived from an ongoing prospective longitudinal study on ovarian function (202). 172 healthy female volunteers were recruited in 1996 and 1997. Women could be included if they were between 25 and 46 years of age and had a regular menstrual cycle with a mean length of 21-35 days and the next menstrual period predictable within a 7 day frame. All women had proven natural fertility, which was defined as having established at least one pregnancy within 1 year after discontinuing contraceptives, resulting in a normal delivery at term. If a woman used hormonal contraceptives, this had to be discontinued at least 3 months before the start of the study. Exclusion criteria were ovarian surgery or ovarian abnormalities.

The second cohort consisted of 90 healthy volunteers that were recruited between 1999 and 2001 for a prospective longitudinal study on pregnancy prediction in the normal population. Inclusion criteria were age between 18-46 years, 2 ovaries, no adnexal surgery in the past and a regular menstrual cycle with a mean length between 21 and 35 days. Couples attempting both first and second pregnancy could participate as long as no previous history of infertility was present. Hormonal contraceptives were discontinued at least 3 months before the measurement of the ovarian reserve tests.

The third cohort of 40 normo-ovulatory women were recruited between 1983-1992 as normal controls for studies in relation to ovarian dysfunction in polycystic ovary syndrome (203-205), and subsequently were asked to participate in a prospective longitudinal study on ovarian function in the year 2000 (71). Inclusion criteria were age 20-35 years, regular menstrual cycle (mean cycle length 26-31 days), body mass index of 19-26 kg/m2, absence of endocrine disorders or any other relevant disease, no hormonal treatment for at least 3 months before the study and no prior treatment for infertility.

All three studies had been approved by the institutional review boards of the University Medical Center Utrecht or the Erasmus Medical Center Rotterdam. Written informed consent was obtained from each participant.

Study design

Volunteers visited the clinic for the first time (T1) during the early follicular phase of the menstrual cycle (on cycle day 2, 3 or 4) for assessment of the number of antral follicles (2-10mm) by transvaginal ultrasonography and to provide blood samples. The ultrasound scans were performed by a limited group of physicians, well trained in transvaginal sonography. The ovary was examined

by scanning from the outer to the inner margin. Round or oval echo-free structures in the ovaries were regarded as follicles and were counted and measured as such. The numbers of follicles in both ovaries were added to compute the antral follicle count. Serum and plasma samples were separated and stored at -20°C until assay of AMH and FSH.
In the period 2008-2010 (T2) all women were contacted again and asked to fill out a standardized questionnaire. Participants were questioned on whether they were still menstruating, on the mean cycle length and the variability of the cycle length. In addition, data on use of hormones, medication, surgical treatment on uterus or ovaries and reproductive history were collected. All completed questionnaires were judged independently by two medical doctors before recording in an electronic database. Both medical doctors were blinded for the results of the ovarian reserve tests. All participants were then placed in one of five subgroups, according to their cycle status or use of sex steroid hormones: regular cycle, menopausal transition, menopause, use of exogenous estrogens or surgical removal of uterus and/or ovaries.

Definitions
Menopause was defined as no menstrual period in the last 12 consecutive months. No uniform definition for the transition to menopause (cycle irregularity) is available, but some definitions based upon increasing variability in cycle patterns have been proposed (206). We defined menopausal transition according to these STRAW criteria, as follows: (A) mean cycle length less then 21 or more than 35 days during the previous half year or longer, or (B) mean cycle length between 21 or 35 days, but the next menstrual period not predictable within a 7 days time frame. A regular cycle was defined as a mean cycle length of 21-35 days and the next menstrual period predictable within a 7 days time frame. Women who were using hormone therapy for medical reasons or as contraceptives or hormonal replacement therapy were excluded. Also, women who underwent surgery leading to removal of the uterus and/or one or both ovaries were excluded from the analysis.

Hormone Assays
Blood sampling was performed on the same day as the transvaginal sonography at T1 (1991/2001). Hormone concentrations were measured in plasma (FSH) and serum (AMH). Specimens were stored at -20°C until processing. Concentrations of FSH were measured with the use of the MEIA technology on a fully automated AxSYM immunoanalyser (Abbott Laboratories) according to the manufacturer's instructions. The World Health Organization Second Inter-

national Reference Preparation for human FSH (78/549) was used as a standard in the FSH assay. For FSH, interassay coefficients of variation were found to be 5.7%, 5.7%, and 7.8% at the levels of 5, 26 and 79 IU/L, respectively (n = 80). The detection limit for the FSH assay was 0.03IU/l.
In the first cohort the AMH levels were measured using an enzyme-immunometric assay (Diagnostic Systems Laboratories, Inc., Webster, TX.). Inter-assay and intra-assay coefficients of variation were less than 5% at the level of 3.0 ng/ml and less than 11% at the level of 13.0 ng/ml. The detection limit of the assay was 0.026 ng/ml. Repeated freezing and thawing of the samples or storage at 37C for 1 h did not affect the results of the assay (73). In the second and third cohort, AMH levels were measured with an ultrasensitive immunoenzymometric assay (Immunotech-Coulter, Marseille, France)(207). The limit of detection (defined as blank + 3SD of blank) was 0.05 ng/ml. Intra and interassay coefficients of variation were less than 5% and less than 8%, respectively.
For the comparison and pooling of the AMH levels a correction coefficient was applied. The AMH levels measured with the Beckman Coulter had to be corrected with a factor of 0.5 to be translated into the AMH levels measured using the DSL assay as we described in an earlier study (36).

Statistical analysis

Based on the average age at follow-up and the expected number to be excluded because of use of hormones or surgical removal of uterus and/or ovaries, a number of 50 women in menopause was anticipated. This number would be sufficient to allow for reliable analysis of five predictive variables for the association with age at menopause, according to the ten events per variable rule of thumb (208).
First, baseline characteristics of the women in the 3 cohorts were compared using the Kruskal-Wallis test or the Chi square tests. Moreover, the baseline characteristics and ovarian reserve tests were compared for women divided into subgroups according to their cycle status at T2.
Then, univariate and multivariate Cox regressions for time to menopause, with follow up time from T1 to menopause or T2 as time axis and the occurrence of menopause as event were performed to assess the predictive capacity of age and ovarian reserve tests. For the multivariate analysis, a forward selection with a P value of less then 0.05 for entry was applied. The effects of the variables were expressed as hazard ratios per 1 standard deviation change in order to allow for a better comparability between the effect sizes of the different tested variables. The C-statistic was calculated to inform on the ability to correctly predict the time to menopause.

For ORTs that significantly added to female age in the prediction of timing of menopause, a prediction model was built. We used a Weibull survival model having age of the women on the time axis, with delayed entry at the age at T1, and percentiles of the ORT as a single covariate. Participants were divided into percentiles for their age specific ORT level by fitting a flexible spline function to the scatter plot of the ORT with age at T1, and assuming a normal distribution of residuals around this fitted curve. Therefore, prior to this analysis, ORT values were log (AMH) or square root (AFC) transformed. For each percentile the curve of the predicted distribution of age at menopause was plotted. Per age category, ORT levels corresponding with the different percentiles will be shown as well as the corresponding median, p5, p25, p75 and p95 of the predicted age at menopause distribution. Data were analyzed with SPSS 15.0 (SPSS Inc., Chicago, Il) and R version 2.9.0. (http://www.r-project.org/).

Results

The 3 cohorts together comprised 302 women. The questionnaire could not be sent to 21 women because they had passed away during the follow up period, moved abroad or correct contact address information could not be obtained. The questionnaire was thus sent to 281 women. Of these women, 24 were either not willing to participate or did not respond to the questionnaire, in spite of repeated mailing and efforts to make telephone contact. In total, 257 women could be included with a follow up rate of 91.5%. The baseline characteristics of these women are shown in *Table 1*. It becomes clear that the women in the three cohorts differ in age distribution, while other possible confounders for age at menopause, such as smoking where not different. Strong confounders such as ovarian abnormalities or surgery have been controlled for by the selection criteria.

Table 1. Baseline Characteristics

	Total	Cohort 1	Cohort 2	Cohort 3	P-value
	(n=257)	*(n=153)*	*(n=71)*	*(n=33)*	
Age at T1 (years)	35.5 ± 5.9	38.0 ± 5.4	32.8 ± 4.5	30.1 ± 4.0	<0.001
Age menarche (years)	12.7 ± 2.0	12.8 ± 2.2	12.7 ± 1.4	12.1 ± 2.5	0.211
BMI (kg/m^2)	24.0 ± 4.0	24.1 ± 4.1	24.3 ± 4.0	22.3 ± 2.8	0.052
Smoking (number (%))	43 (16.7 %)	29 (19.0%)	7 (9.9 %)	7 (21.2%)	0.180

Means and Standard Deviation or numbers (percentages) are shown. P-values calculated between the different cohorts, using Kruskal-Wallis or Chi square test.
Note: All three cohorts consist of healthy volunteers, with a regular cycle, age 18-46 years, no history of ovarian abnormalities or surgery and no hormonal treatment for at least 3 months before entrance into the study. Therefore the most important confounders affecting age at menopause, such as ovarian surgery, have been controlled for. The comparability between the cohorts is further reflected in this table. Only age differs significantly between the cohorts, which is corrected for in the Cox regression by delayed entry for age.

From the questionnaires it appeared that 57 women (22%) were using hormonal therapy and that 15 women (6%) had undergone surgical removal of the uterus or at least one ovary. These 72 women were excluded from the analysis. The remaining 185 women were subdivided in groups of women who still had a strictly regular cycle (n=95 (37%)), women in the menopausal transition (n=42 (16%)) and women in the postmenopause (n=48 (19%)). The proportion of postmenopausal women at T2 was 19%, with a mean age at T2 for the study population of 46.5 years. This is in line with estimates from existing studies on age at menopause distributions, where the proportion of women that had reached menopause at the age of 46 years was 16% (19). The mean interval between T1 and T2 was 11.2 years and did not differ significantly across the different subgroups ($p = 0.051$). The women that were excluded from the analysis were comparable to the women that were included in the analysis, except that the women using hormones were somewhat younger. 12 participants with missing data were discarded from the analysis. Missing data occurred in AFC, AMH, and FSH in 8, 8, and 5 participants, respectively.

Patient characteristics and ORTs at T1 were compared between the subgroups based on the cycle status at T2. A significant difference in age upon initial screening was found between the women with a regular cycle, in menopausal transition or in postmenopause at T2 ($p < 0.001$). AMH levels and the AFCs were significantly lower with increasing loss of cyclicity, while basal FSH concentrations were higher (p value for all ORTs < 0.001). There were no differences in body mass index and percentage of smokers.

The results of the Cox regression of the predictive power of age and the ovarian reserve tests for time until the occurrence of menopause are depicted in *Table 2*. In the univariate analysis it is clear that age, AMH, AFC and basal FSH are all significantly correlated with the time to menopause. Moreover, age, AMH and AFC demonstrated an adequate predictive capacity (C-statistic for proportion of correct predictions of 0.87, 0.86, 0.84, respectively). FSH only showed a moderate predictive capacity (C-statistic of 0.70). From the two most significantly predicting test (AMH and AFC) AMH revealed the strongest hazard ratio per unit of standard deviation, indicating the strength of this predictor.

The analysis of independent effects of the ORTs next to age at T1 in the prediction of time until menopause, revealed that only AMH significantly added predictive information ($p < 0.001$), with an improvement of the c-statistic to 0.90. FSH also showed a significant association, however, the c-statistic did not improve in comparison to the prediction based on age alone, demonstrating the lack of added value for this test. As smoking could be a potential confounder, these analyses were also performed taking smoking into account. Smoking was

Table 2. Predictive capacity for ovarian reserve tests for time to menopause.

	Hazard Ratio	95% CI	P-value	C-statistic
Univariate analysis				
Age (per 5.85 years)	7.989	4.649 - 13.730	<0.001	0.87
AMH (per 0.89 ng/ml)	0.033	0.010 - 0.110	<0.001	0.86
AFC (per 6.94)	0.126	0.068 - 0.230	<0.001	0.84
FSH (per 4.47 IU/L)	1.725	1.464 - 2.030	<0.001	0.70
Multivariate analysis (ajdusted for age)				
AMH (per 0.89 ng/ml)	0.092	0.025 - 0.340	<0.001*	0.90
AFC (per 6.94)	0.559	0.260 - 1.200	0.135	0.88
FSH (per 4.47 IU/L)	1.350	1.112 - 1.640	0.002	0.88

The hazard ratio, as estimated with the Cox Regression, is the effect of the variable on the risk of menopause occurring at a certain time point in the observation period (mean 11 years). The C statistic is the proportion correctly predicted events. The * indicates significant difference in comparison to age alone. Effects are depicted per unit of Standard Deviation of age and ovarian reserve test.

not significantly associated with time to menopause (p = 0.075) in a model with AMH and age, and did not change the predictive capacity of AMH.

As only AMH showed a significant added value to age, a nomogram of age and AMH for the prediction of age at menopause was constructed using a Weibull model. There was a good agreement between the Weibull model and the non-parametric Kaplan-Meier curve for age at menopause. Women were divided into percentiles of AMH level for their age category. Women with a relative low AMH level for their age are in the lower percentiles, women with a relative high AMH level for their age are in the higher percentiles. The distribution of age at menopause was then plotted for each percentile (*Figure 1*). *Figure 1* shows that an age specific AMH level will shift the normal expected distribution of age at menopause to a considerable extent. This becomes more obvious when the data are presented in a forecast table where combined information from age and AMH was linked to predicted age at menopause (*Table 3*). Per age category, AMH levels associated with a certain AMH percentile. For each percentile the predicted p5, p25, P50 (median), p75 and p95 of age at menopause are presented.

For example, a 30 year old woman with an AMH concentration close to 0.15 ng/ml is associated with 5th percentile, therefore her predicted median age at menopause will be 48.8 years (p5 to p95 is 42.1-53.0 years). On the other hand, a 30 year old woman with an AMH concentration close to 4.38 ng/ml is associated with the 95th percentile, therefore her predicted median age at menopause will be 55.3 years (p5 to p95 is 47.7-60.1 years) (*Table 3*).

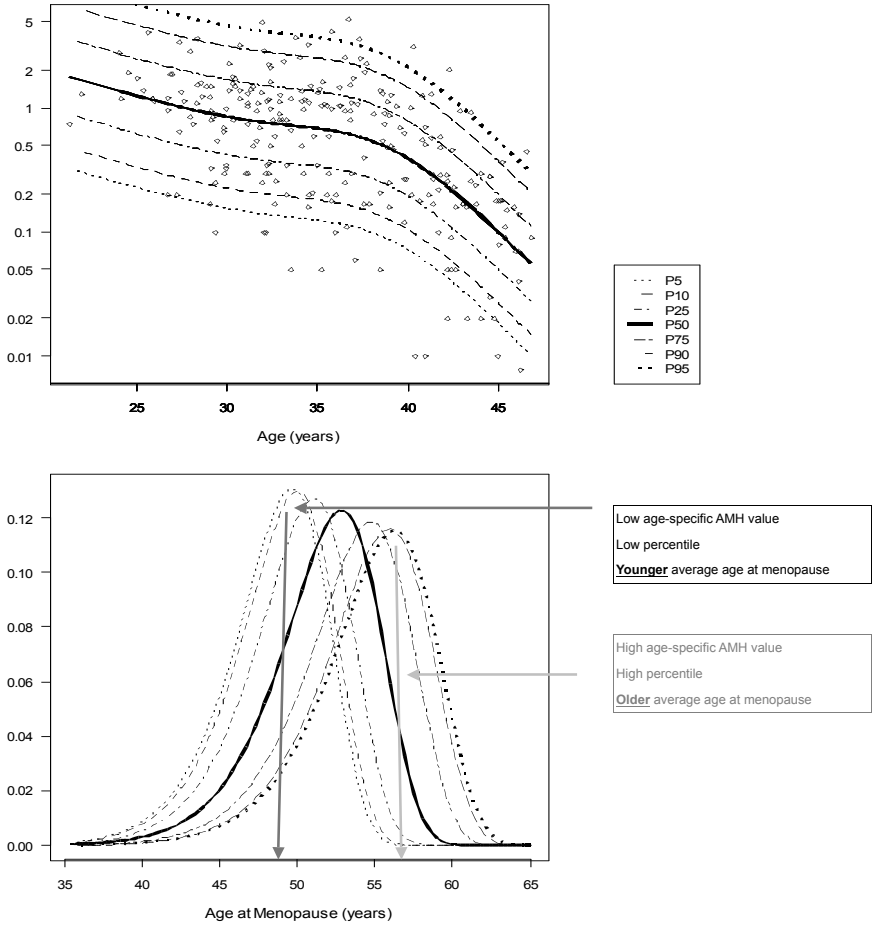


Figure 1. Nomogram for the relation between age specific AMH concentrations and the distribution of Age at Menopause.
In the upper panel the AMH levels measured at entry of the study for women at the given age are shown, measured ~11 years prior to cycle status assessment. The lines represent the upper margins of the different percentiles of AMH. Women thus can be placed in a percentile category based on their AMH concentration at a given age. The lower panel depicts the variation of age at menopause for different percentiles of AMH. Women with a low AMH level for their age will enter a low percentile, for example the P5 which is represented in the small dotted line. For women in a low percentile the predicted distribution of age at menopause shifts towards a younger age. Women with a high AMH level for their age will enter a high percentile, for example the P95 which is represented in the big dotted line. For women in a high percentile, the predicted distribution of age at menopause shifts towards an older age. Note that the median age at menopause in this population is 52 years; this is due to the selection based on cycle regularity at the entry of the study for women up to 46 years old, which will shift the overall age at menopause to a later time.

Table 3. Age specific AMH and corresponding percentiles for AMH and Predicted Age at Menopause

AMH level (ng/ml)													Percentile AMH	Predicted Age at menopause				
														p5	*p25*	*median*	*p75*	*p95*
9.78	8.26	6.85	5.77	4.96	4.38	4.02	3.73	3.28	2.56	1.70	1.00	0.54	95	47.7	52.6	55.3	57.5	60.1
6.72	5.68	4.71	3.96	3.41	3.02	2.76	2.56	2.26	1.76	1.17	0.69	0.37	90	47.4	52.3	54.9	57.1	59.7
5.22	4.41	3.66	3.08	2.65	2.34	2.15	1.99	1.75	1.37	0.91	0.53	0.29	85	47.1	51.9	54.5	56.7	59.2
4.27	3.61	2.99	2.52	2.17	1.92	1.76	1.63	1.43	1.12	0.74	0.44	0.24	80	46.7	51.5	54.2	56.3	58.8
3.60	3.04	2.52	2.12	1.82	1.61	1.48	1.37	1.21	0.94	0.63	0.37	0.20	75	46.4	51.2	53.8	55.9	58.4
3.08	2.60	2.16	1.82	1.56	1.38	1.27	1.18	1.03	0.81	0.54	0.31	0.17	70	46.1	50.8	53.4	55.5	58,0
2.67	2.26	1.87	1.57	1.35	1.20	1.10	1.02	0.90	0.70	0.46	0.27	0.15	65	45.8	50.5	53,0	55.1	57.6
2.33	1.97	1.63	1.37	1.18	1.04	0.96	0.89	0.78	0.61	0.41	0.24	0.13	60	45.4	50.1	52.7	54.8	57.2
2.04	1.73	1.43	1.20	1.04	0.92	0.84	0.78	0.69	0.54	0.36	0.21	0.11	55	45.1	49.8	52.3	54.4	56.8
1.79	1.52	1.26	1.06	0.91	0.80	0.74	0.68	0.60	0.47	0.31	0.18	0.10	50	44.8	49.4	51.9	54.0	56.4
1.58	1.33	1.10	0.93	0.80	0.71	0.65	0.60	0.53	0.41	0.27	0.16	0.09	45	44.5	49.1	51.6	53.6	56.0
1.38	1.17	0.97	0.81	0.70	0.62	0.57	0.53	0.46	0.36	0.24	0.14	0.08	40	44.2	48.7	51.2	53.2	55.6
1.21	1.02	0.85	0.71	0.61	0.54	0.50	0.46	0.41	0.32	0.21	0.12	0.07	35	43.9	48.4	50.8	52.9	55.2
1.05	0.88	0.73	0.62	0.53	0.47	0.43	0.40	0.35	0.27	0.18	0.11	0.06	30	43.6	48,0	50.5	52.5	54.8
0.90	0.76	0.63	0.53	0.45	0.40	0.37	0.34	0.30	0.23	0.16	0.09	0.05	25	43.3	47.7	50.1	52.1	54.5
0.75	0.64	0.53	0.44	0.38	0.34	0.31	0.29	0.25	0.20	0.13	0.08	0.04	20	43.0	47.4	49.8	51.8	54.1
0.62	0.52	0.43	0.36	0.31	0.28	0.25	0.24	0.21	0.16	0.11	0.06	0.03	15	42.7	47.0	49.4	51.4	53.7
0.48	0.40	0.34	0.28	0.24	0.21	0.20	0.18	0.16	0.13	0.08	0.05	0.03	10	42.4	46.7	49.1	51.1	53.3
0.33	0.28	0.23	0.19	0.17	0.15	0.14	0.13	0.11	0.09	0.06	0.03	0.02	5	42.1	46.4	48.8	50.7	53.0
20-22	**22-24**	**24-26**	**26-28**	**28-30**	**30-32**	**32-34**	**34-36**	**36-38**	**38-40**	**40-42**	**42-44**	**44-46**						
					Age (years)													

Results of prediction of menopause are represented in a tabular form, in which women can combine their current age and AMH level to find their AMH percentile with it's corresponding range of predicted age at menopause. AMH levels are represented in ng/ml as measured by the DSL assay. Given a certain age frame, the AMH level closest to the concentration measured in a particular women can be found; to the right find the corresponding percentile of age corrected AMH level, which further to the right will give a prediction of the p5, p25, median (P50), p75 and p95 of age at menopause for women in that percentile.

Discussion

This prospective study is the first to report on long-term follow-up of ovarian reserve status in normo-ovulatory women. It demonstrates that AMH is capable of predicting future age at menopause for a given woman. This finding opens new avenues for the primary prevention of infertility and menopause related conditions.

The rationale for the predictive value of AMH for menopause timing is based on the age related decline in follicle number. Serum AMH levels have been shown to strongly correlate with the number of antral follicles (72;73), and are well capable of predicting ovarian response to hyperstimulation for IVF (114). From earlier follow-up studies it has become evident that serum AMH represents the best endocrine marker to assess the age-related decline of follicle number (137). This has been based on the solid cross sectional relation with age and the consistent decline with time both for group data as well as within individuals.

The present findings build upon earlier studies on the relationships between ovarian reserve markers and menopause. In a cross-sectional study (36) the relation between age specific AMH levels and menopause could be demonstrated from a comparison of the distribution of menopausal age based on true observations, with that of the distribution predicted from an AMH decline model. In a short term follow up design in normal women, the possible role of AMH as a predictor for the occurrence of the menopausal transition, independent of age was demonstrated for the first time (37). Subsequently the usefulness of AMH levels as predictors of menopause in women followed in their late reproductive period (38). Moreover, a linear decline of AMH to undetectable levels over a 9 year period in perimenopausal women to some 4-5 years before the occurrence of menopause has been demonstrated (138). These very low AMH levels could reflect the exhaustion of the ovarian follicle pool, resulting in the loss of steady cyclic ovarian function in the menopausal transition in individual women. In the present study, the wide range of ages at initiation of the observation period emphasizes the possibility of long term prediction at those stages of life, where relevant decisions on preventive management are still feasible.

The results of the present study show the unique abilities of AMH compared to other known tests for ovarian reserve. As for the AFC, a cross sectional study has suggested a possible relation with the timing of menopause (14). In the present study, as well as in the earlier report by van Rooij et al., however, both the AFC as well as basal FSH have failed to show a significant predictive capacity in predicting age at menopause compared to the prediction on basis of age alone (37). Age at menopause is linked to loss of natural fertility, occurring at a mean

age of 41 years, with a distribution curve very similar to the one for menopausal age (7;14;20). If individualized predictions of the menopausal age range could be given early in life, a tool for individualized preventive management of age related infertility could be developed. Such advanced knowledge could lead to important strategy decisions, such as individual planning to attempt conception earlier or preservation of fertility by banking oocytes (209).

Age at menopause is also related to women's health in general. It has been shown that bone loss accelerates following menopause. The earlier menopause occurs, the lower bone density will be later in life (31;32). Furthermore, data have also consistently shown an increased risk for cardiovascular disease for women experiencing premature menopause (33). Also, an increased risk for cognitive impairment or dementia was shown for women experiencing premature menopause (34). At the other end of the spectrum, late age at menopause increases the risks for development of breast and endometrial cancer (35). From this knowledge, preventive management regarding cardiovascular, reproductive and neurological health could be targeted, based on menopause prediction at an early stage of life.

In the current study, women between the age of 21 and 46 with still regular cycles at T1 were included. This is likely to explain both the somewhat higher predicted median age at menopause of 52 years in this cohort, as well as the strong predictive effect of age at T1 in the prediction of menopause. If women with any cycle status would have been included in the cohort, the relation between age at T1 and time to menopause would have been weaker. For AMH, the relationship with menopause timing would have become reinforced by also selecting women with both lower AMH and earlier menopausal ages.

A possible limitation of this study is that it is composed of 3 separate, though quite similar, cohorts. Each of the cohorts had been set up for studies on the status and decline of normal fertility in the human. Moreover, the protocols for the selection of these volunteers were highly comparable. All three cohorts reflect healthy volunteers, with a regular cycle, age between 18-46 years, no history of ovarian surgery, no previous or current ovarian abnormalities and no hormonal treatment for at least 3 months before entrance into the study. The difference in age distribution between the cohorts, will not affect the comparability of the cohorts. Moreover, this distribution difference is taken into account in the analysis by delayed entry for age in the Cox regression. Furthermore, subgroups analyses have been performed, showing no differences in the predictive performance of the ORTs within the three subgroups (AMH $p = 0.58$, AFC $p = 0.65$, FSH $p = 0.95$, test for interaction). As such, we feel that it is justified to combine

these 3 cohorts into a single study population. Another limitation for both research and clinical practice is the usage of two different AMH assays. However, subgroup analysis comparing the performance of AMH measured with the different assays revealed no significant difference ($p = 0.28$, test for interaction). In previous studies the results of Beckman Coulter and DSL assays showed a good correlation, although the translation of results from one assay to the other is not based on large bodies of published data (90;173). In our own laboratory setting a consistent correction factor of 2 has been used also in previous studies (36;210). Interpretation of the data in the forecast *Table 3* may only be translated freely into clinical practice, after thorough assay standardization, but rather more offers insight into the possible future use of this marker.

In summary, AMH is highly predictive for the time interval until the occurrence of menopause. Using age and AMH, the age range in which menopause will occur, can be individually predicted. Correct prediction of age at menopause could open avenues of individualized prevention of age related infertility and menopause related conditions, like cardiovascular disease and breast carcinoma. Long term follow up studies starting at age 20 need to show whether predictions over periods longer than 11 years will be possible, using not only endocrine and ultrasound factors, but also genetic information to be linked to future reproductive events.

Acknowledgements

Hereby we thank all participants for volunteering and putting time and effort into this longitudinal study. Thanks to you, research on natural fertility in healthy women is possible.

Chapter 8

General discussion

Ovarian ageing represents the age related decline of the quantity and quality of the oocytes residing within the follicles present in the ovarian cortex. The changes in quantity and quality will lead to four milestones in the reproductive lifespan: the onset of decreasing fertility, the subsequent loss of natural fertility, menopausal transition and finally menopause. The normal process of ovarian ageing varies considerably among women. This implicates that woman of the same age differ in their reproductive capacity. Despite the variation with age, it is hypothesized that a fixed temporal relationship is present among the four reproductive events, with the occurrence of the end of natural fertility some 10 years before menopause (7). Because of the individual variability, the development of tests that correctly forecast the ovarian reserve status is desirable.
Such test are early follicular Follicle Stimulating Hormone (FSH) concentration (40), the Antral Follicle Count (AFC) as measured by transvaginal ultrasound (41), and Anti-Müllerian Hormone (AMH) levels (42).
In this thesis we have therefore evaluated the real accuracy of these ovarian reserve tests (ORTs) in the prediction of the current and future ovarian reserve status. The aims of this thesis can be listed as follows:

1. Study the accuracy of ovarian reserve tests in the prediction of outcome in ART
2. Study the added accuracy of ovarian reserve tests in the prediction of outcome in ART, when baseline patient characteristics such as female age are taken into account
3. Assess the added value of ovarian reserve tests in the prediction of age at menopause

Current ovarian reserve status

The first two aims reflect the capacity of the ORTs in the prediction of the current ovarian reserve status. The studies presented in chapters 2-6 focus on these aims. It was demonstrated that the ORTs AFC and AMH have an accurate predictive value for the outcome ovarian response, both for poor and excessive response. These tests are superior over female age and a single test of AMH or in combination with the AFC shows a comparable accuracy to any other multivariable model. Although they can predict ovarian response, we have demonstrated that these tests clearly fail in the prediction of an ongoing pregnancy. Female age is still the best predictor of pregnancy and no ORT or combination of ORTs can improve this accuracy. The use of ORTs for the prediction of a possible pregnancy has now been proven to be redundant. Since ovarian response to controlled hyperstimulation is considered as an expression of the

quantitative aspect of the ovarian reserve, while the realization of an ongoing pregnancy is considered as an expression of the qualitative aspect of the ovarian reserve, we can conclude that the ORTs solely reflect the quantitative aspect of the current ovarian reserve status.

The reason why ORTs do not predict pregnancy could be that pregnancy prospects depend on more factors than the oocyte alone. The chance of pregnancy after IVF/ICSI is also dependent on the embryo quality, transfer technique and endometrial receptivity (140). Furthermore, it can be said that only in consecutive cycles a true estimation of the pregnancy potential of a couple can be studied. One study did analyze the capacity of the ORTs in multiple cycles, but demonstrated that in a multivariable model age is the only independent predictor of pregnancy in 3 subsequent cycles (106). This indicates that age currently is the only suitable marker for the qualitative ovarian reserve. Since age is indicative of both the quantitative and qualitative aspect of the ovarian reserve, it is reasonable to assume that the quantitative and qualitative aspects of the ovarian reserve are at least linked in some way. There are several arguments for an indirect or direct relationship between these aspects. For example, women with a poor response in IVF have a lower spontaneous and ART related pregnancy rate (23). It has also been demonstrated that a low number of retrieved oocytes in an IVF/ICSI treatment is related to miscarriage and trisomic pregnancy, although this effect becomes more explicit at higher female age (211;212). These results suggest that there is a link between quantity and quality. However, it has also been demonstrated that young poor responders have reasonable to good pregnancy rates (213;214), even if an elevated FSH level is present (215;216). This indicates that the poor prospects in poor responders may be much more age dictated, and that early decline in follicle quantity does not necessarily affect oocyte quality to the same extent.

In the limited oocyte hypothesis, the quality of oocytes is assumed to be directly related to the quantity (217;218). At the same time there is some support for the theory that an age-dependent accumulation of damage to the oocyte exists, this would be on the basis of for example cumulating oxidative stress (219). A recent study shows alterations in abundance and activity of MCAK, a gene involved in meiosis, in aged oocytes. This may contribute to the loss of control of cell cycle and chromosomal behaviors, thus increasing risk for errors in chromosome segregation and aneuploidy. However, further studies need to be performed to unravel the process of qualitative decline of the oocytes with age and the relation to the quantitative decline of the ovarian reserve. The hypothetical relationship between quantitative and qualitative ovarian ageing is shown in *Figure 1*.

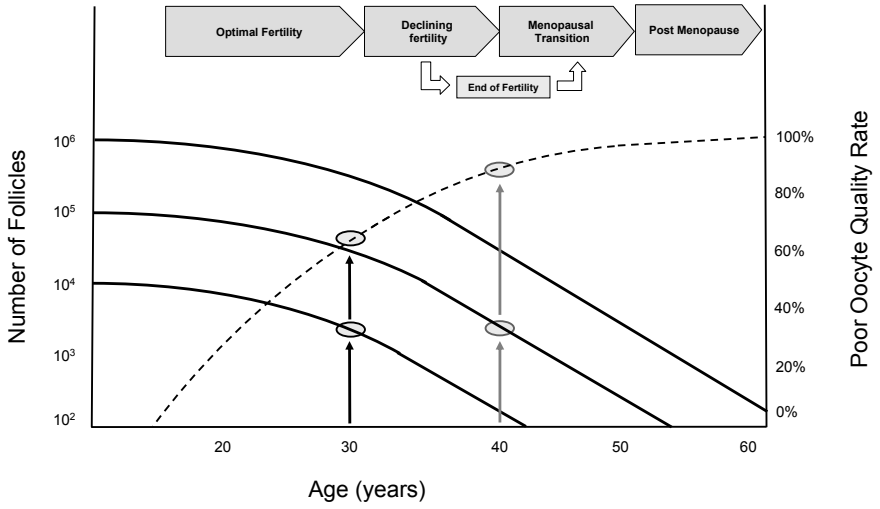


Figure 1. The hypothetical relationship between quantitative and qualitative ovarian ageing
The black lines indicate variability in the quantitative decline based on a different endowment at birth. The black dotted line indicates the increase of the proportion of poor quality oocytes and therefore the decline of quality of the oocytes, based on the theory that there is an age-dependent accumulation of damage to oocytes. This will explain that age determines quality and that young women with low quantity have better pregnancy prospects than older women with a low quantity, due to a lower decline in quality at a younger age.
For example, a 30 year old woman with a diminished quantitative ovarian reserve has a lower percentage poor quality oocytes (black arrows) as a 40 year old woman with the same decrease in the quantitative ovarian reserve (grey arrows).
Note that the rate of the decrease of the quantitative and qualitative ovarian reserve may also vary.

Regardless of the true nature of the link between the quantitative and qualitative aspect of the ovarian reserve, we have demonstrated that through measurement of the quantity aspect no real information about the quality aspect (pregnancy) can be obtained. Consequently, we can contemplate whether there is a real clinical value in measuring solely the quantitative ovarian reserve prior to IVF/ICSI treatment.

Although the prediction of ovarian response categories is accurate, the clinical value of this finding is depends on the consequences that the outcome of these tests have on patient management. The question of which management options could be chosen based on the test result, and the cost-effectiveness of using the test with its subsequent management changes needs to be evaluated. Clinical implications of abnormal test results could vary from counseling the patient on the expected response to ovarian hyperstimulation, to changing patient management by for example FSH dosage adjustments. Since no individualized information regarding pregnancy prospects can be given, the value of counseling is probably limited and most likely not cost-effective.

The most interesting consequence of ovarian response prediction is its potential for individualizing the stimulation protocol with focus on FSH dosage adaptation or the use of mild stimulation approaches in an antagonist system. To date, studies have provided contradictory results (101;102;148-150). In a randomized study doubling the starting dose of gonadotrophins from 150 to 300 IU per day in predicted poor responders, based on an AFC<5, did not lead to an improvement of the response during IVF treatment nor pregnancy prospects (102). In a comparable, but pseudo-randomized design, it was demonstrated that increasing the starting dose of FSH stimulation in potential poor responders based on low AMH values did not alter response or pregnancy rates (148). Moreover, the effect of two high dose FSH treatment arms (300 versus 400 IU daily) in predicted poor responders based on basal FSH levels was studied. Despite an apparent good response in both dosage arms, the outcome at all stages of the IVF treatment was still equally poor and clearly poorer than in women with normal FSH levels (149). In contrast to these three studies, another study revealed that using an individual starting dose based on a response predicting algorithm, did in fact narrow the distribution of ovarian responses and reduce the incidence of patients with a poor or excessive response (101). These results were confirmed by a study demonstrating that with an individual dose fewer cancellation were needed because of an excessive response (150). A large, well designed randomized controlled trial is necessary to study the true value of individualization of the FSH dosage for both the prevention of a poor as an excessive response, while simultaneously observing the cost-effectiveness of this strategy.
Currently, in the Netherlands such a trial is starting, the OPTMIST trial (OPTIMisation of cost effectiveness through Individualized FSH Stimulation dosages of IVF Treatment: a randomized trial, registration nr: NTR2657). The results of this trial will determine whether ORTs prior to IVF/ICSI treatment have clinical value or if they are redundant.
A new possibility for predicted poor responders could be the supplementation of androgens. Several preliminary studies have been performed on this subject suggesting an increase in the ovarian reserve, by an increase in AMH levels (220), higher number of oocytes retrieved and day 3 embryo's after IVF treatment (221), reduced embryo aneuploidy rates (222), and lower miscarriage rates (223) after supplementation of DHEA. This studygroup speculates that DHEA may positively affect recruitment from the dormant primordial follicular pool, it may reduce apoptosis or it may selectively and directly affect granulosa cells and thus increase AMH output. This would suggest progressively more pre-antral follicles are accumulating and/or improvement of follicle quality overall, resulting in the AMH increase. They do not speculate on how DHEA

would affect non-dysfunctional events. Several randomized trials have been performed, but with conflicting results. One study shows no significant beneficial effects of androgen administration on the ovarian response (224). But on the contrary another study did show a significant difference in oocyte yield, but not in pregnancy prospects (225). Finally, a significant difference in live birth after treatment with androgens was demonstrated (226), although these result and statistical analysis are questioned (227). A Cochrane review regarding LH supplementation, describes that there was no evidence for statistical differences in pregnancy outcomes. Noteworthy is that all these trials contain limited numbers of patients and the effect independent of age has not been analyzed. Therefore, to this point it is not clear whether the favorable outcomes that have been reported, were already to be expected, since studies have described acceptable (pregnancy) prospect for especially young poor responders.
Finally, a possible valuable application for ORTs is to use them in case of a poor response in the first IVF attempt, to confirm whether there is a truly a diminished ovarian reserve or other putative causes like underdosing based on FSH polymorphisms or gross obesity (63;228-230). This would imply that routine testing prior to starting ART would be abandoned and the first cycle would be seen as a trial of ovarian capacity. Women with a poor response and a confirmation of diminished ovarian reserve with an ORT could be counseled to refrain from further treatment (61;231). Such application needs to be studied in larger cohorts and the OPTMIST trial (OPTIMisation of cost effectiveness through Individualized FSH Stimulation dosages of IVF Treatment: a randomized trial, registration nr: NTR2657) will offer new data to judge the applicability.

Future ovarian reserve status

If a woman's future ovarian reserve status or reproductive lifespan can be predicted early in life it may have a significant influence on the choices women will make regarding their career and family planning. Mass campaigns propagating earlier child birth have not been successful, but perhaps individual advice on attempting conception earlier in life might be more successful. Especially, since new options are arising for women with a predicted early menopause/end of natural fertility. If these women do not want to attempt conception early in life, cryopreservation of oocytes could be a fair alternative (209;232). Feasibility and desirability studies need to be performed to demonstrate whether women incorporate their predicted reproductive lifespan in their family planning or fertility management like oocyte banking. If these results will not follow, then the knowledge of ovarian ageing will aid in infertility care and women's health in general, since menopause is also related to other systems in the human body.

Early menopause is associated with osteoporosis, cardiovascular disease, cognition, well-being and sexual health, while late menopause on the other hand is associated with breast, ovarian and endometrial cancer.

To study the value of ORTs in the assessment of a woman's future ovarian reserve status, an extended follow up study was performed of which the results are described in Chapter 7. We have demonstrated in the first long term follow-up study that an age specific AMH value can be used to calculate an individual age range in which menopause will subsequently occur. In this follow-up study a time interval of 11 years was studied. Further studies need to confirm if AMH is still a good predictor of age at menopause with a time interval of around 20-30 years. Since the real value of predicting the future fertility status is for woman of an age around 20 years, so that they can be properly informed about their reproductive lifespan. Moreover, we used a nomogram for age-specific AMH values, but it needs to be demonstrated that AMH is constantly declining and that women remain in the same AMH percentile during their reproductive lifespan. Only if this is demonstrated, one measurement of AMH at a young age will give information about a women's entire reproductive lifespan.
Some studies have been performed on the age related decline of AMH. A recent study, showed an age-related nomogram, in which AMH declines in a non-linear way, which is best described by a quadratic equation for women aged 25-45 years (233). This decline is consistent with a maintained relationship among the non-growing follicular cohort size, age and reproductive ageing, with an increased heterogeneity of circulating AMH concentrations in younger women. The heterogeneity in young women may implicate that AMH does not decline that much in women aged 20-30 years. However, an age related decline that was best fitted by a polynomial function was also demonstrated together with normal values for different age categories (234). Another study does show an AMH decline in women aged 25-45 years and a normal distribution of AMH values in assumed healthy women, but does not describe the pattern of decline (235). In a healthy cohort of girls and women in the Netherlands, it was demonstrated that in the pre-pubertal years AMH serum levels rise, despite the enormous reduction of the primordial follicle pool in this period. After an initial rise, AMH levels seem to remain constant until about 25 years of age, after which AMH levels start to decline (data in press, Lie Fong et al.). A large follow up study of women from around 20 years, of which AMH levels are measured regularly and information of age at menopause is present, is needed to confirm whether AMH levels even at a young age can be predictive of the reproductive lifespan. The MORGEN cohort in the Netherlands could be such a cohort. In this cohort a

randomly selected group of more than 10,000 women is followed every 5 years to collect information about their menstrual pattern, to investigate fertility and to draw blood samples.

With new insights on the mechanisms dictating ovarian ageing and its individual variation, identification of other possible markers for the individualized prediction of the reproductive lifespan could be an option.

Genetic factors have proven to play a major role in determining the variation in menopausal age, as demonstrated in several mother-daughter, twin and sib-pair studies. Estimation of heritability in menopausal age range from 31-87% (236-239). Next to genetic factors, several environmental and life-style factors like smoking, body mass index, use of alcohol and parity have claimed to influence menopausal timing as well (19;240-242). Thus, menopausal age is a complex trait. A recent systematic review summarizes all the studies that have been performed on genes that possibly encode menopausal age (243). From this review it became apparent that a number of genetic regions and variants involved in several possible pathways underlying timing of age at menopause could be identified. Two possible interesting regions (9q21.3 on chromosome 8 at 26cM) in linkage analyses were found and genome wide association studies have identified two genomic regions (19q13.42 and 20p12.3), containing two promising candidate genes (BRKS1 and MCM). Another recent study demonstrated that genes involved in initial follicle recruitment may also be associated with age at menopause; variation in the AMHR2 gene modifies the relationship between parity and age at natural menopause and BMP15 was associated with menopausal age. An easy accessible expression of the genetic influence of ovarian ageing could be maternal age at menopause. A recent study showed that maternal age at menopause predicts the AFC and its decline, and thereby likely the size of the primordial follicle pool (244). Moreover, the association of urinary FSH of women and their mother's age at menopause was already demonstrated (245). Identification of the genes and environmental factors that contribute to the endowment and wastage of follicles in the ovaries and thus timing of menopause will add to the understanding of the physiological mechanism of this trait. It could lead to possible new markers that represent the expectation of the reproductive lifespan.

Vascular factors have also been mentioned as possible markers of ovarian ageing, since the association between early menopause and vascular disease as a possible causative factor has recently received attention. Infertile women with reduced ovarian reserve as expressed by a poor response to ovarian hyperstimulation during ART, have demonstrated to reach menopause earlier and have an increased rate of vascular complications in subsequent pregnancies

(246-248). For women with premature menopause who become pregnant after oocyte donation, the same pattern of vascular compromise has become obvious (249). It has also been suggested that pregnancies after IVF are more likely to occur in those women with a favorable vascular status (250;251). Moreover, associations have been found between age at natural menopause and a variant of the APO-E gene, which is associated with longevity and atherosclerosis (252;253). Finally, cardiovascular risk factors such as hypertension, obesity and hypercholesterolemia have shown to be associated with a decreased age at natural menopause (254). If vascular factors are shown to be a major causative factor in timing of ovarian ageing, a breakthrough in long term prediction may come from early markers for vascular quality, as well as from genetic factors driving these vascular changes.

In summary, in the process of ovarian ageing genetic, environmental and vascular factors seem to play a role. The actual process and interaction of all these components in the mechanisms of reproductive ageing needs to be elucidated. Ovarian reserve tests can be seen as an expression of this individual constitution.

In conclusion, this thesis has demonstrated that ORTs are capable of predicting the current quantitative ovarian reserve status. The clinical value of these predictions will depend on future studies that determine whether individualizing the stimulation protocol has benefits for the patient and if this is a cost-effective strategy. ORTs do not predict pregnancy after IVF/ICSI treatment, and therefore the current qualitative ovarian reserve can not be determined with the ORTs. Female age remains the most important predictor of the qualitative aspect of the current ovarian reserve. Since female age is also related to the quantitative aspect of the ovarian reserve it is believed that the quantitative and qualitative aspects of ovarian ageing are somehow related. The mechanisms of mainly qualitative ovarian ageing and the relationship with quantitative ovarian ageing still need to be unraveled further.

Moreover, we have demonstrated that AMH is highly predictive of timing of menopause and that an age-specific AMH value can calculate the individual age range in which menopause will occur. Further long term follow up studies, preferably from the age of 20 years and up need to demonstrate whether we can predict the future fertility or reproductive lifespan for young women. The influence of genetic and vascular factors on the reproductive lifespan need to be taken into account in such a study. Prediction of the reproductive lifespan will lead to the possibility of exploring the primary prevention of age related infertility and menopause related conditions.

Chapter 9

Summary

This thesis aims to evaluate the true value of ovarian reserve tests in the assessment of the current and future ovarian reserve status. We focused on evaluating the available evidence through conventional meta-analysis. Using individual patient data meta-analysis the continuous nature of these test and the possibilities to study these tests in multivariable models and subgroups was further evaluated. For the assessment of the future ovarian reserve status an extended follow study was performed to demonstrate the value of these tests in the prediction of age at menopause.

The introduction, **Chapter 1**, addresses the concept of ovarian ageing, the purpose of ovarian reserve testing and the several ovarian reserve tests. Ovarian ageing is the gradual decline in the number of oocytes and the simultaneous decrease of the quality of the remaining oocytes. The changes in quantity and quality will lead to four milestones in the reproductive lifespan: subfertility, the end of natural fertility, menopausal transition and menopause. It is hypothesized that a fixed interval exists between these milestones, with the end of natural fertility occurring 10 years before menopause. The normal process of ovarian ageing varies considerably among women, with an age range of about 20 years. Therefore, accurate prediction of the current and future ovarian reserve status is desirable. This thesis evaluates the true value of ovarian reserve tests in the assessment of the current and future ovarian reserve status.

In **Chapter 2**, the value of the ovarian reserve tests Antral Follicle Count (AFC) and Anti-Müllerian Hormone (AMH) in the prediction of a poor response or ongoing pregnancy were assessed in a conventional systematic meta-analysis. A total of 13 studies were found reporting on AMH and 17 on the AFC. Summary receiver operating characteristics (ROC) curves were estimated and compared. The curves for the prediction of a poor response showed that both AMH and AFC had a good accuracy and indicated no significant difference between the performances of AMH and AFC. For the prediction of pregnancy a poor performance was found for both AMH and AFC.

Chapter 3 is an individual patient data meta-analysis (IMPORT study) regarding the ORTs in the prediction of a poor response or pregnancy. Original data was collected from previously published studies. In total 24 authors contributed 28 databases, regarding 5,705 patients undergoing IVF/ICSI treatment. It was demonstrated that AFC and AMH do have added value to female age in the prediction of a poor response. Yet the predictive value of any multivariable model was not significantly better than that of either AMH or AFC alone. A single test of AMH or AFC could therefore be considered sufficient for the prediction of a

poor response. For the prediction of an ongoing pregnancy ORTs do not add to the limited predictive capacity of female age.

In **Chapter 4**, the performance of the ORTs in different clinical subgroups is assessed in an individual patient data meta-analysis (IMPORT study). It was studied whether age, BMI or duration of subfertility influence the accuracy of the ORTs. For the prediction of poor response the results suggest that AMH is less accurate in older women and that the AFC is less accurate when the duration of subfertility is longer. But an obvious improvement or decline of poor response prediction in one of the subgroups could not be found. It shows that an accurate prediction of an ongoing pregnancy is not feasible, and no clinical subgroup could be identified in which the performance of one of ovarian reserve tests excels. For poor response prediction, ORTs remain applicable in unselected populations indicated for IVF.

Chapter 5 summarizes the results of the existing literature on the ovarian reserve tests AFC and AMH in the prediction of an excessive response in a systematic conventional meta-analysis. In total 9 studies were found reporting on AMH and 5 reporting on AFC. Comparison of the summary estimates of sensitivity, specificity and the summary ROC curves showed that both tests are accurate predictors of an excessive response to ovarian hyperstimulation and there is no significant difference between these tests. Moreover, it was shown that both tests have clinical value for the prediction of an excessive response.

Chapter 6 is an extended version of the IMPORT study and contains individual patient data for meta-analysis on excessive response prediction, the EXPORT study. The search was updated and another 5 databases could be added to the IMPORT study, therefore the EXPORT study contains a total of 33 databases regarding 6,852 patients undergoing IVF/ICSI treatment. It was demonstrated that AFC and AMH add value to female age in the prediction of an excessive response. Interestingly, a single test of AMH or AFC yielded a comparable accuracy. A model combining these two tests provides the vast majority of the predictive information, and female age has no additional value. The influence of patient characteristics on the prediction of an excessive response was also studied and it was shown that age influences the accuracy of AFC and FSH. The accuracy of AFC was higher with increasing age, the accuracy of FSH is lower with increasing age. However, an obvious improvement or decline of the performance of the ORTs could not be found. This implies that ORTs remain applicable for excessive response prediction in unselected populations indicated for IVF.

In **Chapter 7**, the value of the ovarian reserve tests in the prediction of the future ovarian reserve status was assessed. An extended follow up study was performed in 257 normo-ovulatory women. The ORTs FSH, AFC and AMH were assessed at Time 1. At Time 2, around 11 years later, the cycle status of these women (strictly regular, menopausal transition or postmenopause) and age at menopause was assessed. A total of 48 women had reached postmenopause at Time 2. Age, AMH and AFC were significantly related with time to menopause and showed a good percentage of correct predictions. After correction for age, only AMH added to the prediction. Therefore, a nomogram of age specific AMH values was constructed. From the constructed nomogram it appeared that the normal distribution of age at menopause will shift considerably, depending on the age specific AMH value. Therefore, the age range in which menopause will subsequently occur, can be individually calculated.

Chapter 8 discusses the conclusions that can be drawn from this thesis.
In this thesis it was demonstrated that the ORTs AFC and AMH are capable of predicting the current quantitative ovarian reserve status. The clinical value of this prediction will depend on future studies that demonstrate if adjustment of clinical management can be justified based on these predictions. ORTs do not predict pregnancy after IVF/ICSI treatment, and therefore the current qualitative ovarian reserve can not be determined with ORTs. Since age is indicative of both the quantitative and qualitative aspect of the ovarian reserve, it is reasonable to assume that the quantitative and qualitative aspects of the ovarian reserve are at least linked in some way. The mechanisms of mainly qualitative ovarian ageing and the relationship with quantitative ovarian ageing, however, still need to be unraveled further.
For the prediction of the future ovarian reserve status we have demonstrated that an age-specific AMH value can give an individualized prediction of the age range in which menopause will occur. A long term follow up study, from the age of 20 years and up, needs to demonstrate if AMH can give a prediction of the reproductive lifespan at such a young age. Moreover, the influence of genetic and vascular factors on the reproductive lifespan needs to be taken into account in such a study. Prediction of the reproductive lifespan will lead to the possibility of exploring the primary prevention of age related infertility and menopause related conditions.

Chapter 10

Nederlandse samenvatting

References

Dankwoord

Curriculum Vitae

Het doel van dit proefschrift is het evalueren van de waarde van ovariële reserve testen in de beoordeling van de huidige en toekomstige ovariële reserve. We hebben ons in eerste instantie geconcentreerd op de evaluatie van beschikbare studies door conventionele meta-analyses. Met behulp van individuele patiënten data meta-analyses kon de continue aard van deze testen en de mogelijkheid om deze testen te onderzoeken in multivariabele modellen en subgroepen verder bestudeerd worden. Voor de beoordeling van de toekomstige ovariële reserve status werd een verlengde follow-up studie verricht om aan te tonen of deze testen de menopauze leeftijd kunnen voorspellen.

In de introductie, **Hoofdstuk 1**, wordt het concept ovariële veroudering, het doel van ovariële reserve testen en de verschillende ovariële reserve testen besproken. Ovariële veroudering is de geleidelijke afname van het aantal eicellen en de gelijktijdige afname van de kwaliteit van de overgebleven eicellen. Deze veranderingen in kwantiteit en kwaliteit leiden uiteindelijk tot de volgende vier mijlpalen: verminderde vruchtbaarheid, onvruchtbaarheid, de menopauzale transitie en de menopauze. Er wordt verondersteld dat er een vast tijdsinterval bestaat tussen deze mijlpalen, waarbij onvruchtbaarheid ongeveer 10 jaar voor de menopauze optreedt. Het normale proces van ovariële veroudering verschilt aanzienlijk tussen vrouwen, met een leeftijdsvariatie van ongeveer 20 jaar. Daardoor is een accurate voorspelling van de huidige en toekomstige ovariële reserve status gewenst. Dit proefschrift onderzoekt dan ook de werkelijke waarde van de ovariële reserve testen (ORTs) in de voorspelling van de huidige en toekomstige ovariële reserve status.

In **Hoofdstuk 2** wordt de waarde van de ovariële reserve testen AFC en AMH in de predictie van een slechte respons op ovariële stimulatie en doorgaande zwangerschap beoordeeld in een conventionele meta-analyse. Er werden in totaal 13 studies gevonden die rapporteerden over AMH en 17 over AFC. De samenvattende Receiver Operating Characteristic (ROC) curven werden geschat en vergeleken. De curven voor de predictie van een slechte respons toonden voor zowel AMH als AFC een goede accuratesse en toonden geen significant verschil tussen de prestaties van AMH en AFC. Zowel AMH als AFC lieten een slechte voorspelling voor een doorgaande zwangerschap zien.

Hoofdstuk 3 is een individuele patiënten data meta-analyse (de IMPORT studie) over ovariële reserve testen in de predictie van een slechte response op ovariële stimulatie en/of doorgaande zwangerschap. De originele data van tevoren ge-

publiceerde studies werd verzameld. Vierentwintig auteurs stelden in totaal 28 studie-databases ter beschikking, samen rapporteerden deze studies over 5705 patiënten die een IVF/ICSI behandeling ondergingen. Er kon worden aangetoond dat AFC en AMH toegevoegde waarde hebben op leeftijd in de voorspelling van een slechte respons op ovariële stimulatie. Maar de voorspellende waarde van enig multivariabel model was niet significant beter dan de voorspellende waarde van AMH en AFC. Een enkele AMH of AFC test kan daarom als voldoende worden beschouwd voor de voorspelling van een slechte respons. In de predictie van een doorgaande zwangerschap voegen ORTs geen waarde toe aan de beperkte voorspellende waarde van leeftijd van de vrouw.

In **Hoofdstuk 4** wordt in een individuele patiënten data meta-analyse (de IMPORT studie) de waarde van de ORTs in verschillende klinische subgroepen onderzocht. Er werd bestudeerd of leeftijd, BMI en de duur van de onvruchtbaarheid invloed hebben op de accuratesse van de ORTs. De resultaten voor de predictie van een slechte respons op ovariële stimulatie suggereren dat de accuratesse van AMH daalt bij hogere leeftijd, en dat de accuratesse van AFC daalt bij een langere duur van de onvruchtbaarheid. Maar een duidelijke verbetering of verslechtering van de voorspelling in één van de subgroepen kon niet worden aangetoond. Er werd aangetoond dat de voorspelling van een doorgaande zwangerschap niet mogelijk is, ook niet in een bepaalde subgroep van patiënten. Voor de voorspelling van een slechte respons blijven de ORTs bruikbaar in niet geselecteerde populaties met een indicatie voor IVF.

Hoofdstuk 5 is een systematische conventionele meta-analyse die de beschikbare studies over de ovariële reserve testen AFC en AMH in de voorspelling van een te hoge (hyper) respons samenvat. In totaal rapporteerden 9 studies over AMH en 5 over de AFC. Vergelijking van de samengevatte schatting van de sensitiviteit, specificiteit en de samengevatte ROC curven, toonde dat beide ORTs accurate voorspellers zijn van een hyper respons. Er werd geen significant verschil tussen beide ORTs aangetoond. Daarnaast werd aangetoond dat beide ORTs klinische waarde hebben als voorspellers van een hyper respons.

Hoofdstuk 6 is een uitbreiding van de IMPORT studie en bevat individuele patiënten data voor meta-analyse voor de predictie van een hyper respons, de EXPORT studie. De zoektocht naar geschikte artikelen werd bijgewerkt en 5 studie databases konden worden toegevoegd aan de IMPORT studie. De EXPORT studie bevat dan ook in totaal 33 databases, met een totaal van 6852

patiënten die een IVF/ICSI behandeling ondergingen. In de voorspelling van een hyper response werd aangetoond dat AFC en AMH toegevoegde waarde hebben op de voorspellende waarde van leeftijd. Hoewel slechts een enkele AMH of AFC bepaling vrijwel een gelijke accuratesse heeft. Een model dat beide testen combineert, geeft het merendeel van de voorspellende waarde weer, en leeftijd heeft geen toegevoegde waarde op dit model. De invloed van patiëntkarakteristieken op de accuratesse van de testen werd ook bestudeerd. Leeftijd beïnvloedt de accuratesse van AFC en FSH in de voorspelling van een hyper response; de accuratesse van AFC is hoger bij een hogere leeftijd en de accuratesse van FSH is lager bij een hogere leeftijd. Een duidelijke verbetering of verslechtering van de voorspellende waarde in de verschillende subgroepen kon echter niet worden aangetoond. ORTs blijven daarom bruikbaar in niet geselecteerde IVF populaties voor de voorspelling van een hyper respons.

In **Hoofdstuk 7** wordt de waarde van de ovariële reserve testen in de voorspelling van de toekomstige ovariële reserve status onderzocht. Een verlengde follow-up studie werd verricht bij 257 normo-ovulatoire vrouwen. De ovariële reserve testen FSH, AFC en AMH werden op tijdstip 1 bepaald. Op tijdstip 2, gemiddeld 11 jaar later, werd de status van de menstruele cyclus (strikt regulair, menopauzale transitie en post-menopauze) en de menopauzeleeftijd beoordeeld. Een totaal van 48 vrouwen was post-menopauzaal op tijdstip 2. Leeftijd, AMH en AFC waren significant gerelateerd aan de tijd tot het optreden van de menopauze en hadden een goed percentage correct voorspelden. Na correctie voor leeftijd, bleek dat alleen AMH toegevoegde waarde had in de voorspelling van de menopauze status. Daarom werd een nomogram geconstrueerd van leeftijdspecifieke AMH waarden. Uit het nomogram bleek dat afhankelijk van de leeftijdspecifieke AMH waarde de normale verdeling van de menopauze leeftijd aanzienlijk verschoven kon worden. Hierdoor biedt AMH mogelijkheden om de verwachting ten aanzien van de menopauze leeftijd individueel bij te stellen.

Hoofdstuk 8 bespreekt de conclusies die uit dit proefschrift getrokken kunnen worden. In dit proefschrift wordt getoond dat de ovariële reserve testen AFC en AMH in staat zijn om de huidige kwantitatieve ovariële reserve status te voorspellen. De klinische waarde van deze voorspellingen is afhankelijk van toekomstige studies, die moeten aantonen of het klinisch beleid zou kunnen worden aangepast op basis van deze voorspellingen. Ovariële reserve testen zijn geen goede voorspellers van een zwangerschap na een IVF/ICSI behan-

deling en dit betekent dat de ORTs niet in staat zijn om de huidige kwalitatieve ovariële reserve status te voorspellen. Aangezien leeftijd indicatief is voor zowel de kwantitatieve en kwalitatieve aspecten van de ovariële reserve, is het aannemelijk dat de kwalitatieve en kwantitatieve aspecten van de ovariële reserve in meer of mindere mate aan elkaar gerelateerd zijn. Het mechanisme van vooral de kwalitatieve ovariële veroudering en de relatie met kwantitatieve ovariële veroudering moet echter nog verder ontrafeld worden.
Voor de predictie van de toekomstige ovariële reserve status hebben we aangetoond dat een leeftijdspecifieke AMH waarde een individuele voorspelling kan geven van de leeftijd spreiding waarop de menopauze zal optreden. Een lange termijn follow-up studie van 20 jarige leeftijd moet aantonen of AMH ook op jonge leeftijd een voorspelling kan geven van het reproductieve leven van een vrouw. Voorspelling van het reproductieve leven van een vrouw zal leiden tot de mogelijkheid om de primaire preventie van leeftijdsgerelateerde onvruchtbaarheid en menopauze gerelateerde ziekten te exploreren.

Chapter 10

Nederlandse samenvatting

References

Dankwoord

Curriculum Vitae

Reference List

1. Block E. Quantitative morphological investigations of the follicular system in women; variations at different ages. Acta Anat (Basel) 1952:14;108-123
2 Block E. A quantitative morphological investigation of the follicular system in newborn female infants. Acta Anat (Basel) 1953:17;201-206
3. Baker TG. A quantitative and cytological study of germ cells in human ovaries. Proc R Soc Lond B Biol Sci 1963:158;417-433
4. Markstrom E, Svensson EC, Shao R et al. Survival factors regulating ovarian apoptosis -- dependence on follicle differentiation. Reproduction 2002:123;23-30
5. Eurostat. Population and Social Conditions Database. http://http://epp eurostat ec europa eu/cache/ITY_OFFPUB/KS-RA-07-021/EN/KS-RA-07-021-EN PDF 2006:
6. Thum MY, Abdalla HI, and Taylor D. Relationship between women's age and basal follicle-stimulating hormone levels with aneuploidy risk in in vitro fertilization treatment. Fertil Steril 2008:90;315-321
7. te Velde E.R. and Pearson P.L. The variability of female reproductive aging. Hum Reprod Update 2002:8;141-154
8. Richardson SJ, Senikas V, and Nelson JF. Follicular depletion during the menopausal transition: evidence for accelerated loss and ultimate exhaustion. J Clin Endocrinol Metab 1987:65;1231-1237
9. Faddy MJ, Gosden RG, Gougeon A et al. Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. Hum Reprod 1992:7;1342-1346
10. Hansen KR, Knowlton NS, Thyer AC et al. A new model of reproductive aging: the decline in ovarian non-growing follicle number from birth to menopause. Hum Reprod 2008:23;699-708
11. Wallace WH and Kelsey TW. Ovarian reserve and reproductive age may be determined from measurement of ovarian volume by transvaginal sonography. Hum Reprod 2004:19;1612-1617
12. Research on the menopause in the 1990s. Report of a WHO Scientific Group. World Health Organ Tech Rep Ser 1996:866;1-107
13. Treloar AE. Menstrual cyclicity and the pre-menopause. Am J Human Biol 1981:3;249-264
14. Broekmans FJ, Faddy MJ, Scheffer G et al. Antral follicle counts are related to age at natural fertility loss and age at menopause. Menopause 2004:11;607-614
15. Magursky V, Mesko M, and Sokolik L. Age at the menopause and onset of the climacteric in women of Martin District, Czechoslovkia. Statistical survey and some biological and social correlations. Int J Fertil 1975:20;17-23
16. Bengtsson C, Lindquist O, and Redvall L. Menstrual status and menopausal age of middle-aged Swedish women. A population study of women in Goteborg 1968--69 and 1974--75. Acta Obstet Gynecol Scand 1981:60;269-275
17. Hagstad A, Johansson S, Wilhelmsson C et al. Gynaecology of middle-aged women--menstrual and reproductive histories. Am J Human Biol 1985:7;99-113

18. Luoto R, Kaprio J, and Uutela A. Age at natural menopause and sociodemographic status in Finland. Am J Epidemiol 1994:139;64-76
19. Van Noord PA, Dubas JS, Dorland M et al. Age at natural menopause in a population-based screening cohort: the role of menarche, fecundity, and lifestyle factors. Fertil Steril 1997:68;95-102
20. Broekmans FJ, Soules MR, and Fauser BC. Ovarian aging: mechanisms and clinical consequences. Endocr Rev 2009:30;465-493
21. Lisabeth L, Harlow S, and Qaqish B. A new statistical approach demonstrated menstrual patterns during the menopausal transition did not vary by age at menopause. J Clin Epidemiol 2004:57;484-496
22. de Boer EJ, den Tonkelaar I, te Velde ER et al. Increased risk of early menopausal transition and natural menopause after poor response at first IVF treatment. Hum Reprod 2003:18;1544-1552
23. Lawson R, El Toukhy T, Kassab A et al. Poor response to ovulation induction is a stronger predictor of early menopause than elevated basal FSH: a life table analysis. Hum Reprod 2003:18;527-533
24. Nikolaou D, Lavery S, Turner C et al. Is there a link between an extremely poor response to ovarian hyperstimulation and early ovarian failure? Hum Reprod 2002:17;1106-1111
25. den Tonkelaar I, te Velde ER, and Looman CW. Menstrual cycle length preceding menopause in relation to age at menopause. Am J Human Biol 1998:29;115-123
26. Leridon H. Demographic effects of the introduction of steroid contraception in developed countries. Hum Reprod Update 2006:12;603-616
27. Leridon H. Can assisted reproduction technology compensate for the natural decline in fertility with age? A model assessment. Hum Reprod 2004:19;1548-1553
28. Habbema JD, Eijkemans MJ, Nargund G et al. The effect of in vitro fertilization on birth rates in western countries. Hum Reprod 2009:24;1414-1419
29. Lutz W. Fertility rates and future population trends: will Europe's birth rate recover or continue to decline? Int J Androl 2006:29;25-33
30. Ziebe S and Devroey P. Assisted reproductive technologies are an integrated part of national strategies addressing demographic and reproductive challenges. Hum Reprod Update 2008:14;583-592
31. De Vos M, Devroey P, and Fauser BC. Primary ovarian insuffiency. Lancet 2010:375;
32. Gallagher JC. Effect of early menopause on bone mineral density and fractures. Menopause 2007:14;567-571
33. Shuster LT, Rhodes DJ, Gostout BS et al. Premature menopause or early menopause: long-term health consequences. Am J Human Biol 2010:65;161-166
34. Rocca WA, Shuster LT, Grossardt BR et al. Long-term effects of bilateral oophorectomy on brain aging: unanswered questions from the Mayo Clinic Cohort Study of Oophorectomy and Aging. Womens Health (Lond Engl) 2009:5;39-48
35. Hartge P. Genetics of reproductive lifespan. Nat Genet 2009:41;637-638
36. van Disseldorp J., Faddy MJ, Themmen AP et al. Relationship of Serum Anti-Mullerian Hormone Concentration to Age of Menopause. J Clin Endocrinol Metab 2008:93;2129-2134

37. van Rooij IA, Tonkelaar I, Broekmans FJ et al. Anti-mullerian hormone is a promising predictor for the occurrence of the menopausal transition. Menopause 2004:11;601-606
38. Tehrani FR, Solaymani-Dodaran M, and Azizi F. A single test of antimullerian hormone in late reproductive-aged women is a good predictor of menopause. Menopause 2009:16;797-802
39. Gougeon-A, Echochard-, and Thalabard-JC. Age-related changes of the population of human ovarian follicles: increase in the disappearance rate of non-growing and early- growing follicles in aging women. Biol-Reprod 1994:50;653-663
40. Scott RT, Toner JP, Muasher SJ et al. Follicle-stimulating hormone levels on cycle day 3 are predictive of in vitro fertilization outcome. Fertil Steril 1989:51;651-654
41. Broekmans FJ, de Ziegler D., Howles CM et al. The antral follicle count: practical recommendations for better standardization. Fertil Steril 2010:94;1044-1051
42. Broekmans FJ, Visser JA, Laven JS et al. Anti-Mullerian hormone and ovarian dysfunction. Trends Endocrinol Metab 2008:19;340-347
43. Klein NA, Houmard BS, Hansen KR et al. Age-related analysis of inhibin A, inhibin B, and activin a relative to the intercycle monotropic follicle-stimulating hormone rise in normal ovulatory women. J Clin Endocrinol Metab 2004:89;2977-2981
44. Pache T.D., Wladimiroff J.W., de Jong F.H. et al. Growth patterns of nondominant ovarian follicles during the normal menstrual cycle. Fertil Steril 1990:54;638-642
45. Hansen KR, Hodnett GM, Knowlton N et al. Correlation of ovarian reserve tests with histologically determined primordial follicle number. Fertil Steril 2011:95;170-175
46. Cate RL, Mattaliano RJ, Hession C et al. Isolation of the bovine and human genes for Mullerian inhibiting substance and expression of the human gene in animal cells. Cell 1986:45;685-698
47. Jost A. Recherches sur la differenciation sexuelle de l'embryon de lapin. Arch Anat Microsc Morphol Exp 1947:36;271-315
48. Josso N, Cate RL, Picard JY et al. Anti-mullerian hormone: the Jost factor. Rec Prog Horm Res 1993:48;1-59
49. Rajpert-De Meyts E, Jorgensen N, Graem N et al. Expression of anti-Mullerian hormone during normal and pathological gonadal development: association with differentiation of Sertoli and granulosa cells. J Clin Endocrinol Metab 1999:84;3836-3844
50. Munsterberg A and Lovell-Badge R. Expression of the mouse anti-mullerian hormone gene suggests a role in both male and female sexual differentiation. Development 1991:113;613-624
51. Lee MM and Donahoe PK. Mullerian inhibiting substance: a gonadal hormone with multiple functions. Endocr Rev 1993:14;152-164
52. Visser JA and Themmen AP. Anti-Mullerian hormone and folliculogenesis. Mol Cell Endocrinol 2005:234;81-86
53. Fanchin R, Schonauer LM, Righini C et al. Serum anti-Mullerian hormone dynamics during controlled ovarian hyperstimulation. Hum Reprod 2003:18;328-332
54. La Marca A, Stabile G, Artenisio AC et al. Serum anti-Mullerian hormone throughout the human menstrual cycle. Hum Reprod 2006:21;3103-3107

55. Tsepelidis S, Devreker F, Demeestere I et al. Stable serum levels of anti-Mullerian hormone during the menstrual cycle: a prospective study in normo-ovulatory women. Hum Reprod 2007:22;1837-1840
56. Hehenkamp WJ, Looman CW, Themmen AP et al. Anti-Mullerian hormone levels in the spontaneous menstrual cycle do not show substantial fluctuation. J Clin Endocrinol Metab 2006:91;4057-4063
57. Jayaprakasan K, Deb S, Batcha M et al. The cohort of antral follicles measuring 2-6 mm reflects the quantitative status of ovarian reserve as assessed by serum levels of anti-Mullerian hormone and response to controlled ovarian stimulation. Fertil Steril 2009:
58. Broekmans FJ, Kwee J, Hendriks DJ et al. A systematic review of tests predicting ovarian reserve and IVF outcome. Hum Reprod Update 2006:12;685-718
59. Spira A. The decline of fecundity with age. Am J Human Biol 1988:Suppl 1;15-22
60. Wood JW. Fecundity and natural fertility in humans. Oxf Rev Reprod Biol 1989:11;61-109
61. Klinkert ER, Broekmans FJ, Looman CW et al. A poor response in the first in vitro fertilization cycle is not necessarily related to a poor prognosis in subsequent cycles. Fertil Steril 2004:81;1247-1253
62. Hendriks DJ, Kwee J, Mol BW et al. Ultrasonography as a tool for the prediction of outcome in IVF patients: a comparative meta-analysis of ovarian volume and antral follicle count. Fertil Steril 2007:87;764-775
63. de Koning CH, Benjamins T, Harms P et al. The distribution of FSH receptor isoforms is related to basal FSH levels in subfertile women with normal menstrual cycles. Hum Reprod 2006:21;443-446
64. Lashen H, Ledger W, Lopez-Bernal A et al. Poor responders to ovulation induction: is proceeding to in-vitro fertilization worthwhile? Hum Reprod 1999:14;964-969
65. van DJ, Eijkemans MJ, Klinkert ER et al. Cumulative live birth rates following IVF in 41- to 43-year-old women presenting with favourable ovarian reserve characteristics. Reprod Biomed Online 2007:14;455-463
66. Hendriks DJ, Mol BW, Bancsi LF et al. Antral follicle count in the prediction of poor ovarian response and pregnancy after in vitro fertilization: a meta-analysis and comparison with basal follicle-stimulating hormone level. Fertil Steril 2005:83;291-301
67. Lee MM, Donahoe PK, Hasegawa T et al. Mullerian inhibiting substance in humans: normal levels from infancy to adulthood. J Clin Endocrinol Metab 1996:81;571-576
68. Weenen C, Laven JS, Von Bergh AR et al. Anti-Mullerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. Mol Hum Reprod 2004:10;77-83
69. La Marca A, Malmusi S, Giulini S et al. Anti-Mullerian hormone plasma levels in spontaneous menstrual cycle and during treatment with FSH to induce ovulation. Hum Reprod 2004:19;2738-2741

70. Hudson PL, Dougas I, Donahoe PK et al. An immunoassay to detect human mullerian inhibiting substance in males and females during normal development. J Clin Endocrinol Metab 1990:70;16-22
71. de Vet A, Laven JSE, de Jong FH et al. Anti-Mullerian Hormone serum levels: A putative marker for ovarian aging. Fertil Steril 2002:77;357-362
72. Gruijters MJ, Visser JA, Durlinger AL et al. Anti-Mullerian hormone and its role in ovarian function. Mol Cell Endocrinol 2003:211;85-90
73. van Rooij IA, Broekmans FJ, te Velde ER et al. Serum anti-Mullerian hormone levels: a novel measure of ovarian reserve. Hum Reprod 2002:17;3065-3071
74. Seifer DB, MacLaughlin DT, Christian BP et al. Early follicular serum mullerian-inhibiting substance levels are associated with ovarian response during assisted reproductive technology cycles. Fertil Steril 2002:77;468-471
75. van Rooij IA, de Jong E, Broekmans FJ et al. High follicle-stimulating hormone levels should not necessarily lead to the exclusion of subfertile patients from treatment. Fertil Steril 2004:81;1478-1485
76. Hazout A, Bouchard P, Seifer DB et al. Serum antimullerian hormone/mullerian-inhibiting substance appears to be a more discriminatory marker of assisted reproductive technology outcome than follicle-stimulating hormone, inhibin B, or estradiol. Fertil Steril 2004:82;1323-1329
77. Littenberg B and Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making 1993:13;313-321
78. Midgette AS, Stukel TA, and Littenberg B. A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. Med Decis Making 1993:13;253-257
79. Moses LE, Shapiro D, and Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993:12;1293-1316
80. Muttukrishna S, McGarrigle H, Wakim R et al. Antral follicle count, anti-mullerian hormone and inhibin B: predictors of ovarian response in assisted reproductive technology? BJOG 2005:112;1384-1390
81. Muttukrishna S, Suharjono H, McGarrigle H et al. Inhibin B and anti-Mullerian hormone: markers of ovarian response in IVF/ICSI patients? BJOG 2004:111;1248-1253
82. Penarrubia J, Fabregues F, Manau D et al. Basal and stimulation day 5 anti-Mullerian hormone serum concentrations as predictors of ovarian response and pregnancy in assisted reproductive technology cycles stimulated with gonadotropin-releasing hormone agonist--gonadotropin treatment. Hum Reprod 2005:20;915-922
83. Tremellen KP, Kolo M, Gilmore A et al. Anti-mullerian hormone as a marker of ovarian reserve. Aust N Z J Obstet Gynaecol 2005:45;20-24

84. Eldar-Geva T, Ben Chetrit A, Spitz IM et al. Dynamic assays of inhibin B, anti-Mullerian hormone and estradiol following FSH stimulation and ovarian ultrasonography as predictors of IVF outcome. Hum Reprod 2005:20;3178-3183
85. McIlveen M, Skull JD, and Ledger WL. Evaluation of the utility of multiple endocrine and ultrasound measures of ovarian reserve in the prediction of cycle cancellation in a high-risk IVF population. Hum Reprod 2007:22;778-785
86. Ficicioglu C, Kutlu T, Baglam E et al. Early follicular antimullerian hormone as an indicator of ovarian reserve. Fertil Steril 2006:85;592-596
87. La Marca A, Giulini S, Tirelli A et al. Anti-Mullerian hormone measurement on any day of the menstrual cycle strongly predicts ovarian response in assisted reproductive technology. Hum Reprod 2007:22;766-771
88. Ebner T, Sommergruber M, Moser M et al. Basal level of anti-Mullerian hormone is associated with oocyte quality in stimulated cycles. Hum Reprod 2006:21;2022-2026
89. Smeenk JM, Sweep FC, Zielhuis GA et al. Antimullerian hormone predicts ovarian responsiveness, but not embryo quality or pregnancy, after in vitro fertilization or intracyoplasmic sperm injection. Fertil Steril 2007:87;223-226
90. Freour T, Mirallie S, Bach-Ngohou K et al. Measurement of serum anti-Mullerian hormone by Beckman Coulter ELISA and DSL ELISA: comparison and relevance in assisted reproduction technology (ART). Clin Chim Acta 2007:375;162-164
91. Kwee J, Schats R, McDonnell J et al. Evaluation of anti-Mullerian hormone as a test for the prediction of ovarian reserve. Fertil Steril 2007:90;737-743
92. Cook CL, Siow Y, Taylor S et al. Serum mullerian-inhibiting substance levels during normal menstrual cycles. Fertil Steril 2000:73;859-861
93. Hansen KR, Morris JL, Thyer AC et al. Reproductive aging and variability in the ovarian antral follicle count: application in the clinical setting. Fertil Steril 2003:80;577-583
94. Scheffer GJ, Broekmans FJ, Bancsi LF et al. Quantitative transvaginal two- and three-dimensional sonography of the ovaries: reproducibility of antral follicle counts. Ultrasound Obstet Gynecol 2002:20;270-275
95. Pache TD, Wladimiroff JW, de Jong FH et al. Growth patterns of nondominant ovarian follicles during the normal menstrual cycle. Fertil Steril 1990:54;638-642
96. Kwee J, Elting MW, Schats R et al. Comparison of endocrine tests with respect to their predictive value on the outcome of ovarian hyperstimulation in IVF treatment: results of a prospective randomized study. Hum Reprod 2003:18;1422-1427
97. Boomsma CM and Macklon NS. What can the clinician do to improve implantation? Reprod Biomed Online 2006:13;845-855
98. Templeton A, Morris JK, and Parslow W. Factors that affect outcome of in-vitro fertilisation treatment. Lancet 1996:348;1402-1406
99. van der Gaast MH, Eijkemans MJ, van der Net JB et al. Optimum number of oocytes for a successful first IVF treatment cycle. Reprod Biomed Online 2006:13;476-480

100. Mol BW, Verhagen TE, Hendriks DJ et al. Value of ovarian reserve testing before IVF: a clinical decision analysis. Hum Reprod 2006:21;1816-1823
101. Popovic-Todorovic B, Loft A, Bredkjaeer HE et al. A prospective randomized clinical trial comparing an individual dose of recombinant FSH based on predictive factors versus a 'standard' dose of 150 IU/day in 'standard' patients undergoing IVF/ICSI treatment. Hum Reprod 2003:18;2275-2282
102. Klinkert ER, Broekmans FJ, Looman CW et al. Expected poor responders on the basis of an antral follicle count do not benefit from a higher starting dose of gonadotrophins in IVF treatment: a randomized controlled trial. Hum Reprod 2005:20;611-615
103. Tarlatzis BC, Zepiridis L, Grimbizis G et al. Clinical management of low ovarian response to stimulation for IVF: a systematic review. Hum Reprod Update 2003:9;61-76
104. de Koning CH, Popp-Snijders C, Martens F et al. Falsely elevated follicle-stimulating hormone levels in women with regular menstrual cycles due to interference in immunoradiometric assay. J Assist Reprod Genet 2000:17;457-459
105. Lambalk CB and de Koning CH. Interpretation of elevated FSH in the regular menstrual cycle. Am J Human Biol 1998:30;215-220
106. Hendriks DJ, te Velde ER, Looman CW et al. The role of poor response in the prediction of the cumulative ongoing pregnancy rate in in vitro fertilisation. In: Dynamic and basal ovarian reserve tests for outcome prediction in IVF: comparisons and meta-analyses 2005:Academic Thesis, Utrecht;162-179
107. Broekmans FJ, Knauff EA, te Velde ER et al. Female reproductive ageing: current knowledge and future trends. Trends Endocrinol Metab 2007:18;58-65
108. Hefler LA, Grimm C, Bentz EK et al. A model for predicting age at menopause in white women. Fertil Steril 2006:85;451-454
109. Muasher SJ, Oehninger S, Simonetti S et al. The value of basal and/or stimulated serum gonadotropin levels in prediction of stimulation response and in vitro fertilization outcome. Fertil Steril 1988:50;298-307
110. Scott RT, Toner JP, Muasher SJ et al. Follicle-stimulating hormone levels on cycle day 3 are predictive of in vitro fertilization outcome. Fertil Steril 1989:51;651-654
111. Verhagen TE, Hendriks DJ, Bancsi LF et al. The accuracy of multivariate models predicting ovarian reserve and pregnancy after in vitro fertilization: a meta-analysis. Hum Reprod Update 2008:14;95-100
112. Leeflang MM, Deeks JJ, Gatsonis C et al. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008:149;889-897
113. Broeze KA, Opmeer BC, Bachmann LM et al. Individual patient data meta-analysis of diagnostic and prognostic studies in obstetrics, gynaecology and reproductive medicine. BMC Med Res Methodol 2009:9;22
114. Broer SL, Mol BW, Hendriks D et al. The role of antimullerian hormone in prediction of outcome after IVF: comparison with the antral follicle count. Fertil Steril 2009:91;705-714

115. Broeze KA, Opmeer BC, van d, V et al. Individual patient data meta-analysis: a promising approach for evidence synthesis in reproductive medicine. Hum Reprod Update 2010:16;561-567
116. Janes H, Longton G, and Pepe MS. Accommodating covariates in receiver operating characteristic analysis. Stata Journal 2009:9;17-39
117. Pepe MS, Longton G, and Janes H. Estimation and comparison of receiver operating characteristic curves. Stata Journal 2009:9;1-16
118. Ashrafi M, Madani T, Tehranian AS et al. Follicle stimulating hormone as a predictor of ovarian response in women undergoing controlled ovarian hyperstimulation for IVF. Int J Gynaecol Obstet 2005:91;53-57
119. Yong PY, Baird DT, Thong KJ et al. Prospective analysis of the relationships between the ovarian follicle cohort and basal FSH concentration, the inhibin response to exogenous FSH and ovarian follicle number at different stages of the normal menstrual cycle and after pituitary down-regulation. Hum Reprod 2003:18;35-44
120. Bancsi LF, Huijs AM, Den Ouden CT et al. Basal follicle-stimulating hormone levels are of limited value in predicting ongoing pregnancy rates after in vitro fertilization. Fertil Steril 2000:73;552-557
121. Caroppo E, Matteo M, Schonauer LM et al. Basal FSH concentration as a predictor of IVF outcome in older women undergoing stimulation with GnRH antagonist. Reprod Biomed Online 2006:13;815-820
122. Luna M, Grunfeld L, Mukherjee T et al. Moderately elevated levels of basal follicle-stimulating hormone in young patients predict low ovarian response, but should not be used to disqualify patients from attempting in vitro fertilization. Fertil Steril 2007:87;782-787
123. Erdem M, Erdem A, Gursoy R et al. Comparison of basal and clomiphene citrate induced FSH and inhibin B, ovarian volume and antral follicle counts as ovarian reserve tests and predictors of poor ovarian response in IVF. J Assist Reprod Genet 2004:21;37-45
124. Liu KE and Greenblatt EM. 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 2008:90;297-301
125. Jayaprakasan K, Hilwah N, Kendall NR et al. Does 3D ultrasound offer any advantage in the pretreatment assessment of ovarian reserve and prediction of outcome after assisted reproduction treatment? Hum Reprod 2007:22;1932-1941
126. Klinkert ER, Broekmans FJ, Looman CW et al. The antral follicle count is a better marker than basal follicle-stimulating hormone for the selection of older patients with acceptable pregnancy prospects after in vitro fertilization. Fertil Steril 2005:83;811-814
127. Merce LT, Barco MJ, Bau S et al. Prediction of ovarian response and IVF/ICSI outcome by three-dimensional ultrasonography and power Doppler angiography. Eur J Obstet Gynecol Reprod Biol 2007:132;93-100
128. Ng EH, Tang OS, and Ho PC. The significance of the number of antral follicles prior to stimulation in predicting ovarian responses in an IVF programme. Hum Reprod 2000:15;1937-1942

129. Ng EH, Chan CC, Tang OS et al. Antral follicle count and FSH concentration after clomiphene citrate challenge test in the prediction of ovarian response during IVF treatment. Hum Reprod 2005:20;1647-1654
130. Nelson SM, Yates RW, and Fleming R. Serum anti-Mullerian hormone and FSH: prediction of live birth and extremes of response in stimulated cycles--implications for individualization of therapy. Hum Reprod 2007:22;2414-2421
131. Popovic-Todorovic B, Loft A, Lindhard A et al. A prospective study of predictive factors of ovarian response in 'standard' IVF/ICSI patients treated with recombinant FSH. A suggestion for a recombinant FSH dosage normogram. Hum Reprod 2003:18;781-787
132. Smeenk JM, Stolwijk AM, Kremer JA et al. External validation of the templeton model for predicting success after IVF. Hum Reprod 2000:15;1065-1068
133. Tomas-C, Nuojua-Huttunen-, and Martikainen-H. Pretreatment transvaginal ultrasound examination predicts ovarian responsiveness to gonadotrophins in in-vitro fertilization. Hum-Reprod 1997:12;220-223
134. van Swieten EC, Leeuw-Harmsen L, Badings EA et al. Obesity and Clomiphene Challenge Test as predictors of outcome of in vitro fertilization and intracytoplasmic sperm injection. Gynecol Obstet Invest 2005:59;220-224
135. Vladimirov IK, Tacheva DM, Kalinov KB et al. Prognostic value of some ovarian reserve tests in poor responders. Arch Gynecol Obstet 2005:272;74-79
136. Vladimirov IK, Tacheva DM, and Kalinov KB. Mean ovarian diameter (MOD) as a predictor of poor ovarian response. J Assist Reprod Genet 2004:21;73-77
137. van Rooij IA, Broekmans FJ, Scheffer GJ et al. Serum antimullerian hormone levels best reflect the reproductive decline with age in normal women with proven fertility: a longitudinal study. Fertil Steril 2005:83;979-987
138. Sowers MR, Eyvazzadeh AD, McConnell D et al. Anti-mullerian hormone and inhibin B in the definition of ovarian aging and the menopause transition. J Clin Endocrinol Metab 2008:93;3478-3483
139. Broer SL, Eijkemans MJC, Scheffer G.J. et al. Anti-Müllerian Hormone predicts Menopause: a long term follow-up study in normo-ovulatory women. Academic Thesis S Broer, University Utrecht 2011:
140. Boomsma CM and Macklon NS. What can the clinician do to improve implantation? Reprod Biomed Online 2006:13;845-855
141. Nelson SM and Lawlor DA. Predicting live birth, preterm delivery, and low birth weight in infants born from in vitro fertilisation: a prospective study of 144,018 treatment cycles. PLoS Med 2011:8;e1000386
142. La MA, Nelson SM, Sighinolfi G et al. Anti-Mullerian hormone-based prediction model for a live birth in assisted reproduction. Reprod Biomed Online 2011:
143. Barad DH, Weghofer A, and Gleicher N. Age-specific levels for basal follicle-stimulating hormone assessment of ovarian function. Obstet Gynecol 2007:109;1404-1410

144. Henne MB, Stegmann BJ, Neithardt AB et al. The combined effect of age and basal follicle-stimulating hormone on the cost of a live birth at assisted reproductive technology. Fertil Steril 2008:89;104-110
145. Scott RT, Jr., Elkind-Hirsch KE, Styne-Gross A et al. The predictive value for in vitro fertility delivery rates is greatly impacted by the method used to select the threshold between normal and elevated basal follicle-stimulating hormone. Fertil Steril 2008:89;868-878
146. Shanbhag S, Aucott L, Bhattacharya S et al. Interventions for 'poor responders' to controlled ovarian hyperstimulation (COH) in in-vitro fertilisation (IVF). Cochrane Database Syst Rev 2007:CD004379
147. Sunkara SK, Tuthill J, Khairy M et al. Pituitary suppression regimens in poor responders undergoing IVF treatment: a systematic review and meta-analysis. Reprod Biomed Online 2007:15;539-546
148. Lekamge DN, Lane M, Gilchrist RB et al. Increased gonadotrophin stimulation does not improve IVF outcomes in patients with predicted poor ovarian reserve. J Assist Reprod Genet 2008:25;515-521
149. Harrison RF, Jacob S, Spillane H et al. A prospective randomized clinical trial of differing starter doses of recombinant follicle-stimulating hormone (follitropin-beta) for first time in vitro fertilization and intracytoplasmic sperm injection treatment cycles. Fertil Steril 2001:75;23-31
150. Olivennes F, Howles CM, Borini A et al. Individualizing FSH dose for assisted reproduction using a novel algorithm: the CONSORT study. Reprod Biomed Online 2009:18;195-204
151. Broer SL, Dolleman M, Opmeer BC et al. AMH and AFC as predictors of excessive response in controlled ovarian hyperstimulation: a meta-analysis. Hum Reprod Update 2010:
152. Dinnes J, Deeks J, Kirby J et al. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. Health Technol Assess 2005:9;1-113, iii
153. Midgette AS, Stukel TA, and Littenberg B. A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. Med Decis Making 1993:13;253-257
154. Glas AS, Lijmer JG, Prins MH et al. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003:56;1129-1135
155. Reitsma JB, Glas AS, Rutjes AW et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005:58;982-990
156. Scott RT, Jr., Elkind-Hirsch KE, Styne-Gross A et al. The predictive value for in vitro fertility delivery rates is greatly impacted by the method used to select the threshold between normal and elevated basal follicle-stimulating hormone. Fertil Steril 2008:89;868-878
157. Lee TH, Liu CH, Huang CC et al. Impact of female age and male infertility on ovarian reserve markers to predict outcome of assisted reproduction technology cycles. Reprod Biol Endocrinol 2009:7;100
158. Riley RD, Dodd SR, Craig JV et al. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Stat Med 2008:27;6111-6136
159. Delvigne A and Rozenberg S. Epidemiology and prevention of ovarian hyperstimulation syndrome (OHSS): a review. Hum Reprod Update 2002:8;559-577

160. Fauser BC, Diedrich K, and Devroey P. Predictors of ovarian response: progress towards individualized treatment in ovulation induction and ovarian stimulation. Hum Reprod Update 2008:14;1-14
161. Verberg MF, Macklon NS, Nargund G et al. Mild ovarian stimulation for IVF. Hum Reprod Update 2009:15;13-29
162. Baart EB, Martini E, Eijkemans MJ et al. Milder ovarian stimulation for in-vitro fertilization reduces aneuploidy in the human preimplantation embryo: a randomized controlled trial. Hum Reprod 2007:22;980-988
163. Heijnen EM, Eijkemans MJ, de KC et al. A mild treatment strategy for in-vitro fertilisation: a randomised non-inferiority trial. Lancet 2007:369;743-749
164. de Carvalho BR, Rosa e Silva AC, Rosa E Silva JC et al. Ovarian reserve evaluation: state of the art. J Assist Reprod Genet 2008:25;311-322
165. Nakhuda GS, Sauer MV, Wang JG et al. Mullerian inhibiting substance is an accurate marker of ovarian response in women of advanced reproductive age undergoing IVF. Reprod Biomed Online 2007:14;450-454
166. Nakhuda GS, Chu MC, Wang JG et al. Elevated serum mullerian-inhibiting substance may be a marker for ovarian hyperstimulation syndrome in normal women undergoing in vitro fertilization. Fertil Steril 2006:85;1541-1543
167. Nardo LG, Gelbaya TA, Wilkinson H et al. Circulating basal anti-Mullerian hormone levels as predictor of ovarian response in women undergoing ovarian stimulation for in vitro fertilization. Fertil Steril 2009:92;1586-1593
168. Riggs RM, Duran EH, Baker MW et al. Assessment of ovarian reserve with anti-Mullerian hormone: a comparison of the predictive value of anti-Mullerian hormone, follicle-stimulating hormone, inhibin B, and age. Am J Obstet Gynecol 2008:199;202-208
169. Harbord RM, Deeks JJ, Egger M et al. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 2007:8;239-251
170. Lekamge DN, Barry M, Kolo M et al. Anti-Mullerian hormone as a predictor of IVF outcome. Reprod Biomed Online 2007:14;602-610
171. Lee TH, Liu CH, Huang CC et al. Serum anti-Mullerian hormone and estradiol levels as predictors of ovarian hyperstimulation syndrome in assisted reproduction technology cycles. Hum Reprod 2008:23;160-167
172. Aflatoonian A, Oskouian H, Ahmadi S et al. Prediction of high ovarian response to controlled ovarian hyperstimulation: anti-Mullerian hormone versus small antral follicle count (2-6 mm). J Assist Reprod Genet 2009:26;319-325
173. Bersinger NA, Wunder D, Birkhauser MH et al. Measurement of anti-mullerian hormone by Beckman Coulter ELISA and DSL ELISA in assisted reproduction: differences between serum and follicular fluid. Clin Chim Acta 2007:384;174-175
174. van Disseldorp J., Lambalk CB, Kwee J et al. Comparison of inter- and intra-cycle variability of anti-Mullerian hormone and antral follicle counts. Hum Reprod 2010:25;221-227

175. Lee TH, Liu CH, Huang CC et al. Serum anti-mullerian hormone and estradiol levels as predictors of ovarian hyperstimulation syndrome in assisted reproduction technology cycles. Hum Reprod 2008:23;160-167
176. La Marca A., Broekmans FJ, Volpe A et al. Anti-Mullerian hormone (AMH): what do we still need to know? Hum Reprod 2009:24;2264-2275
177. Kaya C, Pabuccu R, and Satiroglu H. Serum antimullerian hormone concentrations on day 3 of the in vitro fertilization stimulation cycle are predictive of the fertilization, implantation, and pregnancy in polycystic ovary syndrome patients undergoing assisted reproduction. Fertil Steril 2010:94;2202-2207
178. Howles CM, Saunders H, Alam V et al. Predictive factors and a corresponding treatment algorithm for controlled ovarian stimulation in patients treated with recombinant human follicle stimulating hormone (follitropin alfa) during assisted reproduction technology (ART) procedures. An analysis of 1378 patients. Curr Med Res Opin 2006:22;907-918
179. Aboulghar M. Symposium: Update on prediction and management of OHSS. Prevention of OHSS. Reprod Biomed Online 2009:19;33-42
180. Busso CE, Garcia-Velasco J, Gomez R et al. Symposium: Update on prediction and management of OHSS. Prevention of OHSS--dopamine agonists. Reprod Biomed Online 2009:19;43-51
181. Kol S and Dor J. Symposium: Update on prediction and management of OHSS. Prevention of OHSS: GnRH agonist versus HCG to trigger ovulation. Reprod Biomed Online 2009:19;59-60
182. Kosmas IP, Zikopoulos K, Georgiou I et al. Low-dose HCG may improve pregnancy rates and lower OHSS in antagonist cycles: a meta-analysis. Reprod Biomed Online 2009:19;619-630
183. Aboulghar MA and Mansour RT. Ovarian hyperstimulation syndrome: classifications and critical analysis of preventive measures. Hum Reprod Update 2003:9;275-289
184. Simon C, Cano F, Valbuena D et al. Clinical evidence for a detrimental effect on uterine receptivity of high serum oestradiol concentrations in high and normal responder patients. Hum Reprod 1995:10;2432-2437
185. Pena JE, Chang PL, Chan LK et al. Supraphysiological estradiol levels do not affect oocyte and embryo quality in oocyte donation cycles. Hum Reprod 2002:17;83-87
186. Ertzeid G and Storeng R. The impact of ovarian stimulation on implantation and fetal development in mice. Hum Reprod 2001:16;221-225
187. Kodaman PH and Taylor HS. Hormonal regulation of implantation. Obstet Gynecol Clin North Am 2004:31;745-66, ix
188. Devroey P, Bourgain C, Macklon NS et al. Reproductive biology and IVF: ovarian stimulation and endometrial receptivity. Trends Endocrinol Metab 2004:15;84-90
189. Bourgain C and Devroey P. The endometrium in stimulated cycles for IVF. Hum Reprod Update 2003:9;515-522
190. Baart EB, Martini E, van dB, I et al. Preimplantation genetic screening reveals a high incidence of aneuploidy and mosaicism in embryos from young women undergoing IVF. Hum Reprod 2006:21;223-233

191. Nargund G, Fauser BC, Macklon NS et al. The ISMAAR proposal on terminology for ovarian stimulation for IVF. Hum Reprod 2007:22;2801-2804
192. Heijnen EM, Eijkemans MJ, De Klerk C et al. A mild strategy in IVF results in favourable outcomes with regard to term live birth, cost and patient discomfort. Lancet 2006:369;743-749
193. Out HJ, David I, Ron-El R et al. A randomized, double-blind clinical trial using fixed daily doses of 100 or 200 IU of recombinant FSH in ICSI cycles. Hum Reprod 2001:16;1104-1109
194. Baart EB, van dB, I, Martini E et al. FISH analysis of 15 chromosomes in human day 4 and 5 preimplantation embryos: the added value of extended aneuploidy detection. Prenat Diagn 2007:27;55-63
195. Ho HY, Lee RK, Lin MH et al. Estradiol level on day 9 as a predictor of risk for ovarian hyperresponse during controlled ovarian hyperstimulation. J Assist Reprod Genet 2003:20;222-226
196. Broer SL, van Disseldorp J., Broeze KA et al. Added value of ovarian reserve testing on patient characteristics in the prediction of ovarian response and ongoing pregnancy: an individual patient data approach. Academic Thesis S Broer, University Utrecht 2011:
197. Broer SL, van Disseldorp J., Broeze KA et al. Performance of ovarian reserve tests in clinical subgroups of patients undergoing ART: an Individual Patient Data meta-analysis. Academic Thesis S Broer, University Utrecht 2011:
198. Gnoth C, Schuring AN, Friol K et al. Relevance of anti-Mullerian hormone measurement in a routine IVF program. Hum Reprod 2008:23;1359-1365
199. Eijkemans MJ, Habbema J.D.F., and te Velde E.R. Age at last childbirth and fertility at young age. Fertility In Populations and In Patients; M J Eijkemans; Academic Thesis, Erasmus Medical Center Rotterdam, The Netherlands 2005:23-34
200. Noord-Zaadstra BM, Looman CW, Alsbach H et al. Delaying childbearing: effect of age on fecundity and outcome of pregnancy. BMJ 1991:302;1361-1365
201. Schwartz D and Mayaux MJ. Female fecundity as a function of age: results of artificial insemination in 2193 nulliparous women with azoospermic husbands. Federation CECOS. N Engl J Med 1982:306;404-406
202. Scheffer GJ, Broekmans FJ, Looman CW et al. The number of antral follicles in normal women with proven fertility is the best reflection of reproductive age. Hum Reprod 2003:18;700-706
203. van Santbrink EJ, Hop WC, van Dessel TJ et al. Decremental follicle-stimulating hormone and dominant follicle development during the normal menstrual cycle. Fertil Steril 1995:64;37-43
204. Schipper I, de Jong FH, and Fauser BC. Lack of correlation between maximum early follicular phase serum follicle stimulating hormone concentrations and menstrual cycle characteristics in women under the age of 35 years. Hum Reprod 1998:13;1442-1448
205. Hohmann FP, Laven JS, de Jong FH et al. Low-dose exogenous FSH initiated during the early, mid or late follicular phase can induce multiple dominant follicle development. Hum Reprod 2001:16;846-854

206. Soules MR, Sherman S, Parrott E et al. Executive summary: Stages of Reproductive Aging Workshop (STRAW). Fertil Steril 2001:76;874-878
207. Long WQ, Ranchin V, Pautier P et al. Detection of minimal levels of serum anti-Mullerian hormone during follow-up of patients with ovarian granulosa cell tumor by means of a highly sensitive enzyme-linked immunosorbent assay. J Clin Endocrinol Metab 2000:85;540-544
208. Peduzzi P, Concato J, Feinstein AR et al. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995:48;1503-1510
209. Oktay K, Cil AP, and Zhang J. Who is the best candidate for oocyte cryopreservation research? Fertil Steril 2010:93;13-15
210. La Marca A., Sighinolfi G, Radi D et al. Anti-Mullerian hormone (AMH) as a predictive marker in assisted reproductive technology (ART). Hum Reprod Update 2010:16;113-130
211. Haadsma ML, Groen H, Mooij TM et al. Miscarriage risk for IVF pregnancies in poor responders to ovarian hyperstimulation. Reprod Biomed Online 2010:20;191-200
212. Haadsma ML, Mooij TM, Groen H et al. A reduced size of the ovarian follicle pool is associated with an increased risk of a trisomic pregnancy in IVF-treated women. Hum Reprod 2010:25;552-558
213. Biljan MM, Buckett WM, Dean N et al. The outcome of IVF-embryo transfer treatment in patients who develop three follicles or less. Hum Reprod 2000:15;2140-2144
214. Hanoch J, Lavy Y, Holzer H et al. Young low responders protected from untoward effects of reduced ovarian response [see comments]. Fertil Steril 1998:69;1001-1004
215. van Rooij IA, Bancsi LF, Broekmans FJ et al. Women older than 40 years of age and those with elevated follicle-stimulating hormone levels differ in poor response rate and embryo quality in in vitro fertilization. Fertil Steril 2003:79;482-488
216. Check JH, Nazari P, Check ML et al. Prognosis following in vitro fertilization-embryo transfer (IVF-ET) in patients with elevated day 2 or 3 serum follicle stimulating hormone (FSH) is better in younger vs older patients. Clin Exp Obstet Gynecol 2002:29;42-44
217. Henderson SA and Edwards RG. Chiasma frequency and maternal age in mammals. Nature 1968:218;22-28
218. Warburton D. The effect of maternal age on the frequency of trisomy: change in meiosis or in utero selection? Prog Clin Biol Res 1989:311;165-181
219. Tarin JJ. Aetiology of age-associated aneuploidy: a mechanism based on the 'free radical theory of ageing'. Hum Reprod 1995:10;1563-1565
220. Gleicher N, Weghofer A, and Barad DH. Improvement in diminished ovarian reserve after dehydroepiandrosterone supplementation. Reprod Biomed Online 2010:21;360-365
221. Barad D and Gleicher N. Effect of dehydroepiandrosterone on oocyte and embryo yields, embryo grade and cell number in IVF. Hum Reprod 2006:21;2845-2849

222. Gleicher N, Weghofer A, and Barad DH. Dehydroepiandrosterone (DHEA) reduces embryo aneuploidy: direct evidence from preimplantation genetic screening (PGS). Reprod Biol Endocrinol 2010:8;140
223. Gleicher N, Ryan E, Weghofer A et al. Miscarriage rates after dehydroepiandrosterone (DHEA) supplementation in women with diminished ovarian reserve: a case control study. Reprod Biol Endocrinol 2009:7;108
224. Massin N, Cedrin-Durnerin I, Coussieu C et al. Effects of transdermal testosterone application on the ovarian response to FSH in poor responders undergoing assisted reproduction technique--a prospective, randomized, double-blind study. Hum Reprod 2006:21;1204-1211
225. Fabregues F, Penarrubia J, Creus M et al. Transdermal testosterone may improve ovarian response to gonadotrophins in low-responder IVF patients: a randomized, clinical trial. Hum Reprod 2009:24;349-359
226. Wiser A, Gonen O, Ghetler Y et al. Addition of dehydroepiandrosterone (DHEA) for poor-responder patients before and during IVF treatment improves the pregnancy rate: a randomized prospective study. Hum Reprod 2010:25;2496-2500
227. Kolibianakis EM, Venetis CA, and Tarlatzis BC. DHEA administration in poor responders. Hum Reprod 2011:
228. Perez MM, Gromoll J, Behre HM et al. Ovarian response to follicle-stimulating hormone (FSH) stimulation depends on the FSH receptor genotype. J Clin Endocrinol Metab 2000:85;3365-3369
229. Behre HM, Greb RR, Mempel A et al. Significance of a common single nucleotide polymorphism in exon 10 of the follicle-stimulating hormone (FSH) receptor gene for the ovarian response to FSH: a pharmacogenetic approach to controlled ovarian hyperstimulation. Pharmacogenet Genomics 2005:15;451-456
230. Greb RR, Grieshaber K, Gromoll J et al. A common single nucleotide polymorphism in exon 10 of the human follicle stimulating hormone receptor is a major determinant of length and hormonal dynamics of the menstrual cycle. J Clin Endocrinol Metab 2005:90;4866-4872
231. Hendriks DJ, te Velde ER, Looman CW et al. Expected poor ovarian response in predicting cumulative pregnancy rates: a powerful tool. Reprod Biomed Online 2008:17;727-736
232. Stoop D, Nekkebroeck J, and Devroey P. A survey on the intentions and attitudes towards oocyte cryopreservation for non-medical reasons among women of reproductive age. Hum Reprod 2011:
233. Nelson SM, Messow MC, Wallace AM et al. Nomogram for the decline in serum antimullerian hormone: a population study of 9,601 infertility patients. Fertil Steril 2011:95;736-741
234. La MA, Sighinolfi G, Giulini S et al. Normal serum concentrations of anti-Mullerian hormone in women with regular menstrual cycles. Reprod Biomed Online 2010:21;463-469
235. Shebl O, Ebner T, Sir A et al. Age-related distribution of basal serum AMH level in women of reproductive age and a presumably healthy cohort. Fertil Steril 2011:95;832-834
236. Snieder H, MacGregor AJ, and Spector TD. Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. J Clin Endocrinol Metab 1998:83;1875-1880

237. de Bruin JP, Bovenhuis H, Van Noord PA et al. The role of genetic factors in age at natural menopause. Hum Reprod 2001:16;2014-2018
238. van Asselt KM, Kok HS, Putter H et al. Linkage analysis of extremely discordant and concordant sibling pairs identifies quantitative trait loci influencing variation in human menopausal age. Am J Hum Genet 2004:74;444-453
239. Murabito JM, Yang Q, Fox C et al. Heritability of age at natural menopause in the Framingham Heart Study. J Clin Endocrinol Metab 2005:90;3427-3430; 42480
240. Bromberger JT, Matthews KA, Kuller LH et al. Prospective study of the determinants of age at menopause. Am J Epidemiol 1997:145;124-133
241. Torgerson DJ, Thomas RE, and Reid DM. Mothers and daughters menopausal ages: is there a link? Eur J Obstet Gynecol Reprod Biol 1997:74;63-66; 42481
242. Gold EB, Bromberger J, Crawford S et al. Factors associated with age at natural menopause in a multiethnic sample of midlife women. Am J Epidemiol 2001:153;865-874
243. Voorhuis M, Onland-Moret NC, van der Schouw YT et al. Human studies on genetics of the age at natural menopause: a systematic review. Hum Reprod Update 2010:16;364-377
244. Rosen MP, Johnstone EB, Gillham SJ et al. Is antral follicle count a genetic trait? Menopause 2010:17;109-113
245. Steiner AZ, Baird DD, and Kesner JS. Mother's menopausal age is associated with her daughter's early follicular phase urinary follicle-stimulating hormone level. Menopause 2008:15;940-944
246. van Disseldorp J., Eijkemans R, Fauser B et al. Hypertensive pregnancy complications in poor and normal responders after in vitro fertilization. Fertil Steril 2009:
247. van Disseldorp J., Broekmans FJ, Peeters PH et al. The association between vascular function-related genes and age at natural menopause. Menopause 2008:15;511-516
248. Woldringh GH, Frunt MH, Kremer JA et al. Decreased ovarian reserve relates to pre-eclampsia in IVF/ICSI pregnancies. Hum Reprod 2006:21;2948-2954
249. Keegan DA, Krey LC, Chang HC et al. Increased risk of pregnancy-induced hypertension in young recipients of donated oocytes. Fertil Steril 2007:87;776-781
250. Merce LT, Bau S, Barco MJ et al. Assessment of the ovarian volume, number and volume of follicles and ovarian vascularity by three-dimensional ultrasonography and power Doppler angiography on the HCG day to predict the outcome in IVF/ICSI cycles. Hum Reprod 2006:21;1218-1226
251. Ozturk O, Bhattacharya S, Saridogan E et al. Role of utero-ovarian vascular impedance: predictor of ongoing pregnancy in an IVF-embryo transfer programme. Reprod Biomed Online 2004:9;299-305
252. Koochmeshgi J, Hosseini-Mazinani SM, Morteza SS et al. Apolipoprotein E genotype and age at menopause. Ann N Y Acad Sci 2004:1019;564-567
253. Bonomini F, Filippini F, Hayek T et al. Apolipoprotein E and its role in aging and survival. Exp Gerontol 2010:45;149-157
254. Kok HS, van Asselt KM, van der Schouw YT et al. Heart disease risk determines menopausal age rather than the reverse. J Am Coll Cardiol 2006:47;1976-1983

Chapter 10

Nederlandse samenvatting

References

Dankwoord

Curriculum Vitae

Mijn proefschrift is af! Dit boekje had ik niet kunnen maken zonder de steun van iedereen die direct of indirect betrokken was bij dit onderzoek. Daarvoor wil ik degenen die hebben meegewerkt enorm bedanken, enkelen met name.

Allereerst wil ik alle vrouwen bedanken die hebben meegewerkt aan deze onderzoeken. Zonder jullie zou dit proefschrift niet tot stand zijn gekomen.

Prof. dr. F.J.M. Broekmans, beste Frank, een aantal jaren geleden kwam ik voor het eerst langs voor een wetenschappelijke keuzestage bij de fertiliteit. Ik had nog geen benul van wat AMH en ovariële reserve testen nu precies waren. Maar jouw enthousiasme was ontzettend aanstekelijk en al snel werd ik gepakt door het onderwerp. Ik ben dankbaar voor het door jou gestelde vertrouwen in mij, waardoor je mij de mogelijkheid hebt gegeven om dit onderzoek uit te bouwen naar een promotietraject! Verder wil ik je bedanken voor alle support en ondersteuning tijdens dit promotietraject, ik waardeer de persoonlijke manier waarop je mij de afgelopen tijd hebt begeleid. Dank voor de humor en de quotes! Ik heb veel van je geleerd en het is een eer om een van jouw eerste promovenda te mogen zijn!

Prof. dr. B.C.J.M. Fauser, beste Bart, graag wil ik jou bedanken voor de helikopterview die jij mij gaf tijdens dit promotietraject. Elke bespreking wist jij er weer voor te zorgen dat ik de rode draad van dit promotietraject terugvond. Dank voor je scherpzinnigheid, waardoor een enkele opmerking grote gevolgen kon hebben voor het onderzoek. Ik ben blij dat ik deel uit heb mogen maken van jouw onderzoeksgroep.

Dr. M.J.C. Eijkemans, beste René, zonder jou zou promoveren nooit hetzelfde zijn geweest. Dank voor al je hulp en ondersteuning in de statistiek. Als geen ander kan jij statistiek begrijpbaar maken. Dank voor je nuchtere benadering en voor al je geduld tijdens het uitleggen. Jij bent voor mij vaak de held van de dag geweest. En dank voor je humor, waardoor de statistiekuren niet alleen moeilijk, maar ook erg gezellig zijn geworden.

Prof. dr. B.W. Mol, beste Ben-Willem, vanaf mijn eerste ervaringen met de wetenschap was jij bij het onderzoek betrokken. Als leider van de IPD meta-analyses gaf je mij het vertrouwen om mee te werken, waarvoor ik je dankbaar ben. Dank voor je scherpe, maar vaak terechte kritiek en doorzettingsvermogen, waardoor deze stukken zo ontzettend veel beter zijn geworden.

De overige leden van de beoordelingscommissie: Prof. dr. Y.T. van der Schouw, Prof. dr. A. Franx, Prof. dr. F.L.J. Visseren, Prof. dr. E.P.J.G. Cuppen, dank ik voor het kritisch beoordelen van dit proefschrift.

Dr. J. van Disseldorp, beste Jeroen, wat een voetstappen om in te treden…. Vanaf dag één vond ik het prettig om met je samen te werken en het is een eer om de onderzoekslijn voort te hebben mogen zetten. Dank voor een sterke basis en de hulp de afgelopen tijd. Daarnaast ook dank voor alle gezelligheid. Ik waardeer je als onderzoeker, collega, maar zeker ook als vriend.

Prof. dr. P.M.M. Bossuyt, beste Patrick, dank voor het verschil dat jij hebt gemaakt bij de IPD meta-analyses. Ik ben blij dat jij jouw expertise met ons wilde delen. Ik vond het prettig om met je samen te werken, en ik hoop dat in de toekomst te blijven doen.

Dr. B.C. Opmeer, beste Brent, dank voor jouw statistische expertise en ondersteuning bij de (IPD) meta-analyses.

Drs. K.A. Broeze, beste Kimiko, dank voor alle ondersteuning en het delen van de ervaringen bij de IPD meta-analyses.

Geachte Dr. I.A.J. van Rooij, beste Ilse, Dr. G.J. Scheffer, beste Gabrielle en Dr. A. de Vet, beste Annemarie, dank voor de solide basis waarop ik het onderzoek voort kon zetten.

Dear IMPORT and EXPORT collaborators, thank you for sharing your data with us in this IPD meta-analyses.

Lieve collega IVF-artsen, dank dat jullie mij zo warm opgenomen hebben in de groep. Anna een extra maal dank voor alle tijd en moeite die je hebt genomen om mij op te leiden. Het was niet altijd even makkelijk om in de kliniek te werken naast een promotie onderzoek, maar ik had het voor geen goud willen missen!

Beste Piet, Marian, Annelies, Sandra en Angelique, bedankt voor alles wat ik van jullie heb mogen leren.

Lieve IVF verpleegkundigen en alle collega's van het IVF laboratorium, het was me een genoegen om met jullie samen te werken.

Lieve dames van receptie 38, dank voor al jullie ondersteuning, maar ook voor de gezelligheid. Ik zal nog eens wat kokosmakronen langsbrengen.

Ellis, Ingrid en Tessa, wat krijgen jullie het altijd goed geregeld, duizendmaal dank daarvoor. En dank voor de gezelligheid en mooie verhalen.

Mede onderzoekers van hier en de 'overkant', promoveren zonder jullie zou niet hetzelfde zijn, bij lange na niet zo gezellig! Dank jullie wel voor alle maandag-lunches, kroketten, etentjes, theetjes. Dank voor het mogen ventileren van de welbekende promotiefrustraties. Maar ook dank voor alle humor, wat heb ik gelachen. Het einde is gehaald, maar ik ben blij dat het niet het einde is van onze samenwerking!

Studenten die mee hebben gewerkt aan deze of verwante onderzoeken. Bedankt voor de prettige samenwerking. Madeleine, een extra maal dank voor al het meelezen en de verfijning van de Engelse taal.

Arts-assistenten gynaecologie, ik kijk er naar uit om de opleiding te starten en jullie collega te worden.

Lieve Jaguars, bedankt voor alle leuke momenten die we hebben gehad en dat ik alles met jullie kan en mag delen. Dank voor jullie support, ook al ben ik de laatste tijd wat vaker afwezig geweest. Vanaf nu komt er weer meer tijd voor alle leuke Jaguar plannen. Ik ben dankbaar voor alle herinneringen en ik kan niet wachten om er meer te maken!

Lieve oud-huisgenoten van het CBR, wat was het leuk om bij en met jullie te wonen. Dank voor alle gezelligheid en het thuisgevoel op de Goedestraat. En voor nu; beter een goede buur, dan een verre vriend…

Lieve vrienden van de studie, dank voor samen studeren, opdrachten delen, gedeelde tentamenstress en alle leuke momenten, avondjes uit en weekendjes weg die we hebben gehad. Ik ben blij nu als collega's nog steeds zoveel leuke momenten met jullie te delen en ik kijk uit naar interdisciplinaire consulten.

My foreign friends, thank you for your friendship, even if we're miles apart. Lisa and Sarah, I can't wait for the next reunion.

Alle vrienden die ik niet in een 'categorie' kan indelen, bedankt voor jullie vriendschap en interesse in mij en de support tijdens dit promotietraject.

Er zijn nog een aantal vrienden ik eruit wil lichten:

Marjolijn, dank dank dank! Voor alle momenten dat je mij de positieve kant van alles kan laten zien, zelfs als ik ze écht zelf niet meer kan bedenken. Dank voor je creatieve input. De dagelijkse oppeppers. Vanaf de eerste ontmoeting zat het goed, dat hoedje gaat niet op…

Lidy, ook met jou was het vanaf het eerste moment duidelijk, wij zouden goede vriendinnen zijn. Dank voor je support, zelfs al is het van mijlenver! Alles kan ik jou vertellen, dank je wel voor je relativeringsvermogen en dat je me om alles kunt laten lachen. Ik ben je dankbaar voor jouw geloof en vertrouwen in mij.

Celine, jouw enthousiasme is zó aanstekelijk. Je kunt me er altijd aan herinneren waarom ik dit vak ook al weer zo leuk vind. Dank voor alle vrolijkheid die je brengt en voor je begrip, als geen ander weet jij hoe het is om wetenschap en de kliniek te combineren.

Eline, waar werkgroep 23 wel niet goed voor is geweest…een nieuw studiemaatje met dezelfde aanpak, een goed vriendinnetje, een gezellig buurvrouw. Er zijn weinig mensen met wie je 3 maanden in Afrika één kamer of tent kunt delen, maar met jou was het een feest en ik kan alleen maar positief terugkijken naar die tijd. Ik ben blij dat we nog steeds zoveel kunnen delen. Asante sana.

Nadine, lief nichtje maar vooral ook vriendinnetje. Vanaf vroeger zat het al goed; brieven, logeerpartijen en bezoekjes zijn uitgegroeid tot een van de sterkste vriendschappen. Dank je wel voor al je interesse en steun de afgelopen jaren. Het idee dat je er altijd voor me bent, maakt mij sterker.

Sarah, twee hele verschillende personen, maar wel twee handen op één buik. Vroeger, maar nu nog steeds. De eerste jaren alles samen, nu ieder zijn eigen weg. Ondanks dat weet ik dat je altijd achter me staat. Dank voor je creatieve ondersteuning. Wat heb ik zin in onze reis!

Annemieke, een verplicht kaartje op de middelbare school werd samen vakken volgen, samen dansen, samen uit. Nu gestudeerd in een andere stad, maar de vriendschap blijft. Dank voor alle gezelligheid en alle momenten die we hebben kunnen delen.

Lieve dames, dank jullie wel dat jullie zulke fantastische vriendinnen zijn! Jullie zijn me dierbaar!

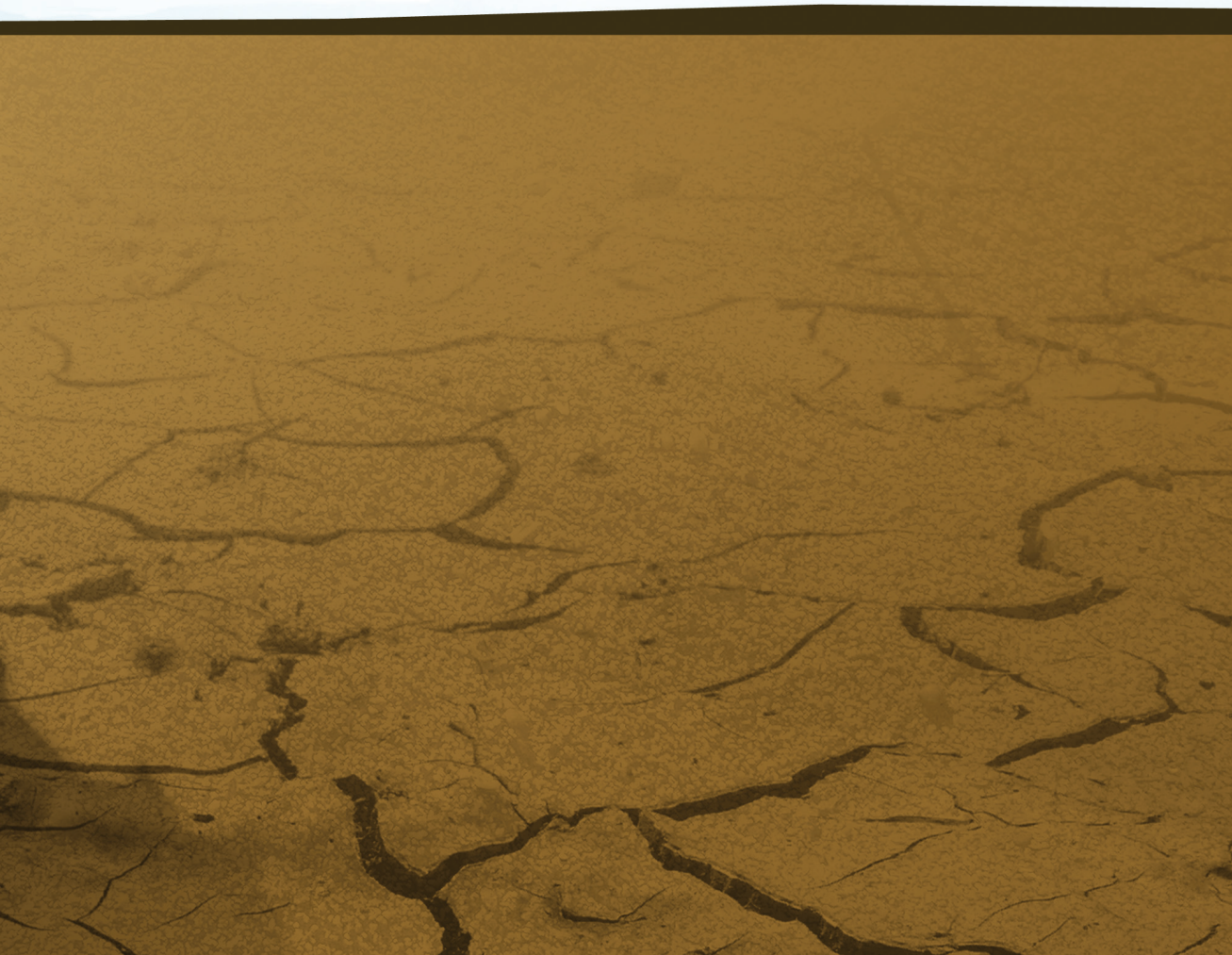
Lieve paranimfen, wat een eer om door jullie, twee zulke sterke en bijzondere vrouwen, bijgestaan te worden vandaag!
Anne, lieve roomy, een kamer met een onbekende delen voor congres is nog nooit zo'n succes geweest. Wat een super week in Barcelona, en dan met een nieuwe vriendin terug naar huis. Dank voor al je adviezen op professioneel en persoonlijk gebied, maar vooral voor alle gezellige momenten hier en in Canada. Ik ben blij dat onze vriendschap steeds sterker wordt.
Jenneke, sis, samen in hetzelfde schuitje. Wat is het fijn om een maatje te hebben, en helemaal eentje zoals jij. Jou hoef ik niks uit te leggen. Avondjes in het UMC, of in de stad, het is altijd gezellig. Jouw enthousiasme en doorzettingsvermogen waardeer ik enorm. Wat ben ik blij dat we behalve collega's nu ook goede vriendinnen zijn.
Dames, bedankt voor alle support tijdens dit promotietraject, en zeker de laatste maanden. Ook jullie zijn me dierbaar!

Lieve familie Broer en Nieuwenhuis, dank voor jullie interesse in mij en dit onderzoek, maar ook voor alle gezellige familiemomenten.

Lieve Wijnand, Marnix, Suzanna en Wessel. Opgroeien met z'n vieren was fantastisch. De 'grote' en de 'kleine' bestaan niet meer, het is één gezellige mengelmoes, waarbij we elk wat anders met elkaar kunnen delen! Wessel wat gezellig dat jij erbij bent gekomen. Ik, toch een beetje de vreemde eend tussen al deze ondernemers, ben dankbaar dat jullie ondanks dit verschil altijd veel interesse getoond hebben in mijn werk en onderzoek. Ook dank voor alle gezelligheid, alle etentjes, drankjes, feestjes, Sinterklaas en de wintersport. Ik hou van jullie.

Lieve papa en mama, ontzettend veel dank voor alles en zoveel meer! Voor het warme nest, de veilige thuishaven, het vertrouwen en geloof dat jullie in mij hebben. Dank voor de 'je kunt meer dan je denkt'. De basis die jullie mij hebben gegeven en de support die ik nog altijd van jullie krijg zorgen ervoor dat ik kan doen wat ik doe! Ik ben blij dat we nog steeds zo'n sterke familieband hebben. Laten we Sinterklaas en de wintersport en alle andere familietradities in ere houden. Ik hou van jullie.

Chapter 10



Simone Louise Broer was born on July 3rd, 1985 in Naarden, as the third of four children. There, she grew up with her brothers Wijnand and Marnix and her sister Suzanna. She attended the gymnasium at the 'Willem de Zwijger College' in Bussum, from which she graduated cum laude in 2003. Subsequently she started medical school at the University Utrecht.

During medical school gynaecology and obstetrics caught her interest and therefore she started her scientific career with her first study on ovarian reserve tests. She then followed an additional internship Gynaecology and Reproductive Medicine at the Universita degli Studi di Siena, Italy in 2007 (supervision: Prof. F. Petraglia). After her return, she continued her studies on ovarian reserve tests, but interrupted these in 2008 for a surgical internship at the Muhimbili National Hospital, Dar Es Salaam, Tanzania. After obtaining her medical degree in 2009, she started working as a fertility physician at the University Medical Center Utrecht and during this time she continued her research project as a PhD-student at the Department of 'Woman and Baby' under the supervision of Prof. dr. F.J.M. Broekmans and Prof. dr. B.C.J.M. Fauser.

In august 2011 she will start her residency in Obstetrics and Gynaecology at the Gelre Hospital in Apeldoorn (supervision: Dr. K.M. Paarlberg).

Simone Louise Broer studied Medicine at the Utrecht University. During medical school she started her PhD research on ovarian ageing. After her graduation she started working at the Department of Reproductive Medicine at the University Medical Center Utrecht as a PhD student and as a fertility physician.

Her thesis aims to evaluate the true value of ovarian reserve tests (ORTs) in the assessment of the current and future ovarian reserve status. It was demonstrated that the ORTs do reflect the current quantitative but not the qualitative aspects of the ovarian reserve status. For the prediction of the future ovarian reserve status it could be demonstrated that an age-specific Anti-Müllerian Hormone value can give an individualized prediction of the age category in which menopause will occur. Prediction of a woman's reproductive lifespan could lead to the exploration of primary prevention of age related infertility and menopause related conditions.

assessment of

Current and future ovarian reserve status

by Simone Broer