Fertility Forecasting: Advances in the Role of Anti-Müllerian Hormone as a Reproductive Biomarker

Madeleine Dólleman

Cover Design: Soha Ehsani & Shamim Ehsani
Cover Photography: L.J. Vasak
Layout: Roy Sanders
Printed by: Gildeprint Drukkerijen BV, Enschede, the Netherlands

ISBN: 978-90-393-6066-8

This thesis was funded by the Department of Reproductive Medicine and Gynaecology, University Medical Center Utrecht, the Netherlands and the department of Obstetrics and Gynaecology, Academic Medical Center Amsterdam.

The author gratefully acknowledges financial support for printing this thesis by the Division Woman & Baby University Medical Center Utrecht, Gilles Hondius Foundation, Gilles Hondium Foundation, Beckman Coulter, Memidis Pharma B.V., Medical Dynamics, BMA B.V., Abbott B.V., Origio B.V., Bayer HealthCare Pharmaceuticals.

Fertility Forecasting: Advances in the Role of Anti-Müllerian Hormone as a Reproductive Biomarker

Voorspelling van Vruchtbaarheid:
De Rol van Anti-Müller Hormoon als Biomarker voor Ovariële Reserve
(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 17 december 2013
des middags te 4.15 uur

door

Madeleine Dólleman
geboren op 1 november 1985 te Nairobi, Kenia

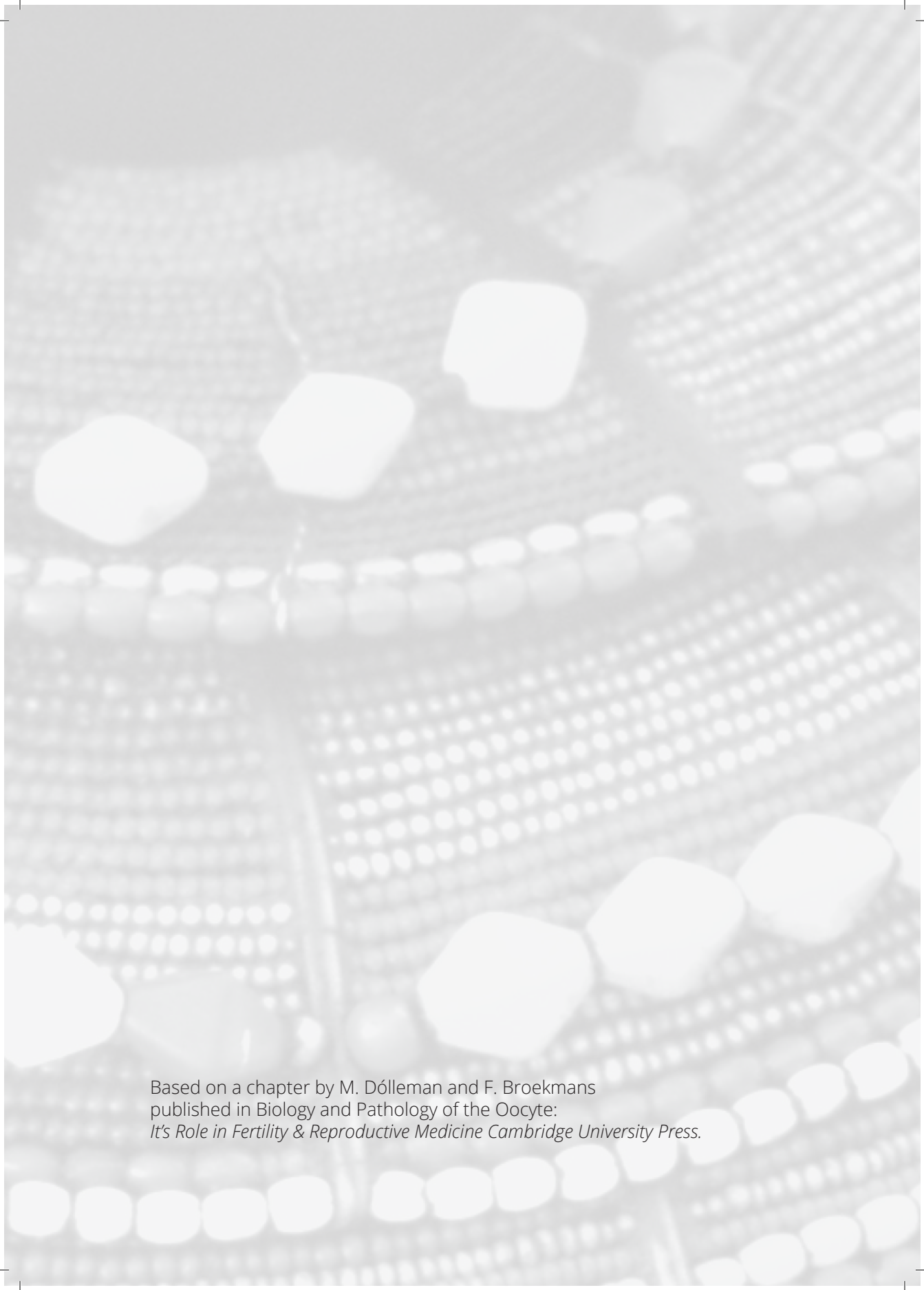
Promotoren: Prof. dr. F.J.M Broekmans
Prof. dr. ir. Y.T. van der Schouw
Prof. dr. B.W.J. Mol

Copromotor: Dr. J. van Disseldorp

Many stories are whispered through the beads of traditional Africa,
one of them is the story of fertility

For my parents

Table of Contents



Based on a chapter by M. Dólleman and F. Broekmans
published in Biology and Pathology of the Oocyte:
It's Role in Fertility & Reproductive Medicine Cambridge University Press.

Chapter 1

General Introduction

"Relative Contribution of Advanced Age and Reduced Follicle Pool Size on Reproductive Success: The Quantity – Quality Enigma"

Introduction

It is a well-known phenomenon that as a woman becomes older, her chances of achieving an ongoing pregnancy decrease. This is largely attributed to ovarian ageing, the age related decline in the quantity and quality of oocytes in the ovaries. At birth every woman has a certain endowment of oocytes. This number of oocytes decreases at various rates during life until the ovarian reserve is exhausted and menopause is reached (1). The gradual decline in oocyte quantity with age is accompanied by a concomitant decrease in oocyte quality.

The age related decrease in female fertility has important repercussions in Western societies as the trend to delay child-bearing continues resulting in more and more women requiring the help of assisted reproductive techniques (ART) such as in vitro fertilisation (IVF). Unfortunately ART is only a successful solution for approximately 50% of women with an unfulfilled child wish underlining the importance of primary prevention of female age related infertility.

The process of female reproductive ageing is characterised by four reproductive milestones: decreasing fertility, the end of the female fertile lifespan, the menopausal transition and finally, menopause (2;3). Of these, menopause is the most noticeable milestone as the absence of menstrual cycles does not go unnoticed while the subtle changes of the first three milestones can be easily overlooked. The reproductive milestones are suggested to happen in a fixed temporal fashion with the end of natural fertility occurring approximately 10 years before the final menstrual period, and the start of declining fertility some 10 years before that (2;3). The individual variability of ovarian ageing between women of the same chronological age is large: the range for the age at natural menopause extends from forty to sixty years with a mean age of fifty-one years (4). Similarly, the age at the end of natural fertility ranges from thirty-one to forty-six with an average age of forty-one years (5). This would mean that some women will already start to experience decreasing fertility in their early twenties while other will remain almost normally fertile until their late thirties. Identifying women with early menopause and thus early subfertility could be used for the primary prevention of infertility by counselling those women to conceive early or to cryopreserve their oocytes. However, the large variability in the reproductive ageing process has taught us that age alone cannot explain the variation in ovarian ageing. These insights have prompted research to find a more reliable marker than chronological age in

predicting age at natural menopause from which the age at the end of the female fertile lifespan can be deduced. One such marker of ovarian reserve (the number of follicles remaining in the ovarian) into which a phenomenal amount of research has been done over the past 10 years, is anti-müllerian hormone (AMH). AMH is a hormone originally known for its role during embryogenesis where it plays an important role in sexual differentiation. While in males, AMH induces regression of the Müllerian ducts (the process to which it owes its name), in females the absence of AMH results in the development of the Müllerian ducts into the uterus, fallopian tubes and the upper part of the vagina (6). AMH in females is produced by the granulosa cells of antral follicles in the ovaries. Next to sexual differentiation AMH plays an important endocrine role, especially in females, where it plays a primary role in primordial follicle recruitment (6). AMH is secreted into the circulation by the granulosa cells of antral follicles in the ovaries. After secretion AMH can be measured in human serum through an Enzyme-linked immunosorbent assay (ELISA). Since the discovery of AMH different assays have emerged with varying sensitivity for AMH, the two main ones being Diagnostics Systems Laboratories assay and the Immunotech Beckman Coulter assay. Due to the use of two different primary antibodies in these assays, there was no clear consensus whether these assays were directly comparable. Driven by the need to have a standardized, comparable measured, these companies fused and have released a newer assay, the Gen-II assay, to replace the 2 older assays (7). Through both its role in cyclic recruitment and its correlation with the ovarian reserve, AMH has been increasingly used in clinical practice for both prognostic purposes, such as prediction of ovarian response during IVF treatment or age at menopause and diagnostic purposes such as the identification of women with premature ovarian insufficiency or polycystic ovary syndrome.

In summary, the increased demand for assisted reproduction in combination with the current fickle economic climate and a medical paradigm that is progressively shifting to a more individualised health care has called for research to establish whether it is possible to use reproductive biomarkers for individual fertility forecasting. 'Fertility forecasting', in this broad sense encompasses the prediction of outcomes after controlled ovarian hyperstimulation during IVF treatment, the prediction of pregnancy in IVF patients and the derivation of the end of a female's fertile lifespan from her individually predicted age at menopause. This PhD thesis focuses on anti-müllerian hormone, a prominent reproductive biomarker, and its value in individualised fertility forecasting.

Introduction to Reproductive Ageing

At birth every woman has a certain endowment of oocytes. This number of oocytes decreases at various rates during life until the ovarian reserve is exhausted and menopause is reached (1). Renewal of the oocyte pool from pluripotent stem cells has so far been denied, but recent studies have elicited possible new insights into this field (8). The gradual decline in oocyte quantity with age is accompanied by a decrease in oocyte quality. This is substantiated by decreased pregnancy rates, increased miscarriage rates and an increase in the rate of aneuploidy leading to offspring with trisomic karyotypes (9;10). Also, a growing incidence of unexplained infertility is apparent in women trying to achieve a pregnancy at a higher age (11).

The introduction of effective contraceptive methods in the 1960s and the growing participation of women in the labour force has resulted in a major change in reproductive behaviour (12). The average age at the birth of the first child has increased from approximately twenty-four years of age in 1970 to the age of thirty or over in recent years (2;13). In addition, the completed fertility rate (number of children born per woman) has decreased considerably and a growing proportion of women seeks the help of assisted reproductive technology to conceive. As modern infertility treatments can only help around fifty percent of these women, a considerable proportion of women will remain childless involuntarily, with increased levels of personal distress and grave effects on relationship stability (14;15). The continuing trend to delay child-bearing does not only have a large impact on population demographics, but the annual costs for society from infertility treatments and ART-related complications, such as multiple pregnancies, are also high (16;17).

This general introduction will discuss on factors that influence the quantitative and qualitative depletion of the ovarian reserve and how this affects cyclic ovarian function as well as the chances for reproductive success. The way in which quantity and quality decline are interrelated will be reviewed. Furthermore, it will address available methods for assessing a woman's individual reproductive age status and how this information can be used to predict her current and future fertility potential.

Follicle quantity and cyclic ovarian function

During the fourth month of fetal development the ovaries contain between six and seven million oocytes. These oocytes are surrounded

by a layer of flat granulosa cells and together they are referred to as the primordial follicle pool (18-20). In the second half of fetal life, the number of oocytes decreases rapidly due to a process of apoptosis. Consequently only one to two million oocytes are left in the ovaries at birth (21). The process of apoptosis continues after birth, but at a slower rate so that at menarche approximately 300,000 to 400,000 primordial follicles remain. In a woman's reproductive years, the gradual decline in follicle quantity is responsible for the occurrence of two reproductive milestones, the onset of overt cycle irregularity and menopause, being the two final events of the quantitative ovarian ageing process (22).

Throughout the period in which follicle numbers decline, the number of continuously present antral follicles remains sufficient to ensure the monthly process of single dominant follicle development and ovulation. It is not until only a few thousand follicles remain in the ovaries that the reduced negative feedback from the ovaries to the hypothalamo-pituitary unit leads to elevated gonadotropin levels, resulting in dysregulated folliculogenesis. Soon thereafter, the availability of sufficient antral follicles for cyclic follicle development can not always be ensured, causing the menstrual cycle to become irregular (23;24). The years before the final menstrual period, when variability in the menstrual cycle is increased, are referred to as the menopausal transition (23). During this period the rate of exhaustion of primordial follicles occurs at an accelerated pace. Eventually, the cycle stops completely, the event known as menopause, which marks the last reproductive milestone (24-27). Menopause is defined as a period of amenorrhea of at least twelve months. The average age at which menopause occurs is fifty-one years, with a range from forty to sixty years (2). Mathematical models that have been designed to portray the age-related depletion of the follicle pool show that the pool has been exhausted when menopause occurs.

A fixed temporal relationship between the last two reproductive milestones is thought to be present (***Figure 1***). Prospective recordings of mean cycle duration and variation around the mean age at menopause have made it possible to assess the onset of the menopausal transition. It was subsequently demonstrated that age at menopause was directly related to age at which cycle irregularity was initiated, with a reasonably fixed interval of approximately 5 years (2;28).

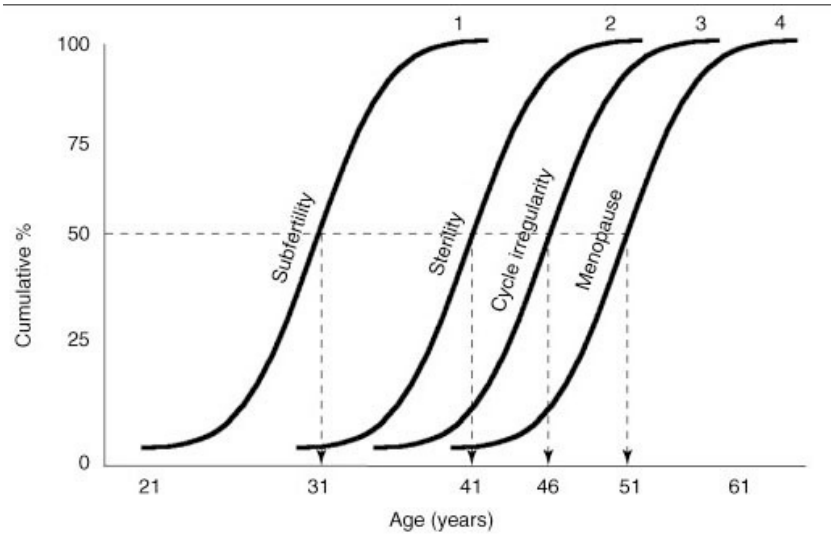


Figure 1. Graphical representation of the variation in the ages at the four reproductive milestones. Curve 1 represents variation in onset of subfertility and is supported by data from Eijkemans (88). Curve 2 shows variation in the end of natural fertility and is based on data from Bouchard (38). Curve 3 represents individual variability in occurrence of cycle irregularity from den Tonkelaar *et al.* The variation in menopausal age has been derived from data by Treloar and Broekmans (4;89) The average age at which the event occurs in the general population is shown on the x-axis. This figure represents the notion that the occurrence of reduced fertility occurs approximately ten years before sterility and that the menopausal transition precedes the onset of menopause by approximately five years. Reproduced from Broekmans *et al.* Female reproductive ageing: current knowledge and future trends, *Trends in Endocrinology & Metabolism*, 18(2), p59, by permission of Oxford University Press (3).

Quality decline and fertility

The age-related decline in follicle numbers is accompanied by a diminishing oocyte quality. Oocyte quality is not uniformly defined. It can be assessed through the ability of achieving an ongoing pregnancy or by assessing the occurrence of aneuploidy or miscarriage. The declining ability to give birth to a healthy child and the increasing time to pregnancy with increasing female age has been documented from numerous sources. With increasing female age, the risk of a pregnancy resulting in early pregnancy loss or the birth of a child with a numerical chromosome abnormality will become more and more substantial (9;10). An increase in the occurrence of meiotic non-disjunction is accepted to lie at the heart of the decreasing oocyte quality. Meiotic non-disjunction leads to frequent aneuploidy in oocytes and the early embryo at higher female ages. In women approaching the age of

forty, the majority of oocytes and embryos are chromosomally abnormal as confirmed by research on chromosome numbers in embryos derived from in vitro fertilisation (IVF) programs (29-32). Several mechanisms are associated with a decline in oocyte quality, including a life-long accumulation of damage to the oocyte and changes in the granulosa cells that surround the oocyte. It is also theorized that there may be inherent differences in the quality of germ cells from which oocytes are formed (2). A two-hit model exists which suggests that there is a first hit due to an inherent reduction in the frequency and pattern of recombination in a fraction of oocytes from the beginning. The second hit encompasses the accumulation of damage to the oocyte and damage to the follicle. This damage results from oxidative stress, lifestyle factors and micro-environmental factors such as a decreased circulation around the leading follicle or impaired functionality of the granulosa cells (33-35). Moreover, accumulation of damage with age results in inadequate reserves of energy due to mitochondrial dysfunction which may lead to chaotic mosaicism in human pre-implantation embryos (36). Furthermore, gene expression patterns appear altered as implied by transcriptome analysis of young and aged metaphase II oocytes from human and mouse suggesting that an abundance of factors in chromosome, cell cycle, spindle regulation and other cellular processes change.

Much like the decline in follicle quantity, the waning oocyte quality will also be reflected in reproductive milestones, although these are not as noticeable for the individual woman as the menopausal transition and menopause.

The first reproductive milestone is the onset of declining fertility at a mean age of 31 years (***Figure 1***). The best notion of this declining natural fertility stems from semen donation studies where the male factor was controlled for (***Figure 2***) (11;37). Next is the advent of natural sterility, defined as the loss of the ability to conceive and give birth to viable offspring, even if well timed exposure is attempted for several years. Knowledge on the advent of natural sterility stems from a nineteenth century population study, in which the highly religious nature of the cohort prohibited the use of contraception and therefore procreation until natural sterility had been reached was ensured. It could be shown that age at last child birth, as a proxy for the loss of natural fertility, was on average at the age of 41 years (38). The gradual decline in fertility during the fourth decade of life passes largely unnoticed from the individual's perspective. Monthly ovulations are believed to ensure that prospects for pregnancy occur, while in fact fecundity rates will distinctly decline after the age of thirty.

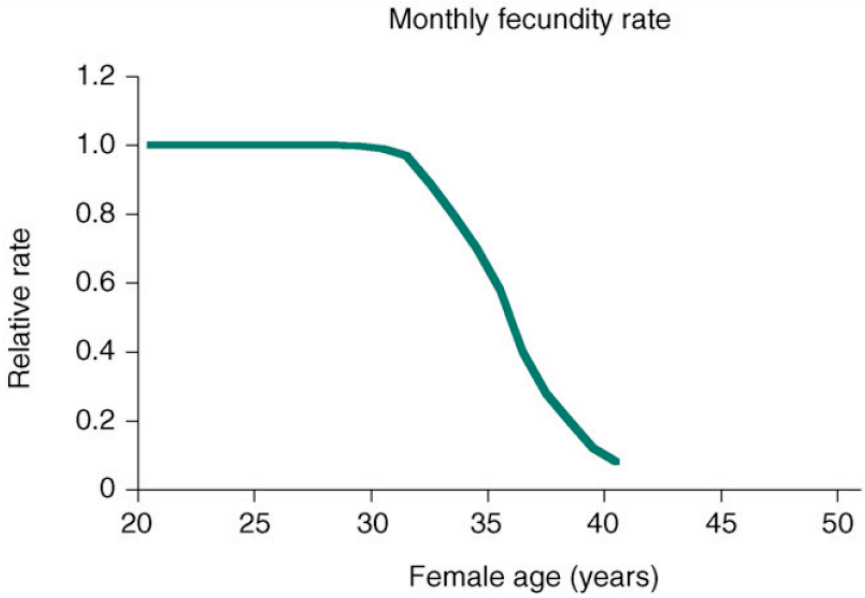


Figure 2. Graphical representation of decreased monthly fecundity rates with increasing age. It demonstrates that after the age of thirty-one years there is a rapid decrease in the monthly fecundity rate. Reproduced from Broekmans *et al.* Female reproductive ageing: current knowledge and future trends, *Trends in Endocrinology & Metabolism*, 18(2), p60, , by permission of Oxford University Press (3).

A total of four reproductive milestones are thus believed to be present, namely decreasing natural fertility, sterility, menopausal transition and menopause. The fixed time interval, as demonstrated by the relationship between the onset of menopausal transition and menopause, may also be true for the milestones of quality decline. The above-mentioned nineteenth century population study also demonstrated that a reduced birth rate in the early stages of marriage (between age twenty and thirty years) was associated with an early age at last child birth (approximately at age of 35 years). This suggests that early sterility is preceded by a decrease in natural fertility from before the age of thirty years (38).

The concept of a fixed-temporal relationship between the quantity and quality milestones, though plausible, has not been substantiated by a vast amount of clinical and experimental observations. However, a study in women who had both reduced pregnancy chances and obviously reduced numbers of antral follicles for their age group, has indicated that

these women will enter menopause earlier than women without these characteristics (39). The same line of evidence has been shown from smaller studies relating either poor response in IVF or elevated basal FSH levels with subsequent occurrence of menopause (3). It is important to realise, however, that ultimate proof for this concept is impossible to obtain, as it would require highly refined assessment of fecundity in a population not applying contraceptive measures and challenging fertility until the age at which natural sterility is reached. Availability of such a cohort in the current era is illusive (39;40).

From the presumed relationship between the four reproductive events it is suggested that the end of natural fertility occurs approximately 10 years before menopause is established. During these 10 years, ovulated oocytes may get fertilised, but due to a high frequency of chromosomal abnormalities either implantation may fail or, if implantation is successful, the implanted embryo may produce a pregnancy that terminates at an early stage (41). The increase in female infertility with increasing age is thus two-fold. First of all, monthly chances of conception are lower and secondly there is a higher probability that the pregnancy will terminate around conception or implantation (2). The former could be due either to depletion of the number of follicles or the demise of quality while the latter can be interpreted predominantly as a problem of oocyte quality. It could be argued that other factors leading to increased miscarriage rates in women of higher chronological age may also play a role. However, in oocyte donation programs, where oocytes from young fertile women are placed in older non-fertile women, both pregnancy and miscarriage rates have been very satisfactory, maintaining the view that it is largely an ovarian factor, and not a uterine or other factor (42).

Variability in reproductive ageing

The individual variability of ovarian ageing between women of the same chronological age is large. The range for the age at menopause extends from forty to sixty years with a mean age of fifty-one years (4). Variation of age at menopause has a Gaussian distribution in which there are fewer cases at the extremes of age. This does not include women diagnosed with premature ovarian insufficiency (women who reach menopause before the age of forty) because they are not thought to represent natural variation in age at menopause (43). Similarly, the age at which women start to notice menstrual cycle irregularity (associated with ovarian reserve

exhaustion) ranges from thirty-five to fifty-four years with an average age of forty-six years (5). The mean age at which a woman reaches natural sterility is difficult to establish. One study assessed the age at which women gave birth to their last child in 1040 women born in Canada in the second half of the nineteenth century. The maternal age at the birth of the last child was used as a proxy for calculating age at the end of natural fertility, and was shown to occur on average at the age of forty-one years with a spread from twenty-three to fifty-one years (38).

The fact that it is not uncommon for women to reach the end of natural fertility at a young age sparked the realisation that chronological ageing (i.e. a woman's age in years) and biological ageing (i.e. the functional age of the ovaries) do not always coincide. From studies performed in IVF populations we have learnt that ovarian response to controlled ovarian hyperstimulation (COH) during IVF treatment, as measured by the number of oocytes retrieved, reflects the quantitative aspect of the ovarian reserve. Young women with a repeated poor response to COH (defined as the retrieval of a small number of oocytes) tend to enter menopause earlier than women with a normal response (39). Such a poor responder is thus seen to have advanced biological ageing for her chronological age and is predisposed to go through all four reproductive milestones earlier than someone of similar age with a normal response to COH. In such women with a young chronological age but an old biological age, other factors must be implicated that have an important influence on the ovarian ageing process.

The quantity-quality enigma

The decline in both quantity and quality has been addressed separately, but neither process alone can really explain the severely decreased fecundity at the point where a woman may first experience cycle changes due to exhaustion of her ovarian reserve. At this point a woman still has approximately 25,000 oocytes (22). This is considerably less than her endowment of six to seven million oocytes during fetal life, but it seems like more than enough to maintain fertility as regular ovulation cycles occur until then. There must then be a complicated interrelationship between oocyte quantity and quality. In a mouse-study it was shown that after the removal of one ovary in newborn mice, thus halving the ovarian reserve, the quality of the oocytes in the other ovary was considerably worse. The unilaterally ovariectomised mice had not only an early onset

of cycle irregularity, but also a much earlier rise in the rate of aneuploidy in the offspring (44). Another mouse-study revealed an increase in the number of metaphase II oocytes with unaligned chromosomes from depleted mouse ovaries, while spindle size decreased with advancing age, irrespective of pool size (45).

In humans it has been suggested that women with a reduced ovarian reserve due to ovarian surgery have an increased rate of trisomy-21 offspring (46;47). This is in line with two other studies that have suggested that women who have an aneuploid conception are often found to have higher levels of FSH (48;49). A third study suggests that women with a history of a Down syndrome pregnancy at a young age, had subtle signs of limited ovarian reserve as measured by slightly higher FSH and slightly lower AMH values (50;51). The increased level of FSH was suggested to represent early depletion of the primordial follicle pool independently from age. All this would support the idea of a close relation between quantity and quality in the ovarian ageing process. However, there are various observations that contradict this one-to-one relation between quantitative and qualitative demise. Follow-up studies on the value of basal FSH levels in predicting aneuploidy failed to clearly confirm this obvious relation after correcting for female age (49;52).

The hypothesis that there is a direct relationship between oocyte quantity and oocyte quality has become known as the limited oocyte pool hypothesis and was introduced by Warburton in 1989. It is based on the idea that in a young woman an oocyte of suboptimal condition would not become the dominant follicle because of an abundance of better quality oocytes. On the contrary, in an older woman with a small ovarian reserve a faulty oocyte would be more likely to become the dominant follicle and be more likely to exhibit chromosome nondisjunction (33). Although feasible, this hypothesis does not address the issue of age-related damage accumulation, and merely assumes a hierarchy in oocyte quality present from the prenatal life phase onwards.

The possible inconsistency in the relationship between quantity and quality in the ovarian ageing process has been underlined by several studies from ART programs. Young women with a low ovarian reserve, as represented by a poor response to controlled ovarian hyperstimulation during in vitro fertilisation, have higher pregnancy rates than older women with a similar poor response (53-55). Similarly, women in whom a higher number of oocytes are retrieved after controlled ovarian hyperstimulation (COH), have higher pregnancy chances than women with a lower number of

eggs retrieved. A combination of young age and a high number of oocytes retrieved is a good measure of reproductive success. However, a young female with a low number of retrieved oocytes still has a higher chance of achieving a pregnancy that an older female with a higher number of retrieved oocytes as shown in ***Figure 3*** (56). This demonstrates that there is not a one to one ratio in quantity versus quality decline and suggests a complex interplay of factors still poorly understood today. For now, the available data seem to point towards the relationship between quantity and quality being female age-modulated, where low quantity leads to poor quality but with either protective or augmenting effects of female age. This is portrayed in ***Figure 4***. Supportive evidence comes from a study performed in women with a reduced ovarian reserve, as measured by extremely low anti-müllerian hormone values, where reasonable pregnancy-chances were observed, especially in those women of a

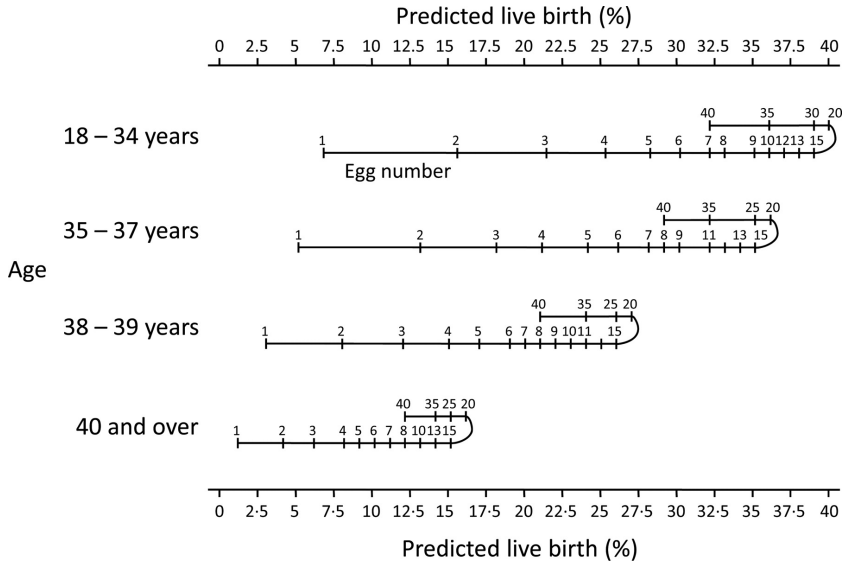


Figure 3. Nomogram of predicted live birth rates according to female age and the number of oocytes retrieved after controlled ovarian hyperstimulation during IVF. It demonstrates three things. Firstly, that live birth rates are influenced by age. Younger women have a better prognosis than older women. Secondly, it shows that a higher number of oocytes is associated with a higher live birth rate, with fifteen oocytes being the optimal number. Lastly it shows that the effect between live-birth rate and the number of oocytes retrieved is age-modulated, where young poor responders have a reasonably good prognosis. Reproduced from Sunkara *et al.*, Association between the number of eggs and live birth in IVF treatment: an analysis of 400 135 treatment cycles, *Human Reproduction, 26(7).p1773,* by permission of Oxford University Press. (56).

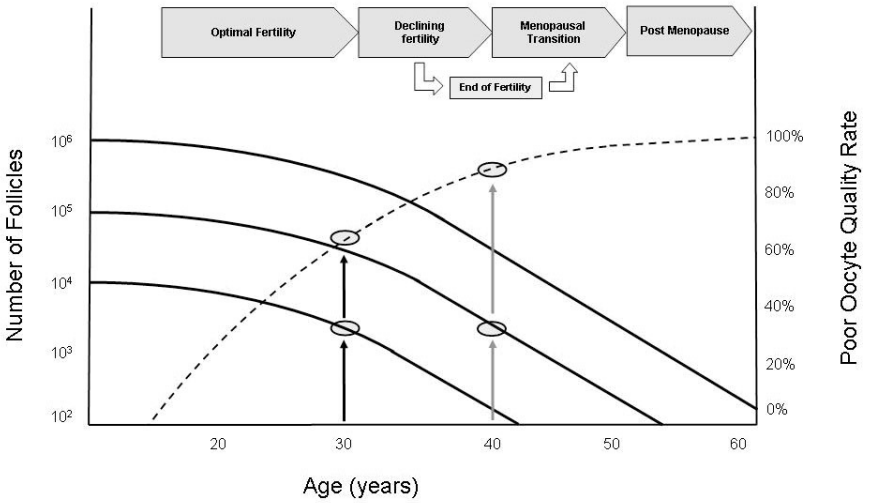


Figure 4. Solid lines represent variability in quantitative decline of oocytes based on different endowments at birth. The dotted line represents the increasing proportion of poor quality oocytes with increasing age based on the theory that damage accumulates with increasing age. It furthermore explains why young women with a low quantity can still have reasonable pregnancy prospects due to relatively good oocyte quality. Reproduced with permission from Broer *et al.* 2011 (87).

younger age (57). Moreover, in women with a low ovarian reserve (as represented by a poor response to COH) the proportion of miscarriages and trisomic pregnancies appeared to become more clearly increased in women at higher age (47;58). A possible explanation may be that quantity in young women is much more variable, so that an observed low quantity in young women is far less consistent and thus less indicative of poor ovarian reserve than a low quantity at high age. The consequences attached to it (i.e. loss in quality) may thus also be less evident. The young poor responder may therefore more often be based on a chance finding (55). To further understand the complex interaction between quantity and quality, genetic factors and environmental influences on reproductive ageing may also need to be considered.

Determinants of reproductive ageing

The variability of age at menopause together with the observation that age at menopause is highly heritable (estimates of heritability range from 30-85%), has led to the search for a gene or set of genes that determines

at what age a woman enters menopause (59-65). It is thought that the identification of such genetic loci would not only explain the timing of menopause but all preceding reproductive milestones as well. For a long time proposed candidate genes included those with a role in fetal ovarian development, primordial follicle maturation, follicular apoptosis, ovarian vascularisation as well as genes associated with premature ovarian insufficiency (3). Currently, the paradigm seems to be shifting towards a search for genes that influence general health and ageing and not only ovarian function. Genome wide association studies (GWAS) have demonstrated that genes involved in DNA repair and function, autoimmunity, neuroendocrine pathways, and genes associated with ovarian function all play a role in regulating age at menopause (66). Other studies have indicated a role for genes associated with vascular support. Both an increase in cardiovascular risk factors like hypertension, obesity, atherosclerosis and hypercholesterolemia and mutations in factor V Leiden, clotting factor VII as well as genes involved in atherosclerosis have been associated with earlier menopause. (67-71). The hunt for genes associated with age at menopause has thus far focussed on identifying variation in the decline of oocyte quantity and not yet on oocyte quality.

The functional lifespan of a woman's ovaries may also be influenced by environmental factors. Of such factors the effect of smoking has been documented most thoroughly. Smoking has been reported to be associated with lower pregnancy rates and earlier onset of menopause and an increased risk for second meiotic nondisjunction causing trisomy 21 may be present in women who smoke and take contraceptives (72;73). Polycyclic aromatic hydrocarbons, found in cigarette smoke and air pollution, interact with the aryl hydrocarbon receptor to cause reproductive defects (74). Smoking accelerates follicle damage and can induce ovarian failure. Furthermore, the identification of DNA repair, as one of the candidate genes involved in menopausal age, could also explain the association between menopausal age and smoking. The damage that results from cigarette smoke activates DNA repair mechanisms that may be associated with age at menopause (66). Women with extreme malnutrition are known to enter menopause earlier showing that nutritionally status influences the rate of reproductive ageing as well, but this has not been elaborately studied (2). In conclusion, the rate of follicle depletion and oocyte quality is an interactive puzzle of inherent and environmental factors.

Assessment of ovarian age

The demand for a critical assessment of ovarian age is increasing as more, older women seek infertility treatment. Adequate evaluation of reproductive age could open doors to individualised patient counselling, personalised treatment protocols (what stimulation dosage to administer for achieving an optimal response) and potentially to abstaining from further treatment in a null-prognosis group of patients. Several markers have been identified that can predict the current ovarian reserve. Ovarian reserve tests (ORTs) reflect the numbers of follicles that are left in the ovaries, therefore these markers are suggested to reflect the quantitative aspect of ovarian age. It is not possible to measure this directly, but it has been shown that by measuring the number of antral follicles in the ovaries (the antral follicle count or AFC) that it is proportionally related to the remaining number of primordial follicles (75;76). There are some endocrine markers that have been shown to do the same. Anti-müllerian hormone, follicle stimulating hormone and Inhibin are endocrine markers that are released from antral follicles. Studies assessing the predictive capacity of these tests show that most tests have an adequate capacity to identify the ovarian reserve (77) and they are therefore used as representatives of the quantitative aspect of ovarian ageing. Through this, anti-müllerian hormone has also been suggested as an adequate predictor of age at menopause (78;79).
Oocyte quality has proven to be more difficult to measure. This is partly due to the absence of a clear definition of oocyte quality. One definition is the ability of an oocyte to produce an ongoing pregnancy. Unfortunately, the capacity of any ORT to predict pregnancy, both after one IVF cycle and cumulative cycles has proven to be limited (80). This is probably due to the fact that ORTS relate fully to quantity aspects of the ovarian ageing process and due to the ambiguous relationship between oocyte quantity and quality. In studies assessing the accuracy of ongoing pregnancy prediction, only small proportions of non-pregnant women were correctly identified with a high occurrence of false-positively predicted cases even when extreme cut-off values were applied (43;75;77;81;82). This makes the test unsuitable for the identification of women who should abstain from further ART treatment due to unfavourable pregnancy prospects; currently the only clinical applicability is limited to the identification of age-specific chance of pregnancy categories where trends towards higher ongoing pregnancy rates with higher ORT values are evident across age categories (this thesis- chapter 3).

Other aspects of quality assessment could include morphological evaluation of oocytes or analysis of polar bodies with fluorescence in-situ hybridization (83). Alternatively, the developmental potential of oocytes has been linked to the appearance and amount of cumulus cells around the oocyte (84) as well as to their gene expression (85). These measures, however, are invasive and can only be used post-hoc during ART treatment where oocytes are aspirated from a stimulated ovary. Furthermore in this scenario the chance of ongoing pregnancy is also dependent on embryo quality, endometrial receptivity and transfer technique (86;87). Ideally, a different assessment of ovarian quality is needed. As elective single embryo transfer becomes more applied globally there will be an even higher demand for new biomarkers of oocyte competence. To this date the most reliable marker for oocyte quality still seems to be age. However, to improve the accuracy of identifying women with a reduced ovarian reserve for their age it may be necessary to combine endocrine markers, ultrasound imaging and genetic tests. If a genetic marker is found, this may improve the accuracy of such a multivariate model even further (3). This thesis focuses on the potential of AMH as both a quantitative and qualitative marker of ovarian reserve.

Conclusions

Ovarian ageing is a multi-factorial process in which genetic factors, other physiological factors and environmental factors all play a role. The way in which these factors interact to cause both a quantitative and qualitative depletion of the ovarian reserve remains largely unknown. Elucidation of these pathways of interaction will create more understanding about the individual variation that exists in women of the same chronological age. Ovarian reserve tests such as AMH can be seen as an expression of this individual constitution. If such tests can adequately predict the ovarian reserve, in the future it may be possible to identify those women who are at risk of early infertility or those at risk of early menopause. This will open doors to primary prevention of infertility by counselling women to conceive early. However, it has been argued that quantity and quality decline do not happen in a one-to-one ratio, therefore it is also necessary that future research aims to identify markers that adequately reflect oocyte quality and not only quantitative aspects of the reproductive ageing process. This thesis focuses directly on the role of AMH in the female reproductive ageing process and value of AMH in fertility forecasting in the general population.

Aims and outline of the thesis

The studies presented in this thesis all analyse the role of AMH as a reproductive biomarker in fertility care today by examining its reflection of both quantitative and qualitative aspects of reproductive decline with increasing age. The goals of this thesis can be summarized in three overall aims.

1. To examine which factors influence age-specific AMH values in the general population.
2. To examine the role for AMH in predicting response to and pregnancy after ART.
3. To examine the role for AMH in predicting age at menopause.

Chapter 2 discusses the results of an individualised patient data meta-analysis which studied the added value of ovarian reserve tests on top of patient characteristics in the prediction of an excessive response to controlled ovarian hyperstimulation in IVF (IPD-EXPORT study). It further studies whether the predictive accuracy of ovarian reserve tests is influenced by an individual's age, BMI or duration of subfertility.

Chapter 3 discusses the results of an individualised patient data meta-analysis which studied the added value of ovarian reserve tests on top of patient characteristics in the prediction of an ongoing pregnancy after IVF treatment (IPD-PROPR study). It further studies whether the predictive accuracy of ovarian reserve tests is influenced by an individual's age, BMI or duration of subfertility.

Chapter 4 addresses which reproductive and lifestyle factors influence age-specific AMH values in a large population-based cohort of women.

Chapter 5 aimed to model age at menopause based on ovarian reserve status derived from age and AMH. By verifying the conformity of predicted and observed distributions of menopausal age it provided further evidence that declining population averages of AMH are associated with menopause.

Chapter 6 investigates the predictive accuracy of mother's age at menopause on daughter's age at menopause in long term follow-up studies. It further establishes AMH as a more accurate predictor of individual age at menopause than mother's age at menopause.

Chapter 7 compares and cross-validates two existing models that use AMH to predict individual age at menopause.

Chapter 8 assesses the value of AMH in the prediction of time to menopause in a large prospective cohort study.

Chapter 9 summarizes and discusses the results of the above studies it discusses the current and future role of AMH in clinical practice, and presents suggestions for future research

Simone L. Broer[§], (MD, PhD) and Madeleine Dólleman (MD)[§], Jeroen van Disseldorp (MD, PhD), Kimiko A. Broeze (MD), Brent C. Opmeer (PhD), Patrick M.M. Bossuyt (PhD), Martinus J.C. Eijkemans (PhD), Ben Willem Mol (MD, PhD) and Frank.J.M. Broekmans (MD, PhD) on behalf of the IPD-EXPORT* study group
§ Both authors contributed equally
* Excessive Response Prediction using Ovarian Reserve Tests

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

Published in Fertil Steril 2013 May 28

Chapter 2

Prediction of an Excessive Response in IVF from Patient Characteristics and Ovarian Reserve Tests and Comparison in Subgroups: an Individual Patient Data Meta-Analysis

Abstract

Objective: To evaluate whether ovarian reserve tests (ORTs) add prognostic value to patient characteristics, like female age in the prediction of excessive response to ovarian hyperstimulation in patients undergoing IVF, and whether their performance differs across clinical subgroups.

Design: Authors of studies reporting on basal FSH, AMH or AFC in relation to ovarian response to ovarian hyperstimulation were invited to share original data. Random intercept logistic regression models were used to estimate added value of ORTs on patient characteristics, while accounting for between study heterogeneity. ROC regression analyses were performed to study the effect of patient characteristics on ORT accuracy.

Setting: IVF clinics

Patients: 4,786 women for the main analysis with a subgroup of 1,023 women with information on all three ORTs.

Intervention: None

Main outcome measures: Excessive response prediction

Results: We included 57 studies reporting on 32 databases. Female age had an area under the ROC curve (AUC) of 0.61 for excessive response prediction. AFC and AMH significantly added prognostic value to this. A model with female age, AFC and AMH had an AUC of 0.85. The combination of AMH and AFC, without age had similar accuracy. Subgroup analysis indicated that FSH performed significantly worse in predicting excessive response in higher age groups, AFC did significantly better and AMH performed the same.

Conclusions: We demonstrate that AFC and AMH add value to female age in the prediction of excessive response and that, for AFC and FSH, the discriminatory performance is affected by female age.

Introduction

In women undergoing in vitro fertilisation (IVF), the development of a large number of oocytes occurs in one third of IVF cycles (90;91). Such an excessive response may lead to poorer quality embryos, lower chances of pregnancy, or cycle cancellation (92-98). Additionally, patients with an excessive response are at risk of developing ovarian hyperstimulation syndrome (OHSS), a potentially life threatening condition (99-101). 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, and to apply effective measures to prevent such an excessive response from occurring.

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 (102). Unfortunately, precise expressions of the predictive accuracy of these characteristics are not available. 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 (75;93;95;103-115). It is not clear, however, whether these ORTs add to predictive and readily available patient characteristics, of which female age is the most important one.

As ovarian reserve decreases with increasing age, it is conceivable that the predictive value of the ORTs is mutually dependent on female age. Alternatively, the accuracy of the AFC may be different in women with a higher BMI. Moreover, BMI could further influence the predictive accuracy by possibly reducing the biologic availability of recombinant FSH for ovarian stimulation, and thereby creating spuriously reduced ovarian responses (116). Most predictive accuracy studies, however, had a limited sample size, lacking the power to evaluate patient characteristics as modifiers of accuracy in specific subgroups and the ability to analyze the added value of the ORTs on patient characteristics.

To overcome the problem of small studies with restricted power, the current study applied an individual patient data (IPD) meta-analysis approach. By aggregating data on the level of the individual patient, more precise estimates of accuracy, evaluations of added accuracy, and identification of accuracy modifiers becomes possible while taking between study heterogeneity into account appropriately.

Material and Methods

Data acquisition

We searched the existing literature for studies on the value of FSH, AFC and AMH in predicting IVF outcome. We expanded searches from conventional systematic reviews on the subject and another IPD meta-analysis (IPD-IMPORT) on poor response prediction; searches were updated to include studies up to the end of 2009. (75;77;78;87;117).

Keywords used in the systematic Medline search included synonyms for In Vitro Fertilisation (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). Studies presenting data on ovarian response to hyperstimulation, at least one ovarian reserve test (ORT) and at least one patient characteristic were eligible for the current review. All titles and abstracts were evaluated for eligibility by two authors (MD and SB or SB and JvD). If necessary the opinion of a third author was decisive (FB).

All authors of potentially eligible primary studies were informed about this individual patient data (IPD) meta-analysis initiative 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 scrutinized on quality and consistency and, whenever possible, converted into a single format. Any issues or inconsistencies were checked with the original author. For a more detailed description of the IPD meta-analysis methodology the reader is referred to previous papers (118;119).

Within all eligible studies, a comparison was made between those studies that could and those that could not be included. Sensitivity and specificity pairs for excessive response prediction were calculated for the ORTs under study, using the thresholds for excessive response that had been set in each study. Spearman correlations were then calculated for sensitivity and specificity pairs across studies, to ascertain that the differences in sensitivity and specificity levels between included and not included studies were likely the result of different threshold levels used, thereby reducing the likelihood of bias in the final analysis.

All original studies either had approval of their local research ethics committee or were exempt from obtaining such approval due to the

nature of the study. We evaluated the quality of the included studies using the QUADAS checklist, supplemented by a number of items to evaluate the risk of bias in prognostic studies. Whenever a particular variable was missing in an individual database or in an individual case within a database, data were not imputed. Baseline characteristics were analyzed in the total IPD dataset and for each of the individual studies.

Definitions

An excessive response was defined as the retrieval of more than 15 oocytes. This cut-off was selected as the definition for excessive responsive in most primary studies varied between more than 14 and more than 16 oocytes(95;105;107;111;120-123). Furthermore, it has been shown that clinical pregnancy rates decline with the retrieval of more than 15 oocytes arguing that is thus an unfavourable condition(56). Duration of subfertility was defined as the period from cessation of oral contraceptives and/or start of unprotected intercourse until the first IVF attempt. In the included studies, patients had been stimulated according to local protocol, resulting in a wide range of daily FSH dosages. In almost all studies a starting dosage of at least 150 International Units (IU) was given. This dosage is considered the optimal daily dosage in expected normal responders; with this dose it may be assumed that all patients received adequate stimulation, creating growth of all follicles sensitive to FSH within the time frame of exposure (124)

Predictive accuracy was defined as the ability of the model to distinguish excessive responders from cases with a normal or poor response. We calculated Areas Under the Receiver–Operator Characteristic Curve (ROC-AUC) for the ORTs in the prediction of excessive response for each individual study and for the pooled studies were calculated as a summary statistic of predictive accuracy.

Statistical analysis

Analyses were done in two steps. First, the added value of ORTs on top of the patient characteristics age, BMI and duration of subfertility was assessed. As a part of this analysis, we assessed whether these results may have been influenced by differences in study characteristics or daily FSH dosage administered. Secondly, we examined whether the predictive performance depends on the patient characteristics age, BMI, and duration of subfertility.

Prediction of an excessive response using ORTs and patient characteristics

To study whether ORTs have an added value on top of patient characteristics in the prediction of an excessive response we used random intercept logistic regression models. The random intercept model takes heterogeneity into account by assuming that included studies are a random sample of a potential universe of studies, and that between-study variation in the incidence of excessive response in this universe can be described by a normal distribution on the log odds scale. These models were created to quantitatively estimate the added value that ORTs have on patient characteristics in predicting an excessive response. It provides both an estimate of the summary predictive effect as well as of the variance of the between study distribution of the incidence of excessive response. Three different sets of models were used for the prediction of excessive response. The first set of models included the patient characteristics female age, BMI, and duration of subfertility. In the second set of models, the predictive capacity of each of the individual ovarian reserve tests (FSH, AFC and AMH) was estimated. In the third set of multivariate models, the added value of combinations of ovarian reserve tests on top of patient characteristics was evaluated.

The next step was to construct receiver operating characteristic (ROC) curves to express the predictive accuracy of each combination of predictive variables in distinguishing excessive responders from the rest. With each of the random intercept logistic regression models, we calculated the probability of an excessive response. By moving the positivity threshold from 0 to 1, we could then calculate sensitivity-specificity pairs for each model. Based on these, we plotted stratified ROC curves with the ROC regression model as proposed by Janes and Pepe (125;126). 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 and AUCs of the models described above and evaluate the statistical significance of any differences. Because not all studies in this meta-analysis had included data for all three ORTs, we constructed prediction models using those databases from the total dataset that included the corresponding 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 subgroup were verified in the total study group.

Because it was not recorded whether studies adjusted FSH dosage according to results of the ovarian reserve tests and as this may have

been different between fertility physicians, correction on the level of the individual study was not considered to be enough and correction on the individual level was necessary. Therefore, we repeated the analyses as described above while adding starting daily FSH dosage as a covariate. In a similar fashion, we included study design features, as identified by the QUADAS checklist, as covariates in our models, in order to evaluate whether differences in daily FSH dosage or study design influenced the observed associations between ORT, patient characteristics and the outcome excessive response (127).

Influence of age, BMI and duration of subfertility on the accuracy of ORTs in excessive response prediction

To study whether the accuracy of ORTs in the prediction of excessive response is modified by patient age, BMI or duration of subfertility we used the ROC regression model proposed by Pepe and Janes (125;126). This model allows us to study the effects of patient or disease characteristics on the classification accuracy of tests. In this model, the ORT ROC curves are modelled as a function of the covariates age, BMI and duration of subfertility. We assumed the effect of the covariate in this meta-analysis 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, thereby correcting for any heterogeneity between studies. The areas under the corresponding ROC curves (AUC) were calculated in order to express the 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/). Random intercept logistic regression prediction models were created with the 'Lme4' library, using the Laplace approximation to the likelihood.

Results

Data acquisition

The MEDLINE search up to the end of 2009 delivered 2551 hits, of which 125 were eligible for inclusion. In 22 studies the authors were untraceable, 33 authors did not reply after repeated effort, in 12 studies the data was lost and 2 studies were not suited for the current analysis. This resulted in a total of 32 databases, used for the preparation of 57 or more manuscripts, which could be included in this IPD-study. Twenty-seven had been previously included in the IPD-IMPORT study (87). Ten additional studies were identified

from the systematic MEDLINE search. We invited these ten extra authors and asked them, as well as the previous 27 studies for permission to use their databases in the present analysis on excessive response prediction. Only four of these authors sent their data (100;101;113;122), one of them submitting two separate databases (113). In total 32 datasets could be included in the EXPORT study project database, with data from 5,251 study participants (***Supplementary Figure 1***) (91;95;98;100;101;103-107;109;112-115;121-123;128-140).

We were able to replicate the primary findings of the original study in 13 databases. In 12 cases, the study database we received contained a number of patients that differed from the publication, whereas in seven other databases there were slight inconsistencies with the baseline data as previously published. These inconsistencies were discussed with the corresponding author and

could be resolved in all cases. Through this process, the level of consistency between the individual data and the data reported in the published manuscripts was regarded sufficient for all included studies.

For the comparison of the 4 included and the 6 not included studies, we attempted to calculate sensitivity and specificity of the ORTs in the prediction of excessive response. However, of the non-included studies only one reported sensitivity and specificity values for AFC in the prediction of an excessive response (111). Therefore, Spearman correlation could not be calculated. Nonetheless, for the majority of the studies this was performed in the IMPORT study (87), a related IPD study from the same research group focused on poor response prediction. In that study it was demonstrated that there was no difference in the correlations between sensitivity and specificity for included and non-included studies on poor response. Since there was no difference in poor response prediction, it is reasonable to assume that there is also no difference for excessive response prediction. We therefore assumed that no obvious bias has occurred for the present analysis by excluding studies based on the availability of primary data. Baseline characteristics of the original studies are summarized in Figures 3A-3D of the online supplementary data.

Data from 4,786 out of the 5,251 women were suitable for the analysis of prediction of excessive response, of which 894 (19%) had an excessive response. In the other 465 women information on the oocyte yield was missing. Baseline characteristics of the total study group are summarized in ***Table 1***. The AUCs of the original studies for excessive response prediction are summarized in the ***supplementary Table 1***.

Table 1. Baseline characteristics from pooled data.

	Total population	Excessive Responders	Non-excessive responder	P value
	n=4,786	n=894	n=3,892	
	Mean(5th–95th percentile)	Mean (5th–95th percentile)	Mean (5th–95th percentile)	
Female age (years)	34.4 (26.0-42.0)	32.5 (25.0-39.9)	34.7 (26.0-42.0)	< 0.001
FSH (IU/L)	7.7 (3.8-14.0)	6.4 (3.5-10.1)	8.7 (3.9-16.0)	< 0.001
AFC (number)	12.1 (3.0-25.6)	17.1 (6.0-32.0)	11.0 (3.0-22.0)	< 0.001
AMH (ng/ml)	2.5 (0.1-7.6)	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)	4.3 (1.3-10.0)	4.3 (1.5-10.0)	4.3 (1.2-10.0)	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. AFC, Antral Follicle Count; AMH, Anti-Müllerian Hormone; FSH, Follicle Stimulating Hormone.

Statistical analyses

Prediction of an excessive response using ORTs and patient characteristics
For the model building exercises, we could use data of 1,023 women from ten datasets for excessive response analysis. This was the number of women for whom all five variables of interest were known: age, AFC, AMH, FSH and the number of oocytes retrieved after stimulation. Of the evaluated patient characteristics, age was the strongest single predictor of excessive response (OR 0.89; 95% CI: 0.85 to 0.93) as shown in ***Table 2***. BMI and duration of subfertility were not significantly predictive of excessive response (***Supplementary Table 2***).
We compared the ORTs using the random intercept logistic regression model in predicting excessive response. The ROC regression analysis showed a high accuracy for AMH (AUC 0.81: 95% CI 0.76 to0.87) 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) (***Figure 1B***).
The multivariable analyses demonstrated 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; p=<0.001). Addition of FSH to this model did not further improve predictive accuracy (AUC 0.85; p = 0.73) (***Figure 1B***). Interestingly, a single AMH or AFC test had a comparable accuracy (AUC 0.81 and 0.79, respectively). Addition of AMH to AFC and of AFC to AMH significantly improved accuracy (p = <0.001 or p=0.003, respectively). A model combining these two tests resulted in an AUC of

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

	Three test study group (N= 1,023)				TOTAL STUDY GROUP (N= 4,786)			
	OR	**95% CI**	**P - value**	**Variance-RI**	**OR**	**95% CI**	**P - value**	**Variance-RI**
Univariable models								
Age (per year)	0,89	0.85 - 0.93	<0.001	0,748	0,90	0.88 - 0.91	<0.001	0,543
FSH (per IU/L)	0,76	0.70 - 0.84	<0.001	1,23	0,83	0.80 - 0.86	<0.001	0,551
AFC (per N)	1,18	1.15 - 1.22	<0.001	0,715	1,14	1.12 - 1.16	<0.001	0,605
AMH (per 1.0 ng/ml)	1,61	1.48 - 1.76	<0.001	0,878	1,59	1.49 - 1.70	<0.001	0,680
Multivariable models								
Age and FSH								
Age (per year)	0,91	0.87 - 0.94	<0.001	0,82	0,91	0.89 - 0.93	<0.001	0,497
FSH (per IU/L)	0,79	0.72 - 0.87	<0.001		0,85	0.82 - 0.88	<0.001	
Age and AFC								
Age (per year)	0,93	0.89 - 1.98	0,003	0,769	0,95	0.92 - 0.98	0,001	0,575
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	0,596	0,92	0.89 - 0.95	<0.001	0,599
AMH (per 1.0 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. The column "Variance RI" denotes the estimated variance of the random intercept in the Random intercept logistic model. It's square root is the estimated standard deviation (SD) , and may be interpreted on the logistic scale. A one SD difference between two studies in the population of studies corresponds to an increase in the Odds on the outcome (excessive response) of exp(SD). E.g. the Age and AMH model for excessive response has variance RI = 0.321, so exp(sqrt(0.321))=1.76, is the relative increase in Odds of excessive response corresponding to a difference between two studies in intercept of one SD. OR (Odds Ratio), 95% CI (95% Confidence Interval)

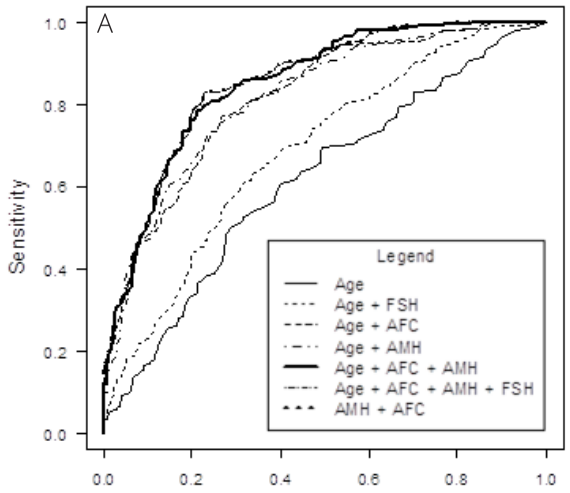
0.85. Age did not add value to this model (p = 0.98). The ROC curves corresponding to the multivariable models are shown in ***Figure 1A.***

Effect of daily FSH dosage and study protocol on excessive response outcome

Patients had been stimulated with a wide range of daily FSH dosages according to their center's local protocol. The mean daily FSH dosage was 204.28 IU (IQR=150-225 IU). Twenty-one women received daily FSH dosages below 150 IU due to an expected excessive response (5 women received 75IU, 14 women received 112.5 IU and 2 women received 125 IU of daily FSH). Women who developed an excessive response tended to have received a lower starting dosage of FSH than women who did not develop an excessive response. The mean dosage was 201.75 IU in those women who developed an excessive response versus a mean dosage of 224.79 IU for women who did not have an excessive response (p-value for difference <0.001). Daily FSH dosage had a significant, negative association with excessive response development. A higher daily FSH dosage was associated with a lower chance of an excessive response in both the three-test study group and in the group as a whole (OR 0.99: p<0.001). In the individual studies it was often not stated whether daily FSH dosage protocols were altered according to the results of the ORTs that were measured. As it is very likely that this occurred and because it is further likely that different physicians acted differently to ORT results adjusting at the level of the individual study was deemed to not be enough and correction on the individual level was necessary. When daily FSH dosage was included in the multivariable model as an additional covariate (in addition to age and the ORTs) the odds-ratios for age and the ORTs, adjusted for FSH dosage, remained basically unchanged. In the multivariable model for age, FSH and daily FSH dosage, FSH had an OR of 0.86 (95%CI 0.83-0.90), the OR for AFC in the multivariable model for age, AFC and daily FSH dosage remained 1.13 (95%CI 1.11-1.15) and the OR for AMH in the multivariable for age, AMH and FSH dosage was 1.55 (95%CI 1.45-1.66).

Study quality characteristics as scored by QUADAS checklist and supplemental questions are shown in ***supplementary Figure 2***. Overall, data were of high quality, with the exception of verification bias. This implies that the test results may have been known to the clinician taking decisions on patient management. None of the study characteristics that were assessed were associated with excessive response development

Figure 1. AUCs and ROC curves of prediction models of age and ovarian reserve tests for the prediction of an excessive response



B

	Three-test study group (N= 1,023)				Total study group (N=4,786)			
	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

1A: 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.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.
1B: AUCs of prediction models of age and ovarian reserve tests for the prediction of an excessive response 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.

(p-value range 0.34-0.89). Similarly, the odds-ratios for age and the ORTs, adjusted for study characteristics, remained basically unchanged.

Influence of age, BMI and duration of subfertility on the accuracy of ORTs in excessive response prediction

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 3***. The accuracy of FSH was significantly lower in women with a higher age (p = 0.01).

For a 20 year old the AUC for FSH was 0.66. In contrast, the AUC for a 30 year old was 0.59 and 0.52 for a 40 year old. The accuracy of AFC was significantly higher in women with a higher age (p = 0.01). For a 20 year old woman the AUC for AFC was 0.64, for a 30 year old it was 0.71 and for a 40 year old it was 0.81. The discriminatory capacity of AMH in response prediction was not significantly influenced by age. BMI and duration of subfertility categories had no significant effect on the ROC curves, for any of the ORTs.

Table 3. 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; Duration= Duration of subfertility.

Discussion

The results of the present IPD meta-analysis, with data from 32 individual studies, demonstrate that both AFC and AMH clearly add value to female age alone in the prediction of excessive response. AMH and AFC in concert have high predictive accuracy, even without adding female age. The results also indicate that the performance of the ORTs may vary across patient subgroups, as determined by female age especially. At a higher female age FSH performs less well, while AFC performs better in older age groups. As FSH performs the least well in excessive response prediction this finding is not very relevant. For AFC the change in predictive accuracy with increasing age is more notable and results in an increased predictive accuracy, in terms of an increase in the area under the curve, of approximately 0.26. However, this increase is only seen with big increments of female age (from 20 to 30 years or 30 to 40 years). With smaller increases in female age such as between 31, 34 and 37 years (the 25th, 50th and 75th percentiles of age and thus the most clinically relevant group) the increase in AUC is much smaller and less clinically relevant. In addition, the gain in predictive accuracy is evenly spread over the entirety of the curve thus limiting the margin of additive clinical value. The results of this IPD meta-analysis are mostly in line with those from a previous, conventional systematic review and meta-analysis of ovarian reserve tests and excessive response (120) and another recent study in which AMH was able to accurately identify 79% of excessive responders (141). Our IPD approach allowed us to evaluate the added value of ORTs on top of female age and, moreover, allowed for the analysis of accuracy in subgroups of women defined by to age, BMI or duration of subfertility. While ORT adds value to female age in predicting excessive response, age adds little to nothing to the accuracy of the prediction based on the ORTs. It does however does seem to influence the accuracy of some ORTs.

The results of this IPD meta-analysis also suggest that age influences the accuracy of AFC and basal FSH. Although ovarian reserve decreases with 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 AFC, the change in accuracy may be significant only from the statistical point of view, without actual implications for clinical practice, and without an obvious explanatory mechanism.

A challenge with the IPD approach is collecting sufficient data. 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 inaccurate contact information or because authors did not reply to the e-mail addresses provided. Older data were often lost or in a format that could no longer be read. Studies to investigate the possibility of combining IPD data with aggregated data are ongoing (142). To compare included and excluded studies we aimed to calculate Spearman correlation coefficients 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. However, for 27 out of 32 studies a Spearman correlation was calculated from a previous IPD meta-analysis on poor response prediction and this showed that there was no difference, (75). Since there is no difference in poor response prediction, it is reasonable to assume that there is also no difference for excessive response prediction. Therefore, we believe that the current number of participants and amount of data allowed us to analyze a valid selection of all the available data. It would have been interesting to add PCOS as a candidate predictor in our uni- and multivariate analyses as women with PCOS have been found to be prone to establishing OHSS after IVF treatment (102). However, in the majority of studies, PCOS was one of the exclusion criteria and from those studies that included and recorded PCOS, a mere 131 women had PCOS.

Although the current IPD meta-analysis included studies up to the end of 2009, the results of more recent studies on the value of ORTs in predicting ovarian response are still in agreement with our findings of this current IPD-meta-analysis. Two recent studies in an IVF setting (141;143) and three studies performed in oocyte donors or breast cancer patients undergoing oocyte cryopreservation all show an AUC of around 0.80 for AMH in excessive response prediction(144-146).

Using original data of a number of studies comes with between study heterogeneity. 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, access to IVF resources, differing treatment protocols, variability in 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

exists to correct for these differences. Consequently, no cut-off values for these tests could be used or mentioned. The most valuable method of obtaining such cut-off values for clinical practice is through randomized controlled trials which are underway at the moment(147). We have used random intercept logistic regression as well as the ROC regression model by Janes and Pepe et al. (125;126) in which pertinent heterogeneity between studies is accounted for.

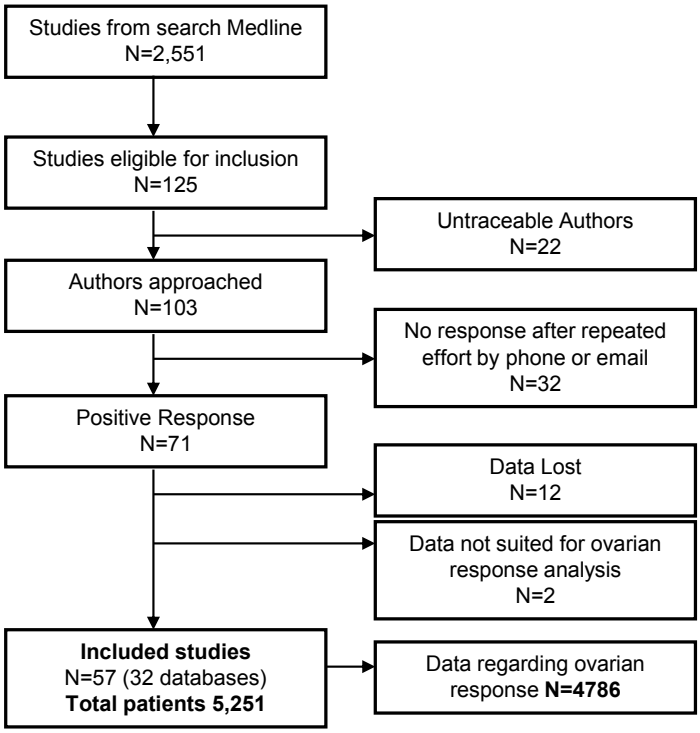
The clinical value of excessive response prediction will depend on the consequences for clinical management. Several studies have looked at the effect of individualised treatment protocols. By providing women with personally tailor-made stimulation protocols, i.e. with a lower daily FSH dosage, it is attempted to keep the oocyte yield between 5-12 oocytes. At present, the evidence is inconclusive upon the effectiveness of such personalised treatment regimens based on a priori prediction of ovarian response (135;148). In the study of Popovic-Todorovic the use of an individualised protocol resulted in a larger number of normal responders but a similar number of excessive responders (135). In contrast, Olivennes et al. demonstrated that lower individualised dosage protocols allow for a similar oocyte yield, implantation rate and pregnancy compared to higher dosage protocols (149). A third study showed no difference in the number of mature oocytes retrieved or in the occurrence of OHSS between patients that were randomly assigned to receive 225 IU or 300 IU of FSH (150). Lastly, it has been suggested that individualization of stimulation protocols dose based on ovarian reserve tests is expected to be cost effective in IVF populations (151).

Based on the current study we cannot speculate about associations between FSH dosage and excessive response prevention. A significant association between daily FSH dosage and excessive response was found, with women with lower daily FSH dosages having higher chances of excessive response. This association reflects physician behaviour, where lower daily FSH dosages are pre-emptively prescribed guided by specific patient characteristics, ORT results, or any comorbidity in anticipation of an excessive response. This suggests a form of selection bias, where the accuracy of ORTs or patient characteristics in the prediction of an excessive response is actually higher than currently reported, as some excessive responses may have been prevented by prescribing lower daily FSH dosages. The high response despite a low daily FSH dosage can be explained by the presence of a large number of follicles with a sensitivity for FSH close to the FSH threshold (152). More prospectively collected

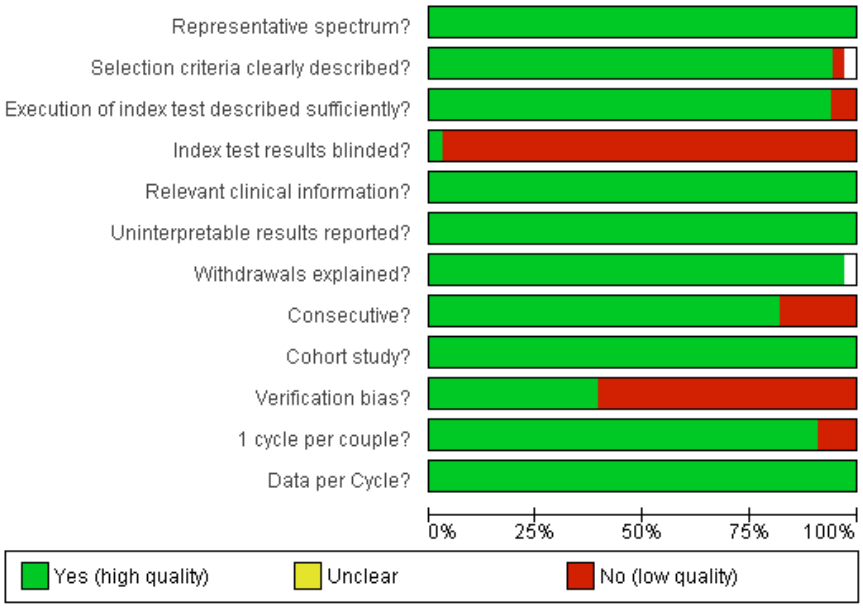
evidence, in the form of large scale randomized control trials is needed to demonstrate whether an individualised treatment protocol based on ORTs and patient characteristics is an truly effective strategy in the prevention of an excessive response, a protocol for such a randomized control trial was recently published (147).

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 good predictive accuracy, without the necessity to include female age. The performance of FSH and AFC, but not AMH, was influenced by female age but not by BMI or duration of subfertility. However, the performance across subgroups with small increments in female age seemed not to be sufficiently altered to be recognized as clinically relevant. The high predictive accuracy for both AMH and 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.

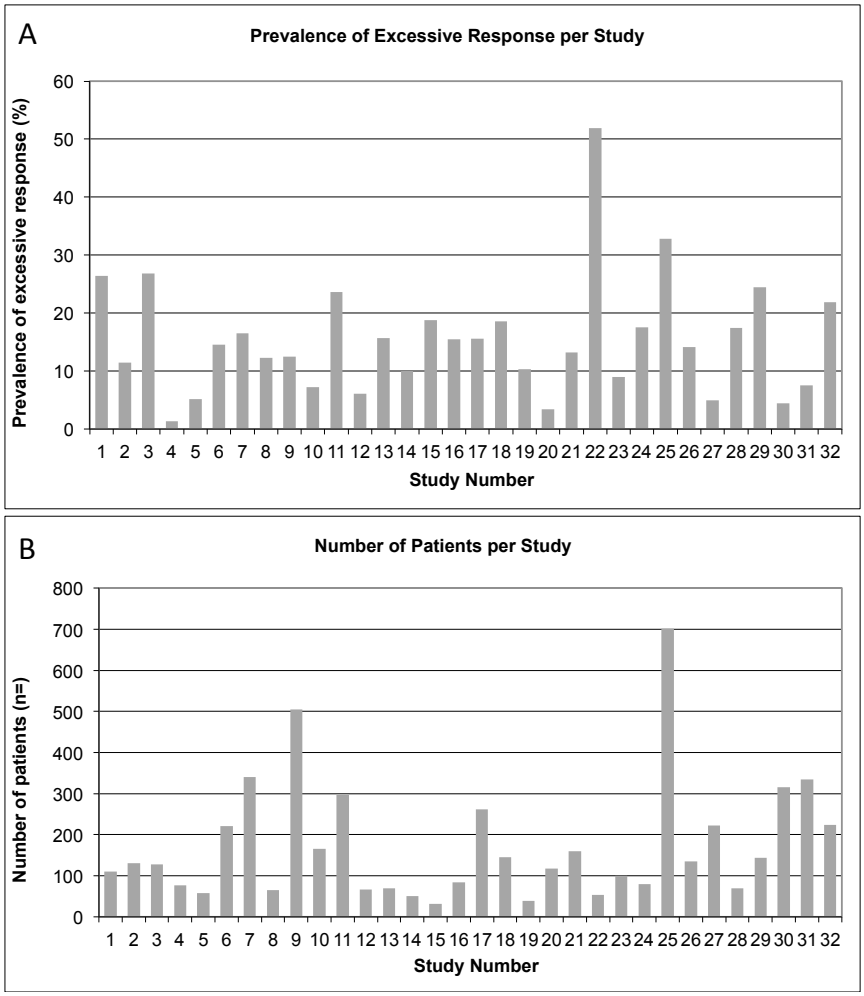
Supplementary material



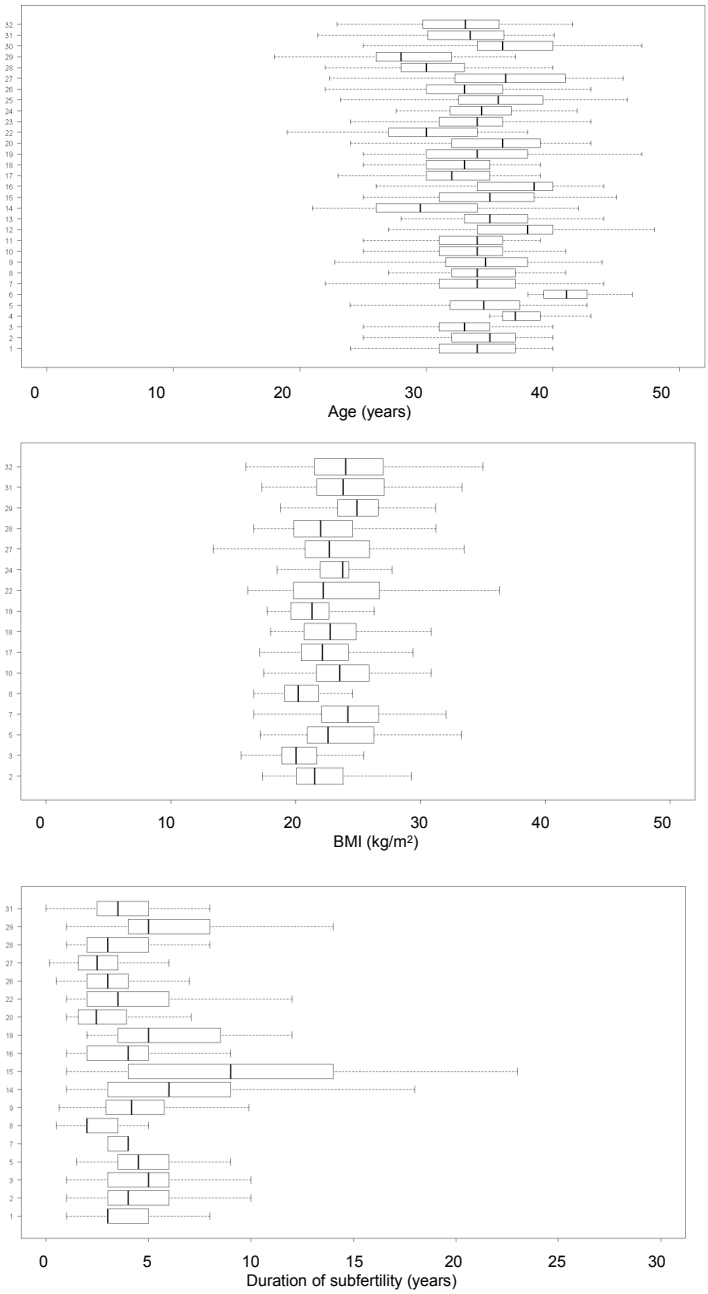
Supplementary Figure 1. Flowchart of included studies



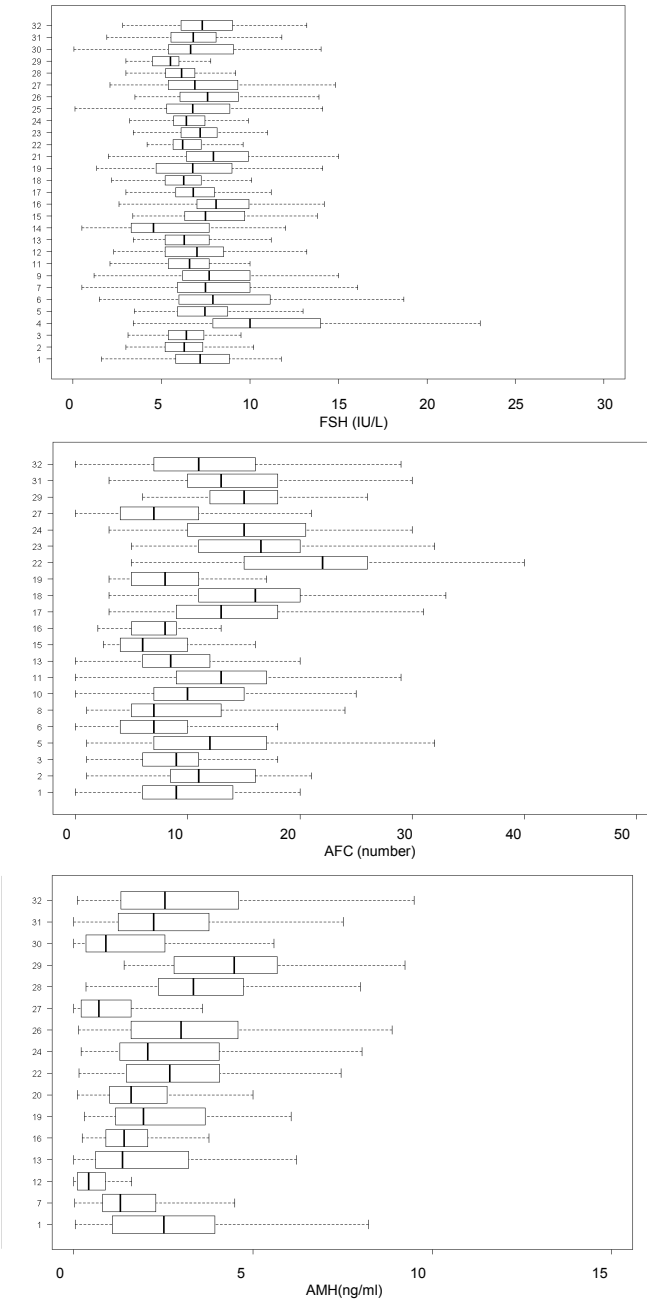
Supplementary Figure 2. Study characteristics according to QUADAS
Characteristics of all included studies evaluated with the QUADAS checklist. Note that QUADAS was set up for diagnostic studies and these are all prognostic studies. Therefore, questions regarding reference test could not be answered. Some questions specific for ovarian reserve testing and fertility studies were added. 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.



Supplementary Figure 3. Baseline characteristics of the included studies
A. The number of patients per study are demonstrated; B. The prevalence of an excessive response per study is demonstrated C. For each individual study the mean, 5th and 95th percentile of the patient characteristics female age, BMI and duration of subfertility are shown; D. For each individual study the mean, 5th and 95th percentile of ovarian reserve tests FSH, AFC and AMH are shown. 1) Kwee, 2) Ng 2000 3) Ng 2005 4) Caroppo 5) Anderson 6) Klinkert 7) Nelson 8) Merce 9) Bancsi 10) Tomàs 11) Greenblatt 12) Muttukrishna 2004 13) Muttukrishna 2005 14) Ashrafi 15) Erdem 16) McIlveen 17) Popovic 2003a 18) Popovic 2003b 19) Vladimirov 20) La Marca 21) van der Linden 22) Eldar-Geva 23) Jayaprakasan 24) Smeenk 2007 25) Copperman 26) Ebner 27) van Rooij 28) Freour 29) Aflatoonian 30) Gnoth 31) Nardo *unpublished 32) Nardo 2008



Supplementary Figure 3. Continued I
C. For each individual study the mean, 5th and 95th percentile of the patient characteristics female age, BMI and duration of subfertility are shown



Supplementary Figure 3. Continued II
D. For each individual study the mean, 5th and 95th percentile of ovarian reserve tests FSH, AFC and AMH are shown.

Supplementary Table 1. AUCs of the included studies in the prediction of an excessive response

Study	FSH			AFC			AMH		
	AUC	N	Assay	AUC	N	Criteria (mm)	AUC	N	Assay
Aflatoonian	0.60 (0.50-0.69)	143	12	0.96 (0.93-0.99)	143	2-6	0.94 (0.90-0.98)	143	DSL
Anderson	0.92 (0.99-1.00)	46	11	0.61(0.67-0.85)	46	2-10	NA		
Ashrafi	0.59 (0.31-0.87)	50	NA	NA			NA		
Bancsi	0.61(0.54-0.68)	505	6	NA			NA		
Caroppo	0.81(0.72-0.90)	76	3	NA			NA		
Copperman	0.65 (0.60-0.69)	570	5	NA			NA		
Ebner	0.61 (0.46-0.75)	127	NA	NA			0.82 (0.74-0.90)	135	BC
Eldar-Geva	0.71(0.57-0.85)	52	5	0.88 (0.75-1.00)	36	2-10	0.75 (0.62-0.88)	54	BC
Erdem	0.77 (0.57-0.97)	24	5	0.85 (0.70-1.00)	24	2-8	NA		
Freour	0.58 (0.41-0.73)	62	NA	NA			0.70 (0.55-0.86)	64	BC
Gnoth	0.64 (0.51-0.78)	122	NA	NA			0.87 (0.79-0.95)	134	DSL
Greenblatt	0.67(0.59-0.74)	261	5	0.69 (0.61-0.77)	223	2-8	NA		
Jayaprakasan	0.74(0.57-0.91)	100	NA	0.82 (0.70-0.95)	100	2-10	NA		
Klinkert	0.42 (0.30-0.55)	212	4	0.45 (0.33-0.57)	221	2-5	NA		
Kwee	0.79 (0.70-0.88)	109	1	0.87 (0.82-0.96)	109	2-10	0.84 (0.76-0.92)	105	DSL
La Marca	NA			NA			0.90 (0.76-1.00)	118	BC
McIlveen	No >15	71	8	No >15	71	2-10	No >15		BC
Merce	NA			0.62 (0.42-0.83)	65	2-5	NA		
Muttukrishna 1	0.81 (0.59-1.00)	66	7	NA			0.92 (0.83-1.00)	66	BC
Muttukrishna 2	0.67 (0.52-0.82)	68	7	0.84 (0.73-0.94)	68	NA	0.73 (0.56-0.91)	68	BC
Nardo 1	0.65 (0.53-0.77)	135	5	0.71(0.59-0.83)	123	2-5	0.74 (0.64-0.83)	135	DSL
Nardo 2	0.68 (0.59-0.77)	145	13	0.71(0.63-0.80)	145	2-5	0.79 (0.72-0.87)	145	DSL
Nelson	0.64 (0.58-0.71)	338	5	NA			0.88 (0.82-0.91)	319	DSL
Ng 1	0.70 (0.56-0.83)	131	2	0.80 (0.70-0.90)	131	NA	NA		
Ng 2	0.72 (0.56-0.83)	109	5	0.77 (0.68-0.85)	127	NA	NA		
Popovic 1	0.62 (0.54-0.71)	256	1	0.71(0.63-0.80)	256	2-5	NA		

Supplementary Table 1. Continued

Study	FSH			AFC			AMH		
	AUC	N	Assay	AUC	N	Criteria (mm)	AUC	N	Assay
Popovic 2	0.62 (0.50-0.73)	143	1	0.76 (0.67-0.86)	143	2-5	NA		
Smeenk 1	0.54 (0.40-0.68)	80	10	0.66 (0.5300.79)	80	2-10	0.71 (0.57-0.84)	80	BC
Tomas	NA			0.82 (0.72-0.91)	160	2-5	NA		
Van Rooij	0.68 (0.58-0.79)	215	10	0.86 (0.79-0.93)	215	2-5	0.87 (0.77-0.97)	215	BC
Van der Linden	0.82 (0.72-0.92)	124	NA	NA			NA		
Vladimirov 2	0.67 (0.48-0.87)	39	9	0.74 (0.52-0.97)	39	2-10	0.80 (0.67-0.93)	39	BC

FSH Assays: 1=Immunometric Delfia finland; 2= automated chemiluminescence ACS180, Bayer USA; 3= immunoradiometric, Immunotech, Marseille, france; 4= immunometric assay Chiron Diagnostics, Tarrytown, NY; 5= Immulite semi-automated DPC, Los Angeles, CA, USA; 6= Enzymun-FSH test Boehringer Mannheim, Mannheim,Germany; 7= immuno-radiometric assay (DPC, Gwynedd, UK); 8= chemiluminescence detection (Adiva Centaur; Bayer, Newbury, UK); 9= electro chemiluminescence immunoassay (Roche Elecsys, Indianapolis, USA); 10=fluorescence immunoenzymometric;AxSYM; Abbott, Hoofddorp; 11= double antibody assay (Organon N.V. Oss Holland); 12= IDCS, Korbach, Germany; 13= Roche E170 automated immunoassay.
AMH assays: DSL= Diagnostic Systems Laboratories (Webster, Texas, USA);BC=Beckman Coulter (Marseilles, France). NA=Not available.

Supplementary Table 2. Univariable and multivariable models of patient characteristics in the prediction of an excessive response

	Excessive Response Prediction					
	Three test study group (n=1,023)			Total study group (n=4,786)		
	OR	**95% CI**	**P - value**	**OR**	**95% CI**	**P - value**
Univariable models						
Age	0.89	0.85 - 0.93	< 0.001	0.90	0.88 - 0.91	< 0.001
BMI	0.98	0.93 - 1.03	0.405	1.00	0.97 - 1.03	0.954
Duration	0.98	0.90 - 1.06	0.555	0.97	0.92 - 1.01	0.156
Multivariable models						
Age and BMI						
Age	0.91	0.87 - 0.95	< 0.001	0.9	0.87 - 0.93	< 0.001
BMI	0.99	0.93 - 1.04	0.616	1.00	0.97 - 1.04	0.976
Age and duration						
Age	0.90	0.85 - 0.94	< 0.001	0.89	0.86 - 0.91	< 0.001
Duration	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.

M. Dólleman (MD)§, S.L. Broer§, ¬(MD, PhD), J. van Disseldorp (MD, PhD), K.A. Broeze (MD), B.C. Opmeer (PhD), P. Bossuyt (PhD), M.J.C. Eijkemans (PhD), B.W. Mol (MD, PhD) and F.J.M. Broekmans (MD, PhD) on behalf of the IPD-EXPORT* study group

The EXPORT* study group:
S.L. Broer and M. Dólleman, J. van Disseldorp, K.A. Broeze, B.C. Opmeer, A. , P. Bossuyt, M.J.C. Eijkemans, B.W. Mol, F.J.M. Broekmans, Aflatoonian, R.A. Anderson, M. Ashrafi, L. Bancsi, E. Caroppo, A.B. Copperman, T. Ebner, T. Eldar-Geva, M. Erdem, T. Freour, C. Gnoth, E.M. Greenblatt, K. Jayaprakasan, N. Raine-Fenning, E. Klinkert, J. Kwee, A. La Marca, C.B. Lambalk, M. McIlveen, L. Mohiyiddeen, L.T. Merce, S. Muttukrishna, L.G. Nardo, S.M. Nelson, H.Y. Ng, B. Popovic-Todorovic, J.M.J. Smeenk, C. Tomás, P.J.Q. Van der Linden, I.A. van Rooij, I.K. Vladimirov *
Excessive Response Prediction using Ovarian Reserve Tests

Manuscript in Preparation

Chapter 3

Predicting Ongoing Pregnancy in Patients With a Poor Response to Ovarian Hyperstimulation in IVF Treatment: an IPD Meta-Analysis

Abstract

Introduction: The aim of the present study was to assess the accuracy of patient characteristics and ovarian reserve tests (ORTs) in predicting ongoing pregnancy in patient with a poor response to controlled ovarian hyperstimulation (COH) during assisted reproduction (ART)

Design: Authors of studies reporting on basal follicle stimulating hormone (FSH), anti-müllerian hormone (AMH) or antral follicle count (AFC) in relation to ovarian response to and, pregnancy after COH were invited to share original data. Random intercept logistic regression models were used to estimate added value of ORTs on patient characteristics, while accounting for between study heterogeneity. ROC regression analyses were performed to study the effect of patient characteristics on ORT accuracy.

Setting: IVF clinics

Intervention: None

Main outcome measures: Ongoing pregnancy prediction

Results: Twenty-seven studies were included, with a total of 5223 patients. There were 430 poor responders (10.7%) in which the overall ongoing pregnancy rate was 11.7% (n=50). Only duration of subfertility and AMH were significantly associated with ongoing pregnancy. The multivariable models of age and AMH or age and AFC both had AUCs of 0.59 (95%CI 0.41-0.78) and 0.61 (0.42-0.80) respectively.

Conclusion: This IPD meta-analysis shows that, in women with a poor ovarian response to COH during ART, any single variable or combination of variables achieves only limited accuracy in the prediction of ongoing pregnancy for the individual.

Introduction

A poor response to ovarian hyperstimulation is a common challenge during assisted reproductive technology (ART) cycles where the small amount of oocytes available at time of ovum pick-up leads to a decreased chance of pregnancy (75;153-156) Nowadays, patients are often counselled according to either their expected or their observed response in the first stimulation cycle. However accurate translation of the ovarian response into prospects for ongoing pregnancy on the individual level remains difficult. Individual predictors for the chance of ongoing pregnancy or live birth in poor responders would allow for patients to be counselled appropriately during IVF treatment and would help physicians to decide when to abstain from further treatment in women with a null-prognosis.
Although studies have shown that ovarian reserve tests (ORTs) are adequate predictors of oocyte quantity (i.e. the number of oocytes retrieved after ovarian hyperstimulation), their ability to reflect oocyte quality (the number of ongoing pregnancies or live births after IVF) is debatable (3;75;81;112;121;132;156-158;158-161) A large individual patient database (IPD) study has demonstrated that poor responders can be identified prior to treatment using a combination of patient age, antral follicle count (AFC) and anti-müllerian hormone (AMH)(162). This combination however, was inadequate in the prediction of ongoing pregnancy in the group as a whole which also contained normal and excessive responders (87). Other studies have investigated the relationship between ORTs and ongoing pregnancy rates but the results on the value of AMH and AFC in this prediction are inconsistent (132;146;161;163-169). Studies that have analyzed pregnancy prospects in poor responders specifically comprise small numbers of patients and most of them do not consider ovarian reserve tests in the analyses. Although the concept that young poor responders have a higher pregnancy rate than older poor responders may be solid (55;153) the added value of ovarian reserve tests or other patient characteristics has not yet been adequately studied. Therefore, it would remain clinically relevant to identify factors that explain the prominent variation in ongoing pregnancy rates amongst poor responders.
The aim of the present study was to assess the accuracy of ovarian reserve tests in the prediction of an ongoing pregnancy in poor responders specifically, during their first cycle of IVF, by performing an IPD meta-analysis.

Material and Methods

Data acquisition

The data search and data acquisition has been thoroughly described elsewhere (162) (170). In brief, we searched the existing literature for studies on the value of FSH, AFC and AMH in predicting IVF outcome. All authors of potentially eligible primary studies were informed about this individual patient data (IPD) meta-analysis initiative 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 scrutinized on quality and consistency and, whenever possible, converted into a single format. Any issues or inconsistencies were checked with the original author. For a more detailed description of the IPD meta-analysis methodology the reader is referred to previous papers (118;119).

Within all eligible studies, a comparison was made between those studies that could and those that could not be included. Sensitivity and specificity pairs for excessive response prediction were calculated for the ORTs under study, using the thresholds for excessive response that had been set in each study. Spearman correlations were then calculated for sensitivity and specificity pairs across studies, to ascertain that the differences in sensitivity and specificity levels between included and not included studies were likely the result of different threshold levels used, thereby reducing the likelihood of bias in the final analysis.

All original studies either had approval of their local research ethics committee or were exempt from obtaining such approval due to the nature of the study. We evaluated the quality of the included studies using the QUADAS checklist, supplemented by a number of items to evaluate the risk of bias in prognostic studies. Whenever a particular variable was missing in an individual database or in an individual case within a database, data were not imputed. Baseline characteristics were analyzed in the total IPD dataset and for each of the individual studies.

Definitions

A poor response was defined as the retrieval of 3 oocytes or less. This cut-off was based on an ESHRE consensus that suggests that a poor response should be defined as the presence of at least two of the following criteria: 1) Advanced maternal age (≥ 40 yrs) or any other risk factor for poor

ovarian response; 2) a previous poor response of ≤ 3 oocytes after standard stimulation; 3) An abnormal ovarian reserve test (171) Duration of subfertility was defined as the time-window between the cessation of oral contraceptives or start of unprotected intercourse and the first IVF attempt. Ongoing pregnancy was defined as a gestational sac with fetal heart action at 6 weeks gestation as seen on an ultrasound. Cases were only included if the stimulation dosage of FSH was known to be 150 IU or more, thereby excluding women who had a poor response due to suboptimal dosing. This dosage is the optimal daily dosage in expected normal responders and with this dose it may be assumed that all patients received adequate stimulation, creating growth of all follicles sensitive to FSH within the time frame of exposure(124). 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. Standardized odds ratios were calculated with logistic regression correcting for patient age and individual study dataset. The area under the curve (AUC) for the prediction of ongoing pregnancy in poor responders by the various ORTs and patient characteristics were calculated for the individual studies as well as the entire database.

Statistical Analysis

Analyses were done in two steps. First, the added value of ORTs on top of the patient characteristics age, BMI and duration of subfertility was assessed. As a part of this analysis, we assessed whether these results may have been influenced by differences in study characteristics or FSH dosage administered. Secondly, we made nomograms in which the chance of achieving an ongoing pregnancy as a poor responder are calculated.

Predicting ongoing pregnancy in poor responders using ORTs and 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 'Lme4' library in R (version 2.9.0. (http://www.r-project.org/), using the Laplace approximation to the likelihood. The random intercept model takes heterogeneity into account by assuming that included studies are a random sample of a potential universe of studies and that between-study variation in the predictive effect in this universe can be described by a normal distribution on the log odds scale. These models were created to quantitatively estimate

the added value that ORTs have on patient characteristics in predicting an ongoing pregnancy amongst poor responders. It provides both an estimate of the summary predictive effect as well as of the variance of the between study distribution of the incidence of ongoing pregnancy.

Three different sets of models were used for the prediction of ongoing pregnancy in poor responders. The first set of models included the patient characteristics female age, BMI, and duration of subfertility. In the second set of models, the predictive capacity of each of the individual ovarian reserve tests (FSH, AFC and AMH) was estimated. In the third set of multivariate models, the added value of combinations of ovarian reserve tests on top of patient characteristics was evaluated.

The next step was to calculate areas under the construct receiver operating characteristic (AUC) curves to express the predictive accuracy of those variables that were significantly associated with ongoing pregnancy from the regression analyses above. The AUCs quantify how well the model distinguishes poor responders with an ongoing pregnancy from those who did not get pregnant. With each of the random intercept logistic regression models, we calculated the probability of an ongoing pregnancy response. By moving a positivity threshold from 0 to 1, we could then calculate sensitivity-specificity pairs for each model. Based on these, we plotted stratified ROC curves with the ROC regression model as proposed by Janes and Pepe (125;126). 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 AUCs of the models described above.

Because it was not recorded whether studies adjusted FSH dosage according to results of the ovarian reserve tests and as this may have been different between fertility physicians, correction on the level of the individual study was not considered to be enough and correction on the individual level was necessary. Therefore, we repeated the analyses as described above while adding starting daily FSH dosage as a covariate. In a similar fashion, we included study design features, as identified by the QUADAS checklist, as covariates in our models, in order to evaluate whether differences in daily FSH dosage or study design influenced the observed associations between ORT, patient characteristics and the outcome excessive response

To investigate whether categorical chances of ongoing pregnancy could be calculated, nomograms were constructed. Subjects were divided into five age groups and five categories according to either their AMH values or the AFC. 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

The MEDLINE search up to the end of 2009 delivered 2551 hits, of which 125 were eligible for inclusion. In 22 studies the authors were untraceable, 33 authors did not reply after repeated effort, in 12 studies the data was lost and 2 studies were not suited for the current analysis. This resulted in a total of 32 databases, used for the preparation of 57 or more manuscripts, which could be included in this IPD-study. For details the reader is referred to IPD-EXPORT (170). Of these 32 datasets, six studies did not look at the outcome measure of ongoing pregnancy or did not report on the number of oocytes retrieved and were disregarded in the current analysis (106;110;130;131;138;139). In one dataset no poor responses occurred (122). In conclusion, a total of 27 datasets with information on 5,080 study cases contributed to the current IPD-meta analysis (Figure 1).

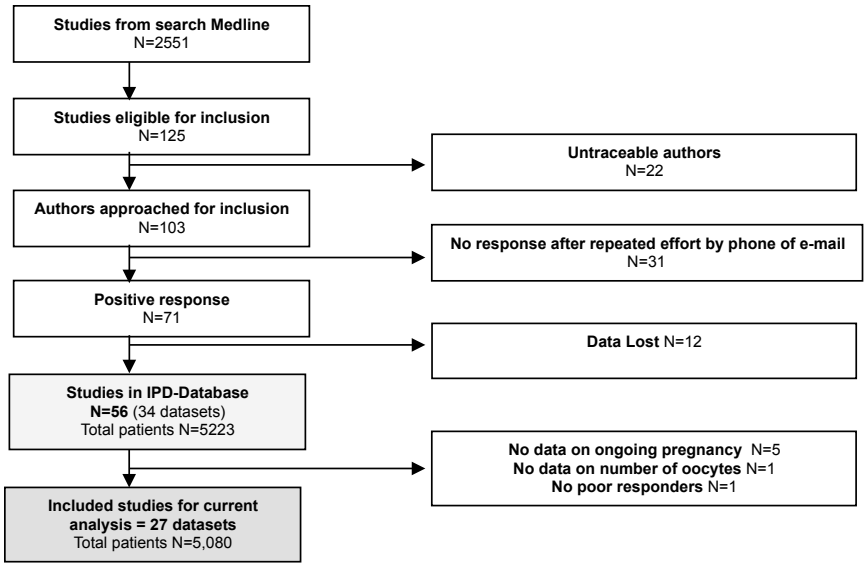


Figure 1. Flowchart of included studies:

Using a definition of poor response of retrieving 3 oocytes or less, there were 471 poor responders (9.3%) and the ongoing pregnancy rate was 11.1%. In this group, there was missing data on pregnancy in 5 women, and in 26 women the cycles were cancelled. After selecting those women who received an adequate FSH dosage of at least 150IU, and excluding those women in whom the cycle was cancelled, 407 women could be included in the analysis. The baseline characteristics for the 407 patients according to the poor response definition are summarized in table 1 along with a comparison between women with and without an ongoing pregnancy. Only AMH differed significantly between ongoing and not-ongoing pregnancies.

Table 1. Baseline characteristics of poor responders

	Poor Responders n=407			
Baseline Characteristic	**Ongoing Pregnancy**	**Not ongoing Pregnancy**	**P-value**	**n=**
N (%)	48 (11,7)	363 (88,3)		
Mean Age	35.8 (4.9)	36.8 (4.5)	0,17	409
Mean BMI	23.5 (4.1)	23.9 (4.1)	0,63	193
Mean Duration	3.4 (2.3)	4.6 (3.1)	0,05	260
Mean FSH (IU/l)	10.5 (4.6)	9..9 (4.5)	0,38	362
Mean AFC (number)	8.4 (7.0)	6.8 (4.5)	0,1	232
Mean AMH (ng/ml)	1.8 (2.9)	1.0 (1.2)	0,01	174
Mean number of oocytes retrieved	2.4 (0.8)	2.3 (0.7)	0,43	424

Baseline characteristics are described are means (SD. The P-value is measured by independent samples t-test to compare baselines between those poor responders with and without an ongoing pregnancy. Abbreviations: BMI= Body Mass index; Duration=duration of subfertility; n=Number of women per group

Statistical analyses

Influence of FSH dosage and study characteristics on pregnancy outcome
Patients were treated with a wide range of FSH dosages according to their local protocol. The mean dosage was 229.07 IU FSH (range 75-900 IU FSH). Only those women who received a FSH dosage of 150 IU or more were selected for the current analysis. In poor responders the dose was higher than in normal responders (mean dosage 253.72 IU vs. 224.88 IU respectively; p-value for difference <0.001). In the prediction of an ongoing pregnancy, FSH dosage was not significantly associated with the

study outcome. Nevertheless, all multivariable models were corrected for FSH starting dose.

Study characteristics according to the QUADAS criteria are summarized in Figure 2. Overall, data were of high quality, with the exception of verification bias. This implies that the test results may have been known to the clinician taking decisions on patient management. When assessing the effect of the study characteristics on study outcome, none of the evaluated study characteristics had a significant association with ongoing pregnancy (p-value range= 0,27-0,99).

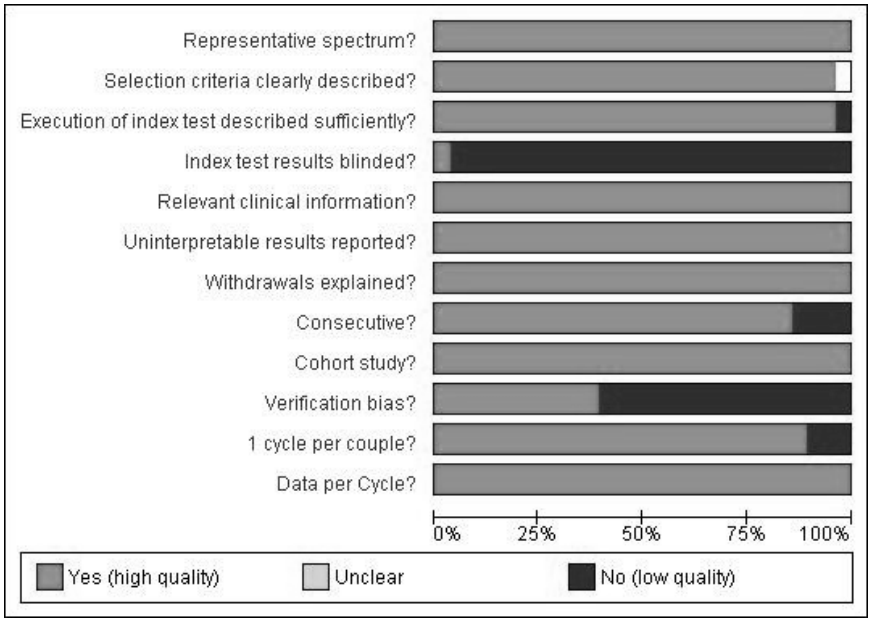


Figure 2. Study characteristics according to QUADAS
Characteristics of all included studies evaluated with the QUADAS checklist. Note that QUADAS was set up for diagnostic studies and these are all prognostic studies. Therefore all questions regarding reference test could not be answered. Some questions specific for ovarian reserve testing and fertility studies were added. 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.

Performance of patient characteristics and ORTs

Of the evaluated patient characteristics in the univariable analysis, only duration of subfertility was significantly associated with ongoing pregnancy (OR=0.82, 95%CI 0.68-0.99, p=0.04). Poor responders with

a shorter duration of infertility had a higher chance of achieving an ongoing pregnancy. Female age and BMI were not significantly associated with ongoing pregnancy. Of the ORTs, AMH was the only factor with a significant association with ongoing pregnancy (OR=1.28, 95%CI 1.04-1.59, p=0.02) suggesting a higher change of ongoing pregnancy amongst poor responders with a higher AMH. FSH and AFC did not show a significant association with ongoing pregnancy.

In the multivariate model, the ORTs were assessed after correction for heterogeneity between individual datasets, FSH starting dose, and either patient age or duration of subfertility to assess the added value of ORTs on top of these patient characteristics. In these multivariate models, AMH and AFC were significantly associated with ongoing pregnancy (OR1.28, 95%CI 1.01-1.62, p=0.04 and OR 0.92, 95%CI 0.85-1.00, p=0.05 respectively) but FSH was not. In the multivariable model where both AMH and AFC were incorporated neither of the ORTs remained significantly associated with ongoing pregnancy. Results for the univariable and multivariable regression are shown in tables 2 and 3. In this same multivariable analysis where duration of subfertility was corrected for in addition to age, none of the ORTs remained significantly associated with ongoing pregnancy. In this multivariable assessment, an association remained evident between a shorter duration of subfertility and a higher chance of achieving an ongoing pregnancy after IVF although this was no longer significant (table 3).

Table 2. Univariate analyses of patient characteristics and ORTs in the prediction of ongoing pregnancy amongst poor responders*

Univariable Models	OR	95% CI	P-value
Patient characteristics			
Age (per year)	0,96	0.90-1.02	0,18
BMI (per kg/m²)	0,97	0.88-1.08	0,62
Duration (per year)	0,82	0.68-0.99	0,04
Ovarian Reserve Tests			
FSH (per IU/L)	1,04	0.97-1.11	0,32
AFC (per N)	1,06	0.99-1.14	0,09
AMH (per ng/ml)	1,28	1.04-1.59	0,02

OR=Odds Ratio; 95% CI= 95% confidence interval; BMI=body mass index; Duration: duration of subfertility in years. N= number of women per category

Table 3. Multivariate analyses of patient characteristics and ORTs in the prediction of ongoing pregnancy amongst poor responders

Multivariable Models	OR	95% CI	P-value
Patient characteristics & Ovarian Reserve Tests			
Age+FSH			
Age (per year)	0,95	0.88-1.02	0,14
FSH (per IU/L)	1,05	0.97-1.13	0,24
Starting dose (per IU)	1,00	1.00-1.00	0,61
Age+AFC			
Age (per year)	0,98	0.88-1.08	0,65
AFC (per N)	1,09	1.00-1.18	0,05
Starting dose	1,00	1.00-1.01	0,25
Age+AMH			
Age (per year)	0,96	0.88-1.05	0,41
AMH (per ng/ml)	1,28	1.01-1.62	0,04
Starting dose (per IU)	1,00	1.00-1.01	0,44
Age+AFC+ AMH			
Age (per year)	0,98	0.85-1.12	0,73
AFC (per N)	1,04	0.92-1.18	0,53
AMH (per ng/ml)	1,12	0.76-1.64	0,56
Starting dose (per IU)	1,04	1.00-1.01	0,35
Age+Duration +FSH			
Age (per year)	0,98	0.90-1.07	0,69
Duration (per year)	0,84	0.69-1.02	0,07
FSH (per IU/L)	1,03	0.93-1.14	0,54
Starting dose (per IU)	1,00	1.00-1.00	0,70
Age+Duration +AFC			
Age (per year)	0,93	0.82-1.06	0,31
Duration (per year)	0,76	0.56-1.03	0,08
AFC (per N)	1,01	0.89-1.16	0,83
Starting dose (per IU)	1,01	1.00-1.02	0,11
Age+Duration+AMH			
Age (per year)	0,95	0.97-1.05	0,32
Duration (per year)	0,76	0.55-1.05	0,09
AMH (per ng/ml)	1,12	0.82-1.53	0,32
Starting dose (per IU)	1,00	1.00-1.01	0,48
Age+ Duration+AFC+AMH			
Age (per year)	0,92	0.79-1.07	0,27
Duration (per year)	0,59	0.34-1.02	0,06
AFC (per N)	0,95	0.78-1.16	0,64
AMH (per ng/ml)	1,07	0.57-2.01	0,83
Starting dose (per IU)	1,01	1.00-1.02	0,17

Multivariate regression analysis where all analyses are corrected for patient age, and stimulation stating dosage. OR=Odds Ratio; 95% CI= 95% confidence interval; BMI=body mass index; Duration: duration of subfertility in years. N= number of women per category

Areas under the curve were only calculated for models with a significant association with ongoing pregnancy. The AUC for age is presented purely for comparison. For age the AUC was 0.54 (95%CI 0.40-0.69), for duration of subfertility the AUC was 0.51 (95%CI 0.32-0.69) and for AMH the AUC was 0.57 (95%CI0.38-0.75). The AUC for a combination of age and AFC was 0.57 (95%CI 0.38-0.78) and for age and AMH it was 0.57 (95%CI 0.38-0.75). For all the above tests the confidence interval was wide and included the no-effect line at 0.50 indicating that no test can adequately differentiate between those poor responders that will and will not experience an ongoing pregnancy. These results are summarized in table 4.

Table 4. Area under the ROC curves for univariable and multivariable models of age and ORTs in the prediction of ongoing pregnancy

	Poor Response Prediction		
Univariable models	AUC	95%CI	n=
Age	0,54	0.39-0.69	314
Duration of subfertility	0,51	0.32-0.69	218
AMH	0,57	0.38-0.75	162
Multivariable models	AUC	95%CI	n=
Age and AFC	0,61	0.42-0.80	179
Age and AMH	0,59	0.41-0.78	162

AUC= Area under the curve; 95% CI= 95% confidence interval N=number of women per group

Categorical chances of ongoing pregnancies according to age and number of oocytes or ORTs

The nomograms which were constructed to calculate categorical chances of ongoing pregnancy according to age and AMH or age and AFC are shown in tables 5A and 5B. Within age categories a general increase in pregnancy chances is seen as the value of AMH or AFC increases. The exception to this trend is the group of women with an AMH of 1.6-2.8 ng/mL or an AFC of 10-15 in which the ongoing pregnancy rates were lower than the previous group. Similarly, a trend towards decreasing pregnancy with increasing age is seen, although much variability is present in the 36-40 year old category. This may be a chance finding due to a small number of women in these groups being present. For all three tables it is evident that the 95% confidence intervals are wide and overlapping making it impossible to make a precise individual prediction based on a woman's age and the value of the ORT.

Table 5. A) Nomogram predicting ongoing pregnancy chances in poor responders according to female age and the antral follicle count

Percentage of ongoing pregnancies across AFC categories						
		Antral Follicle Count (2-10 mm follicles)				
		<4	4-8	8-10	10-15	>15
Age (years)	<31	13,0%	16,2%	25,3%	17,3%	26,3%
	(95% CI)	4-39%	5-45%	7-61%	4-50%	5-70%
	31-35	7,9%	9,9%	15,9%	10,4%	16,6%
	(95% CI)	3-21%	4-23%	6-36%	3-28%	4-52%
	36-38	9,8%	12,3%	19,4%	12,9%	20,2%
	(95% CI)	4-23%	5-28%	7-43%	4-33%	5-57%
	38-40	13,5%	16,8%	25,8%	17,6%	26,8%
	(95% CI)	5-31%	7-37%	9-54%	5-46%	5-70%
	>40	7,7%	9,7%	15,6%	10,3%	16,3%
	(95% CI)	3-18%	4-23%	5-39%	3-31%	3-56%

B) Nomogram predicting ongoing pregnancy chances in poor responders according to female age and the anti-müllerian hormone concentration

Percentage of ongoing pregnancies across AMH categories						
		Antimullerian Hormone (ng/ml)				
		<0 ,4	0 ,4-0 ,8	0 ,8-1 ,6	1 ,6-2 ,8	>2 ,8
Age (years)	<31	17,8%	19,9%	30,3%	25,6%	40,2%
	(95% CI)	6-44%	7-44%	12-58%	7-62%	15-73%
	31-35	13,8%	15,4%	24,2%	20,2%	33,0%
	(95% CI)	5-34%	6-34%	12-43%	6-49%	11-65%
	36-38	8,1%	9 .2%	15,1%	12 .3%	21,5%
	(95% CI)	2-24%	3-24%	6-35%	3-39%	6-54%
	38-40	15 .8%	17 .7%	27 .3%	22 .9%	36 .7%
	(95% CI)	5-41%	6-41%	11-53%	6-59%	10-75%
	>40	8,0%	9,1%	14,9%	12 .2%	21 .3%
	(95% CI)	2-23%	3-27%	5-37%	3-43%	5-57%

Tables A and B show the chance of achieving an ongoing pregnancy according to female age and AFC and AMH categories respectively. The chance of pregnancy increases with younger age and higher ORT values. For these tables poor response was 3 oocytes and less. Chance predictions are depicted in percentages in the grey bars with the 95% confidence intervals in italic in the white bars underneath.

Discussion

Main findings

In this study, the individual patient database (IPD) meta-analysis approach was used to determine the value of patient characteristics and ovarian reserve tests in the prediction of ongoing pregnancy in patients with a poor response (≤ 3 oocytes) to ovarian hyperstimulation. We showed that within poor responders, no single predictor or combination of predictors can adequately discriminate between those women who will and will not have an ongoing pregnancy after IVF. Even in the best scenario, the area under the curve did not exceed 0.57. This means that in just over forty percent of cases no single test nor combination of tests can discriminate between those women who will and those who will not achieve an ongoing pregnancy. We were able to create prognostic categories using female age and AMH or AFC. From these prognostic categories is could be seen that young women with higher ORT values had more favourable pregnancy prospects. However, the differences between the prognostic categories were often small with a very wide and overlapping 95% confidence interval thereby minimizing the clinical applicability of these predictors in this specific ART subgroup.

In relation to literature

Although there is substantial evidence that ORTs are adequate predictors of a poor response to controlled ovarian hyperstimulation the literature on their value to predict oocyte quality is conflicting (77;80;112;120;121;132;157-160). By only selecting poor responders we tried to discern whether a qualitative assessment of oocytes (i.e. ongoing pregnancy) remains after the quantitative aspect (i.e. number of oocytes) is minimized by limiting the definition of a poor response to retrieving three oocytes or less.

Existing studies that assess the value of ORTs in the prediction of ongoing pregnancy, cumulative pregnancy and live birth rates in IVF populations that were not selected on poor response have conflicting conclusions (57;132;146;155;158;162;163;165-167;172). In most studies only small proportions of non-pregnant women were correctly identified using ORTs with a high occurrence of false-positively predicted cases even when extreme cut-off values were applied (43;80;156). Two studies suggest that AMH levels are significantly correlated to the number of live births and that the effect is independent of age (132;171). Other reports suggest

that AFC and AMH have a moderate accuracy in the prediction of ongoing pregnancy (164;165;168;169;172). Although we also found a significant association between both age corrected AFC and AMH with ongoing pregnancy, the accuracy of these tests as measured by the area under the curve remained insufficient. The results of this IPD-meta-analysis are similar to those described in a study by La Marca where age alone had an AUC of 0.55 AMH alone had an AUC of 0.57 and a model combining age and AMH had an accuracy of 0.66 (163). Lastly, a meta-analysis by Oudendijk et al. showed that age and number of oocytes were the best predictors of an ongoing pregnancy in poor responder patients (55). Surprisingly, within our group of poor responders, even female age had only a limited predictive capacity in identifying poor responders (AUC=0.54). The nature of the relationship was that a higher age was association with less ongoing pregnancies but this association was not significant. This is a surprising finding, especially as the nomograms do indicate a trend towards increasing ongoing pregnancy rate in younger women (with exception of women in the category of 38-40 years old). Non-linearity of age was checked with restricted cubic splines but was not significant. There may be several explanations for this. Perhaps those women with an expected poor response due to their age were treated differently than unexpected poor responders who were younger thus masking the effect of age, for example by transferring more embryos. This is something that could not be confirmed in our data due to very limited availability of information on the number of transfers. We did however find a relationship between duration of subfertility and ongoing pregnancy. The positive influence of a short duration of subfertility is well-described in literature and it is upon which prognostic models of natural conception are largely built.

Strengths and weaknesses

This IPD-meta analysis is the first to look at the combined value of patient characteristics and ORTs in predicting pregnancy prospects in women with a poor response during their first cycle of IVF.

Through the IPD method, by aggregating data on the level of the individual patient, a large cohort can be created with sufficient statistical power to generate reliable estimates of accuracy. This method also allows exploration of new determinants of accuracy as well as appropriate correction of heterogeneity and permits analysis of continuous factors such as age. Despite these advantages, IPD studies also have some drawbacks. Accumulation of original data from studies does come with

heterogeneity between different datasets. The integration of ovarian reserve test results in IVF practice has led to selection bias in some study populations and to verification bias in others. Such bias is evident in the observation that poor responders had received a higher starting stimulation dose that women without a poor response. This observation probably reflects physician behaviour, where higher starting dosages were pre-emptively prescribed guided by specific patient characteristics, ORT results, or any comorbidity in anticipation of a poor response. Further heterogeneity found in the included studies concerned differences in IVF indications or access to IVF resources and differing treatment protocols as was evidence by the variance in ongoing pregnancy rates between studies. These types of heterogeneity between datasets were been addressed by correcting for FSH starting dose and correcting for heterogeneity between datasets. Lastly, to check whether between study variability influenced the association between patient characteristics or ORTs and ongoing pregnancy, we repeated the analysis while correcting for study design features as identified by the QUADAS checklist (127). This analysis confirmed that none of the associations mentioned in this paper were influenced by study design.

A limitation that regards attention is the variation in hormone assays, particularly for AMH, and the lack of an international consensus on AFC protocols. Consequently, no cut-off values for these ORTs could be used or mentioned. This may also explain the wide, overlapping confidence intervals in the prediction of ongoing pregnancy using AMH and AFC in the nomograms. We have used the model by Janes and Pepe et al. (125;126) in which the heterogeneity between studies is corrected for. Another challenge with the IPD approach is collecting sufficient data. 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 inaccurate contact information or because authors did not reply to the e-mail addresses provided. Older data were often lost or in a format that could no longer be read. Although the current IPD meta-analysis included studies up to the end of 2009, no recent original studies have looked specifically at predicting pregnancy in poor responders.

Ongoing pregnancy was chosen as the outcome measure in this study. It has been suggested that it may be more clinically relevant to look at cumulative pregnancy rates or live birth rates after transfer of both fresh and frozen embryos, but this was not possible with the available data (173;174) as only information on the fresh embryo transfer was present.

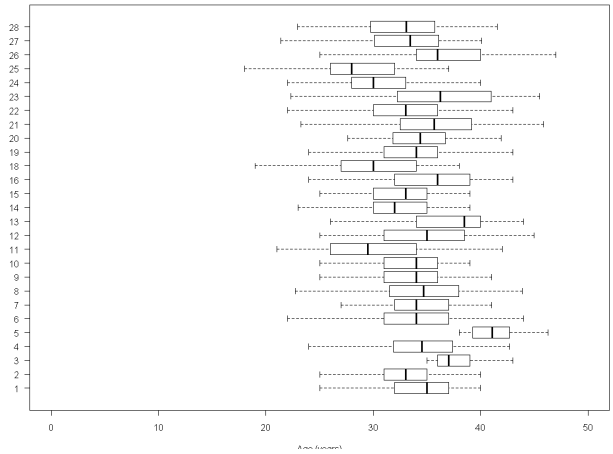
However, the poor responders in our study all had 3 or less oocytes so the number of subsequent frozen transfers would be limited.

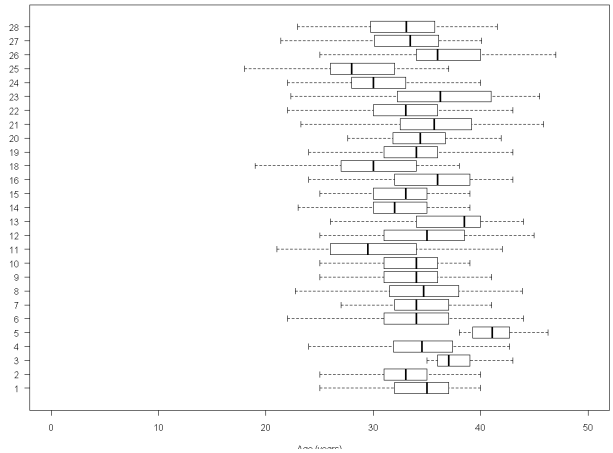
Poor responders are a challenging group of patients with a reduced chance of an ongoing pregnancy. However, a lot of variation seems to occur within poor responders. Therefore, it would be clinically relevant to discern which poor responders have a good chance of conceiving and which poor responders should be denied further treatment due to a negligible chance of an ongoing pregnancy. In clinical practice this means that those women with a favourable profile will not be judged on their poor response alone and will be allowed several IVF treatments. In contrast, in those women with poor pregnancy prospects may be told to abstain from further treatment. To improve the accuracy of identifying women with a reduced ovarian reserve for their age it may be necessary to combine endocrine markers, ultrasound imaging and genetic tests. If a genetic marker is found, this may improve the accuracy of such a multivariate model even further (3).

As elective single embryo transfer becomes more applied globally there will be an even higher demand for new biomarkers of oocyte competence. The desire for such parameters upon which one can abstain from further treatments is especially present in countries that offer fertility treatment coverage as part of the national health insurance system. It has been argued that such fertility treatment should be reserved for women with reasonable chances of achieving an ongoing pregnancy. Unfortunately, this review does not allow us to establish such a favourable profile that distinguishes between poor responders that will and will not achieve and ongoing pregnancy.

With the results of this IPD meta-analysis we can argue that in women with a poor ovarian response to ovarian hyperstimulation during ART, any single variable or combination of variables achieves only limited accuracy in the prediction of ongoing pregnancy for the individual. Identification of women who should abstain from further treatment due to a null chance of pregnancy, therefore, is not feasible. However, prognostic categories using some factors like female age and AMH or AFC, could be identified, but with modest differences between them and large confidence intervals. Clinical applicability of these predictors in this specific ART subgroup may therefore be minimal.

Supplementary Material





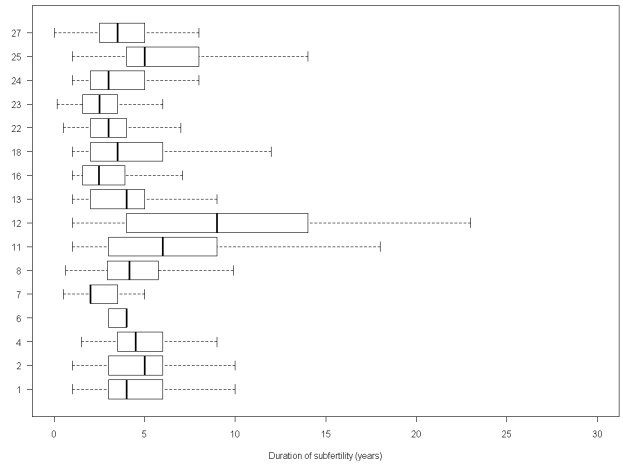
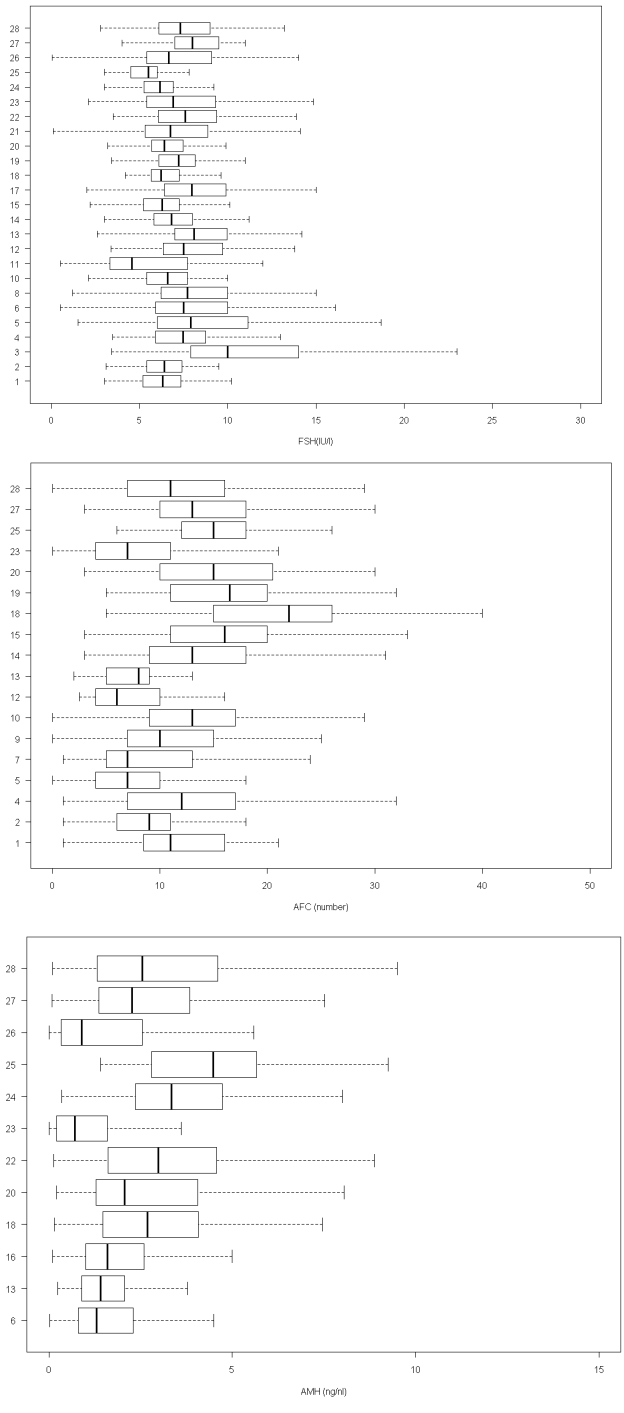
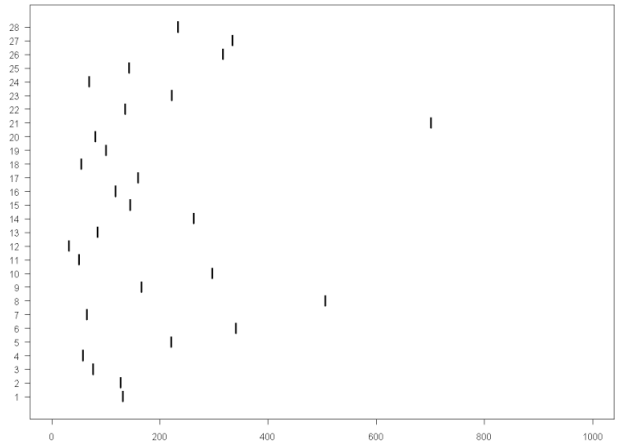
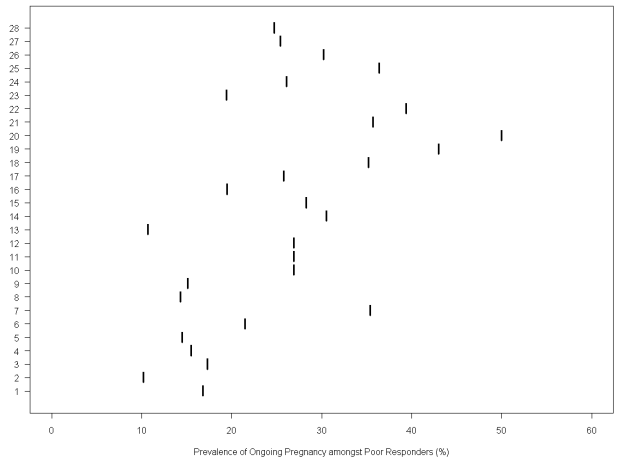


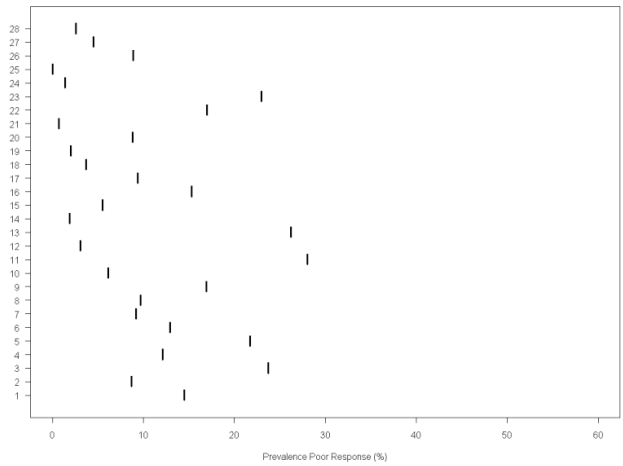
Figure A-I.
Box-and-whisker plots showing the means and interquartile ranges of the raw data per database for each patient characteristic. Study number 25 did not have any poor responders and did not contribute to the analyses in this paper. **A)** Patient Characteristics per Study: Age, BMI and Duration of Subfertility**.**



B) Ovarian Reserve Tests per Study: FSH, AFC and AMH.







C) Study Characteristics per Study : Number of patients per study, prevalence of poor response and prevalence of ongoing pregnancy amongst poor responders

Table A-II. AUCs of the included studies in the prediction of ongoing pregnancy amongst poor responders

Study	Female age (Years)	BMI	Duration of Subfertility (Years)	FSH (IU/l)	AFC (number)	AMH (ng/ml)	Oocytes (number)	Prevalen-cePoor Response % (<4 oocytes)
	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	
Aflatoonian	30,2	25,29	6,9	4,98	10,7	2,14	4	7
Anderson	35,44	24,68	5,3	9,428	8,9		3,3	21,7
Bancsi	35,798		5,169	9,179			1,535	31
Caroppo	38,9	NA	NA	11,927	NA	NA	2,23	39,5
Copperman	37,268			8,572			3,5	2,5
Ebner	34,138		6,063	10,946		1,676	2,828	34,5
Eldar-Geva	34,5	19,287	3	10,7	5	0,646	3,25	8
Erdem	37,625		10,067	8,471	5,125		6,25	50
Freour	31,5	31,86	3,125	5,7		2,348	2	6,6
Gnoth	33,667			4,86		0,443	2	9,4
Greenblatt	35,286			7,675	10,85		2,786	10,8
Jayaprakasan	37			7,929	7,571		1,857	7
Klinkert	41,399			11,563	5,198		2,418	49
La Marca	37,032		3,378			1,555	2,548	26,3
McIlveen	37,333		4,435	8,583	6,106	1,273	3,114	57,1
Merce	35,25	19,903	3,542		3,75		3,083	18,5
Nardo 1	34,307	23,377	3,806	8,333	13,269	15,179	3,167	24,2

Table A-II. Continued

Study	Female age (Years)	BMI	Duration of Subfertility (Years)	FSH (IU/l)	AFC (number)	AMH (ng/ml)	Oocytes (number)	Prevalen-cePoor Response % (<4 oocytes)
	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	
Nardo 2	35,233	25,25		8,767	8,167	1,457	3,167	8,3
Nelson	35,741	24,203	3,541	12,063		0,730	2,138	32,1
Ng 1	35,727	22,725	4,788	7,818	8,939	NA	2,636	25,2
Ng 2	33,9	21,459	5,05	6,826	6,8	NA	2,6	15,7
Popovic 1	34,421	22,186		8,1	9,368		2,938	9,3
Popovic 2	34,182	26,476		6,595	8,636		2,2	7,6
Smeenk 1	34,992	23,647		7,8	12,25	2,072	3,167	15,6
Tomas	34,889	24,137			7,306		2,889	29,3
Van Rooij	37,952	24,539	2,986	10,87	4,63	0,448	2,086	40,1
Van der Linden				11,1			3	15,7

Definition of a poor response: ≤ 3 oocytes retrieved. AUC = Area Under the Curve. FSH = Follicle Stimulating Hormone, AFC = Antral Follicle Count, AMH = Anti-Müllerian Hormone. For those AUCs that are <.50 it means that the predictive effect works in the opposite direction, namely prediction of women who do not have an ongoing pregnancy. FSH Assays: 1=Immunometric Delfia Finland; 2= automated chemiluminescence ACS180, Bayer USA; 3= immunoradiometric, Immunotech, Marseille, france; 4= immunometric assay Chiron Diagnostics, Tarrytown, NY; 5= Immulite semi-automated DPC, Los Angeles, CA, USA; 6= Enzymun-FSH test Boehringer Mannheim, Mannheim,Germany; 7= immuno-radiometric assay (DPC, Gwynedd, UK); 8= chemiluminescence detection (Adiva Centaur; Bayer, Newbury, UK); 9= electro chemiluminescence immunoassay (Roche Elecsys, Indianapolis, USA); 10=fluorescence immunoenzymometric;AxSYM; Abbott, Hoofddorp; 11= double antibody assay (Organon N.V. Oss Holland); 12= IDCS, Korbach, Germany; 13= Roche E170 automated immunoassay. AMH assays: DSL= Diagnostic Systems Laboratories (Webster, Texas, USA);BC=Beckman Coulter (Marseilles, France). NA=Not available.

M. Dólleman (1, 2), W.M.M. Verschuren (3), M.J.C. Eijkemans (2), M.E.T Dollé (3), E.H.J.M. Jansen (3), F.J.M. Broekmans (1), Y.T. van der Schouw (2)

1. Department of Reproductive Medicine and Gynaecology, University Medical Center Utrecht
2. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
3 National Institute for Public Health and the Environment, Bilthoven, The Netherlands

Published in J Clin Endocrinol Metab. 2013 May;98(5):2106-15

Chapter 4

Reproductive & Lifestyle Determinants of Anti-Müllerian Hormone in a Large Population Based Study

Abstract

Context: Anti-müllerian hormone (AMH) is an ovarian reserve marker that is increasingly applied in clinical practice as a prognostic and diagnostic tool. Despite increased use of AMH in clinical practice, large scale studies addressing the influence of possible determinants on AMH levels are scarce.

Objective: To address the role of reproductive and lifestyle determinants of AMH in a large population-based cohort of women

Design: Cross-sectional study in the Doetinchem Cohort. Using general linear modelling with CG-LMS (an established method to calculate growth curves for children) age-specific AMH percentiles were calculated.

Setting: General community.

Participants: 2,320 premenopausal women

Main Outcome Measure: The effect of female reproductive and life-style factors on shifts in age-specific AMH percentiles was studied.

Results: In comparison to women with a regular menstrual cycle, current oral contraceptive (OC) users, women with menstrual cycle irregularity and pregnant women had significantly lower age-specific AMH percentiles (for OC use 11 percentiles lower, for cycle irregularity 11 percentiles lower and for pregnancy 17 percentiles lower (p-value for all <0.0001). Age at menarche and age at first childbirth were not associated with the age-specific AMH percentile. Higher parity was associated with 2 percentiles higher age-specific AMH (p=0.02). Of the lifestyle factors investigated, current smoking was associated with 4 percentiles lower age-specific AMH percentiles (p=0.02), irrespective of the smoking dose. BMI, waist circumference, alcohol consumption, physical exercise and socioeconomic status were not significantly associated with age-specific AMH percentiles.

Conclusions: This study demonstrates that several reproductive and lifestyle factors are associated with age-specific AMH levels. The lower AMH levels associated with OC use and smoking seem reversible, as effects were confined to current use of OC or cigarettes. It is important to give careful consideration to the effect of such determinants when interpreting AMH in a clinical setting and basing patient management upon AMH.

Introduction

Anti-müllerian hormone (AMH) is a marker of female 'ovarian reserve', or the number of follicles remaining in a woman's ovaries (175). AMH is increasingly used in clinical practice for both prognostic purposes, such as prediction of ovarian response during IVF treatment, prediction of age at menopause, and diagnostic purposes such as the identification of women with premature ovarian insufficiency or polycystic ovary syndrome (75;78;117;176;177).

Several studies have presented nomograms for clinical interpretation of AMH-levels in different populations (178-183). However, there are suggestions that AMH varies according to OC use, smoking and BMI, casting doubt on the clinical utility of these nomograms (184-188).

Although it is generally believed that smoking is deleterious to the ovarian reserve and AMH values (184;189;190) , not all studies agree (179;191) and most studies have not been designed to look at dose response and the influence of quitting smoking. Thus far, studies have remained inconclusive about the effect of other potential determinants such as cycle regularity, current pregnancy, oral contraceptive (OC) use, parity, age at menarche, age at first childbirth, Body mass index (BMI), waist circumference, alcohol consumption, physical exercise or socioeconomic status. (184-187;192-195). These are factors that have been suggested to affect either female endocrinology, age at menarche or age at menopause (196-198). The aim of this study is therefore to address the possible role of both reproductive and lifestyle determinants of AMH in a large cohort of women from the general population.

Materials and Methods

Participants

The Doetinchem Cohort Study (DCS) is an ongoing multipurpose prospective study, initially carried out in a random general population sample of men and women aged 20–59 years (1987–1991). Over 95% of the DCS population are Caucasian women (199). The aim of the DCS was to study the impact of (changes in) lifestyle factors and biological risk factors on various aspects of health, such as the incidence of chronic diseases, physical and cognitive functioning, and quality of life. The cohort is re-examined every five years with questionnaires and a physical examination at the local health service. Three follow-up examination rounds were completed during 1993–1997, 1998–2002 and 2003–2007. All participants

gave written informed consent, and the study was approved according to the guidelines of the Helsinki Declaration by the Medical Ethics Committee of the Netherlands Organization of Applied Scientific Research. Details on the DCS have been extensively described elsewhere (200).
For the present study we used a cross-sectional design among women participating in the second examination round of the DCS. Women were excluded from the current study if they were post-menopausal (n=1700), taking hormone replacement therapy (n=23), if they had a history of surgery to the uterus or ovaries (n=26) or if the information on their reproductive status was missing (n=59). After exclusion, 2,320 from the original 4,128 women remained eligible for the current study.

Potential AMH determinants

The reproductive determinants included were age at menarche, OC use, cycle regularity status, pregnancy, parity, and age at first childbirth. The lifestyle determinants encompassed body mass index (BMI), waist-circumference, smoking, alcohol consumption, physical exercise and socioeconomic status. Reproductive history was assessed via extensive questionnaires. Women were asked at what age they had their first menstrual period, if their period was regular, how long their menstrual cycle was in days, and how many menstruations they had in the last 12 months. Furthermore women were asked whether they were currently pregnant, if they had ever been pregnant, if so, how many pregnancies and how many children they had as well as when the children were born. Additionally, women were asked if they had ever used OC or hormone replacement therapy (HRT), how old they were when they started using this medication, if they were using it at the time of blood sampling for AMH measurements and how many years they had used it in total. It was also inventoried whether and at what age women had undergone a hysterectomy and/or if they had ever been operated on one or both ovaries and at what age that occurred. A regular cycle was defined as having a regular cycle with a mean cycle length of 24-36 days. OC use was categorized as previous OC-use, current OC-use and never used OC. The number of years of OC use was divided into <1 years, 1-5, 5-10, 10-15, 15-20 and >20 years. Reproductive status at time of blood sampling was categorized into mutually exclusive categories of regularly cycling, irregularly cycling, currently pregnant or currently taking OC. Parity was assessed as a continuous variable. Body weight and height were measured by trained staff. Body weight was measured to the nearest 100 g on

calibrated scales with participants wearing light indoor clothing without shoes, with emptied pockets (200). BMI was categorized as normal (<25 kg/m²), overweight (25-30 kg/m²) and obese (≥30 kg/m²). Waist circumference was measured to the nearest 0.5 cm according to WHO instructions and categorized into normal (<80 cm), moderately increased (80-88 cm) and abdominal obesity (>88 cm) (201). Ever smokers were identified based on the question 'did you ever smoke regularly'. For ever smokers, information on age at which the respondent started smoking, as well as the total number of years of smoking and average number of cigarettes smoked was assessed, followed by a question on current smoking ('do you smoke at present'). Smoking was categorized as current smoker, ex-smoker and never-smoker. The amount of smoking was categorized into 5 levels: <1 cigarette per month, <1 per day, 1-9 per day, 10-19 per day and >20 per day. Packyears of smoking was divided into 7 strata with 5 year-intervals. Alcohol intake was assessed in the questionnaire as number of glasses consumed per week, and subsequently operationalized as number of beverages consumed daily. Physical activity was assessed using the validated (202) EPIC questionnaire and categorized according to the Cambridge Physical Activity Index (CPAI) into (moderately) active and (moderately) inactive (203). Socioeconomic status (SES) was classified into four categories according to the highest level of education that a woman had reached: primary school (SES level 1), lower secondary or vocational school (SES level 2), intermediate vocational or higher secondary school (SES level 3) and higher vocational or university (SES level 4)(200).

AMH assay

Blood samples were collected on a random day of the menstrual or OC cycle. Serum was frozen on the day of vena cubiti puncture and stored in liquid nitrogen. Prior to the AMH measurement each sample went through one thaw-freeze cycle on ice for 4 hours; intermitted storage until AMH measurement was at -80° for a maximum of 4 weeks. Serum AMH was measured with the AMH Gen-II ELISA (Beckman-Coulter, Sinsheim, Germany) in a single laboratory, by the same experienced lab technician. The precision of assay results was validated with linearity-of-dilution assessment. The limit of detection for this assay is 0.08 ng/ml, and the limit of quantification is 0.16 ng/ml. The inter-assay and intra-assay coefficients of variation were 3.35% and 4.0% respectively.

Statistical analysis
Baseline characteristics were described for the group as a whole with mean values and corresponding standard deviation for continuous variables and frequencies and percentages for categorical data.

Age-specific reference values of AMH
We age-standardized the distribution of AMH using the CG-LMS method. This CG-LMS method, originally described by Cole and Green, is an established method to estimate growth curves for children. The LMS method summarizes the distribution of age-specific AMH by three smoothing splines, which represent the skewness (L), the median (M) and the coefficient of variation (S). These three splines are simultaneously fitted by non-linear regression and the extent of smoothing can be modified by altering the degrees of freedom (204). The fitted model was used to predict the age-specific percentile of AMH of each woman in the database. Data was analysed with SPSS version 20.0 (Inc., Chicago, Il, USA) and with R version 2.13 (http://www.r-project.org/). The "gamlss" package, which is a collection of functions to fit Generalized Additive Models for Location Scale and Shape (GAMLSS), was used to perform the CG-LMS method in R (www.gamlss.org) (205).

Determinants of AMH
The relationships between potential determinants and age-specific AMH percentiles were assessed in the cohort using univariable and multivariable regression analysis. Determinants were investigated for both linear and non-linear associations with AMH by categorizing determinants into the clinically relevant categories as described above. The magnitude of the effect of the determinants was calculated as the number of percentiles of change. As percentiles of age- standardized AMH were used, no further correction for age in a multivariable analysis was needed. The change in percentiles is displayed with the standard error and the corresponding p-value. Individual p-values for categorical values in comparison to the reference value were calculated and the overall p-value was measured with analysis of variance between groups (p-ANOVA). For reproductive status at time of blood sampling, regularly cycling women were chosen as the reference category for the 3 other mutually exclusive categories of women. A p-value of ≤0.05 was considered to be statistically significant. The effect can thus be interpreted for dichotomous variables as *"in the presence of determinant x, age-specific AMH is y percentiles lower/higher" and for continuous variables as "with a one unit higher level of determinant x, the AMH level is y percentiles higher/lower".*

Results

The study cohort comprised of 2,320 women, as depicted in the flowchart in ***Figure 1***, of whom 1,090 were regularly cycling, 908 were taking OC for family planning purposes, 49 were pregnant and 273 had an irregular unpredictable cycle. The baseline characteristics of the cohort are shown in ***Table 1***. Age specific percentiles of AMH as estimated using the CG-LMS model is depicted in ***supplemental Figure 1***.

Table 1. Baseline characteristics of the cohort

	Mean (SD) or % (n)
Lifestyle factors	
Age (years)	37.3 (9.2)
BMI	24.3 (3.9)
Waist circumference	84.6 (10.6)
Smoking	
Current Smokers (%)	32.8% (760)
Ex- Smokers (%)	29.2% (678)
Never Smoked (%)	38% (882)
Number of cigarettes p/day amongst smokers	12.6 (7.5)
Number of pack years amongst smokers	10.2 (9.1)
Alcohol	
Drink alcohol daily (%)	46% (1068)
Amount of alcoholic beverages p/day amongst drinkers	1.1 (0.9)
Physical Activity	
CPAI 1 (inactive)	2.8% (62)
CPAI 2 (moderately inactive)	22.4% (464)
CPAI 3 (moderately active)	31.8% (622)
CPAI 4 (active)	43% (872)
Socioeconomic status	
SES 1	5.2% (120)
SES 2	47.4% (1099)
SES 3	31.7% (736)
SES 4	15.3% (356)
Reproductive Factors	
Age at menarche	13.2 (1.4)
Ever taken oral contraceptives	90.6% (2102)
Parity	2.1 (0.8)
Age at birth of first child	25.4 (4.0)
Reproductive status at AMH measurement	
Regularly cycle	47% (1090)
Irregular cycle	11.8% (273)
Oral contraceptives	39.1% (908)
Pregnant	2.1% (49)

Baseline characteristics. Means and standard deviations (SD) are shown for continuous variables. Percentage and numbers are shown for categorical variables.

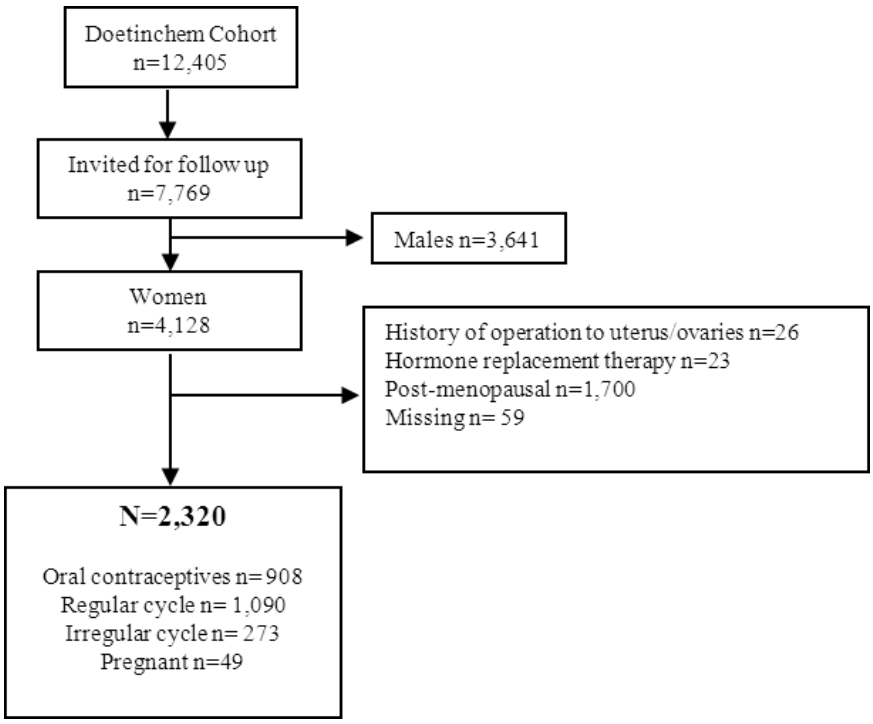


Figure 1. Flowchart of included women

Reproductive determinants

There was no association between age at menarche and AMH (beta= 0.7, p=0.09). In comparison to women with a regular menstrual cycle, current OC use, current pregnancy and having an irregular menstrual cycle were all reproductive factors that were associated with having significantly lower age-specific AMH (***Table 2***). Current OC-users had an age-specific AMH of 11 percentiles lower than women with a regular cycle (beta=11; p <0.0001). When the history of OC-use was assessed in the entire cohort, current OC users had significantly lower AMH values than women who had never used OC (beta=-9; p-value<0.0001) but previous OC use was not associated with lower AMH values (beta=-0.3; p-value=0.9). There was no clear dose-response relationship of duration of OC use with age-specific AMH levels; only women who used OC >20 years had a 8.8 ± 4.2 percentiles lower age-specific AMH compared to women who used OC for a year or less. Being pregnant was associated with age-specific AMH values of almost 17 percentiles lower (beta=16.6; p<0.0001) and having

an irregular cycle with age-specific AMH values of 11 percentiles lower (beta=11, p<0.0001). A significant trend was seen between parity and AMH. Having given birth to a higher number of children was associated with higher age-specific AMH levels (per additional child, beta=2.1,p=0.02). There was no association between the age at first child birth and AMH (beta=-0.2, p=0.39). Results are displayed in ***Table 2***.

Table 2. Number of percentiles that AMH shifts in the presence of a specific reproductive determinant

Determinants	**Linear Regression**		
	Percentiles of Change	**Std error**	**P-value**
Current Reproductive Status			
Regular Cycle	Reference		
Irregular Cycle	-11,0	2,2	<0.0001
Current OC use	-11,0	1,3	<0.0001
Current Pregnancy	-16,6	4,1	<0.0001
History of OC use			
Never	Reference		
Previous	-0,3	2,3	0,88
Current	-9,0	2,3	<0.0001
Years of OC use in current or past users (n=1996)			
< 1 year	Reference		
1-5 years	2,6	3,1	0,39
5-10 years	-2,2	3	0,46
10-15 years	-2,6	3,2	0,42
15-20 years	-3,9	3,5	0,26
>20 years	-8,8	4,2	0,04
Age at menarche	0,7	0,4	0,09
Parity	2,1	0,9	0,02
Age at birth of first child	-0,2	0,2	0,39

The association between reproductive determinants and age-specific AMH is quantified as the number of percentiles that AMH shifts in the presence of a specific determinant. Corresponding standard error (std error) and p-value are shown

Lifestyle determinants

There was no significant association between BMI (beta=-0.2; p=0.16), waist circumference (beta=-0.09; p=0.12) , alcohol consumption (beta=-0.4; p=0.74), physical activity (p-ANOVA=0.62), or socioeconomic status (p-ANOVA=0.62), and age-specific AMH. When determinants were modelled in clinically relevant categories instead of linear variables, there was still no significant association with age-specific AMH (see ***Table 3***). Smoking, however, was associated with lower AMH percentiles: Current smoking was associated with 3.6 percentiles lower age-specific AMH values (p=0.02) in comparison to never-smokers. Ex-smoking was not associated with lower AMH values (beta=-0.6. p=0.67). Daily smoking was associated with significant decreases in age-specific AMH. However, the number of cigarettes smoked daily and AMH does not seem to be related to AMH, as the number of percentiles difference of age-specific AMH with non-smokers was similar for women who smoked 1-9 cigarettes per day (beta= -12.8; p=0.05) as for women who smoked 10-19 cigarettes daily (beta= -12.3; p=0.05) or more than 20 cigarettes daily (beta=-11.8; p=0.06 respectively). Although the linear association between the number of pack years smoked and AMH was statistically significant (beta =- 0.3; p=0.002), there was evidence for a threshold after which the number of pack years becomes significant. Smoking 5-10 pack years was not associated with significantly lower age-specific AMH values than smoking less than five pack years (beta=-3.1,p=0.15), but smoking more than 10 pack years was associated with significantly lower age-specific AMH values. The magnitude of the association after 10-15 pack years (beta=-7.0, p=0.003) was similar to 15-20 pack years (beta=-8.5, p=0.001) or 20-25 pack years (beta-6.6, p=0.05). Results are displayed in ***Table 3***.

Multivariate analysis

Table 4 shows that the variables that were significantly associated with age-specific AMH levels in the univariable analysis, also remained significantly associated after multivariable adjustment. The magnitude of the effects of current smoking, current OC-use, pregnancy, cycle irregularity and parity remained largely the same in the multivariable models as in the univariable analysis. Only parity was no longer significant in the multivariate analysis after correcting for OC-use, cycle irregularity, current pregnancy, and current smoking.

Table 3. Number of percentiles that AMH shifts in the presence of a specific lifestyle determinant

Determinants	Linear Regression Percentiles of Change	Std error	P-value
Anthropometrics			
Body Mass Index	-0,2	0,2	0,16
BMI			
<25 kg/m² (Normal)	Reference		
25-30 kg/m² (Overweight)	-2,5	1,4	0,08
> 30 kg/m² (Obese)	-1,6	2,4	0,52
Waist circumference	-0,09	0,06	0,12
Waist cirumference			
<80 cm (normal)	Reference		
80-88cm (increased)	-1,9	1,49	0,19
> 88 cm (abdominal obesity)	-2,3	1,5	0,12
Smoking			
Never Smoked	Reference		
Previous smoker	-0,6	1,5	0,67
Current smoker	-3,6	1,5	0,02
Number of cigarettes smoked per day in current smokers	-0,3	0,1	0,06
Smoking frequency in currrent smokers			
<1 per month	Reference		
<1 per day	1,3	7,2	0,86
1-9 per day	-12,8	6,5	0,05
10-19 per day	-12,3	6,2	0,05
≥20 per day	-11,8	6,3	0,06
Number of packyears in current/ previous smokers	-0,3	0,09	0,001

Table 3. Continued

Determinants	Linear Regression Percentiles of Change	Std error	P-value
Packyears in categories			
0-5	Reference		
5-10	-3,1	2,1	0,15
10-15	-7	2,3	0,003
15-20	-8,5	2,6	0,001
20-25	-6,6	3,3	0,05
≥ 25	-6	2,9	0,04
Alcohol			
Daily alcohol intake (yes/no)	-0,4	1,3	0,74
Number of alcoholic beverages daily	1,6	1	0,12
Physical Excercise			
CPAI 1 (inactive)	Reference		
CPAI 2 (moderately inactive)	-5,1	4,2	0,23
CPAI 3 (moderately active)	-2,8	4,2	0,50
CPAI 4 (active)	-4,4	4,1	0,29
Social economic status			
SES 1 (lowest education level)	Reference		
SES 2	3,2	3,0	0,28
SES 3	4,0	3,0	0,19
SES 4 (highest education level)	3,4	3,2	0,29

The association between lifestyle determinants and age-specific AMH is quantified as the number of percentiles that AMH shifts in the presence of a specific determinant. Corresponding standard error (std error) and p-value are shown.

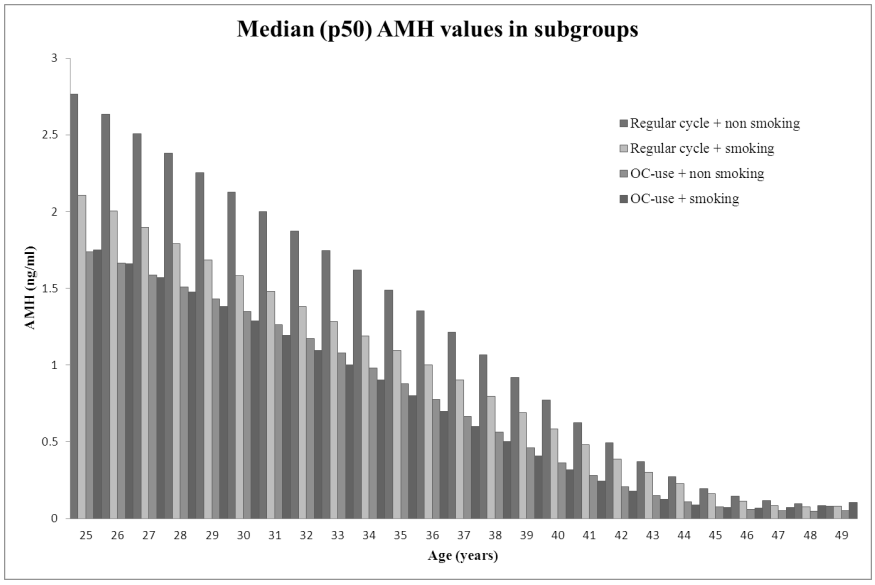
Figure 2 shows the median AMH values according to the CG-LMS model in the four most clinically relevant subgroups: Regularly cycling women and OC-taking women who did and did not smoke at the time of blood sampling for AMH measurement. In this table it is evident that the difference in AMH values between OC-taking women who did and did not smoke at the time of AMH measurement was smaller than for regularly cycling women. A test for interaction between smoking and OC-use was done to see if this would explain this discrepancy, but no significant interaction effect was found (p=0.07).

Table 4. Multivariable analysis of determinants

	Linear Regression		
	Percentiles of Change	**Standard Error**	**P Value**
Determinants			
Irregular cycle vs regular cycle	-10,8	2,4	<0.0001
Current OC-use vs regular cycle	-12,7	1,7	
Current pregnancy vs regular cycle	-17,6	5,1	
Current smoking vs nonsmoking	-4,5	1,6	<0.001
Parity (per child)	1,4	0,9	0,14

Multivariable analysis of the association between reproductive and lifestyle determinants and age-specific AMH. The association is quantified as the number of percentiles AMH shifts with the corresponding standard error (std error) and p-value.

Figure 2. Median (p50) AMH values in subgroups



Bar graph showing mean (p50) AMH values with increasing age according to the CG-LMS model for women in four subgroups.

Discussion:

Main findings

This large cohort study demonstrates that age-specific AMH levels are influenced by both reproductive factors and lifestyle determinants. More specifically, we found two distinct factors strongly associated with

age-specific AMH levels: current use of oral contraceptives and current smoking. This knowledge will ameliorate our understanding of AMH dynamics and its usefulness as a screening or diagnosis tool in clinical practice.

Findings in view of existing literature

OC use at the time of blood sampling was associated with significantly lower age-specific AMH levels. AMH was eleven percentiles lower in current OC users compared to women with a regular cycle and 9 percentiles lower compared to women who had never taken OC. One would expect the effect of current OC use to be a weighted average between previous and current OC use (-0.3 and -9.0 percentiles). However, due to the occurrence of *current* pregnancy in the group of women with *previous* OC-use the effect of current OC use is slightly lower than expected. This effect of OC use was seen across all ages and for all percentiles of AMH. The finding is not entirely consistent with literature. Several studies have suggested AMH to remain constant under the influence of exogenous sex steroids used for contraception (186;192-195). However, these studies consisted of small numbers, and none assessed age-specific AMH levels. Other studies have revealed a significant decrease in AMH during exposure to OC (184;185;187;206). Lastly, one prospective longitudinal study observed an increase in AMH values of approximately 23% after discontinuation of OC which supports our notion of the reversibility of the effect of OC (188).

The observed reduction in AMH amongst OC-users is unlikely to be due to depletion of the primordial follicle pool as OC-users do not experience natural menopause at an earlier age (207;208). Furthermore, in women who were previous OC-users, AMH levels were similar to non-users, suggesting that the effect is reversible. The decrease in AMH amongst current OC-users is therefore more likely to be due to down-regulation of the hypothalamic-pituitary-ovarian axis by OC resulting in diminished FSH and LH production and consequent altered development of the antral follicle cohort(185;209). It has also been argued that AMH must be partially gonadotropin responsive as FSH can directly stimulate AMH production from granulosa cells in vitro (210;211). However, direct stimulation by endogenous or exogenous FSH rises has not elicited any response in serum AMH levels (105). Therefore, most likely, we may theorize that OC-induced arrest of follicular growth and decreased support of growing follicles by FSH causes AMH levels to be diminished. This notion is supported by

several studies that observe a decrease in follicle size and ovarian volume (185;193;194;212) and other studies that observe a decrease in antral follicle numbers (187;194). Furthermore, FSH is an important survival factor for the AMH-producing pre-antral and early antral follicles (21). As these are the stages of follicle development during which most follicles undergo atresia (213), OC-induced suppression of FSH may also result in either increased apoptosis or decreased follicle functionality. This combined effect of smaller numbers of follicles and diminished follicle functionality from decreased FSH support is a very plausible explanation for the lower AMH values observed in this study. All the above arguments may also apply to the association between pregnancy and reduced AMH observed in this study and a previous longitudinal study as pregnancy is a transient condition with severely suppressed endogenous FSH(214). However, during pregnancy the GnRH suppression is induced by endogenously increased levels of estrogens instead of by exogenously administered drugs (57). Further research is needed to assess whether AMH levels measured under OC use are correlated with natural cycle AMH measures. By doing so it can be determined whether AMH under OC use is an adequate reflection of individual ovarian reserve and not just the individual degree of suppression by OC. Interestingly, van den Berg et al. showed AMH levels on day 7 of the pill-free interval to be highly correlated with AMH levels after discontinuation of the pill. Similarly, Nelson et al. showed pregnant AMH measures to be highly correlated with non-pregnant measures (214), indicating a proportional relationship. This study showed that cycle irregularity was associated with significantly lower age-specific AMH levels than AMH levels in women with a regular cycle. This finding is difficult to interpret, as this group encompasses women with an oligomenorrhea associated with PCOS (high age-specific AMH levels), women with irregular menstrual cycles associated with the menopausal transition (low or undetectable AMH levels), women with post-partum cycle arrest or unexplained cycle irregularity. In our study the mean age of women with an irregular cycle was 44.6 years (+/- 9 years) arguing that in a proportion of these women cycle irregularity will be explained by the occurrence of the menopausal transition. This also explains the discrepancy with a recent paper that showed women with cycle irregularity to have higher AMH values: in this population the mean age was 26.6 years and therefore high AMH values and cycle irregularity will more likely pertain to the occurrence of PCOS(215). Our results also showed that women who have given birth to more children are likely to have

significantly higher AMH levels. Recent prospective studies have shown that having a high AMH is associated with a higher age at menopause (78;79;216). As a timed interval of approximately 10 years is suggested to be present between the final menstrual period and the end of natural fertility, it is conceivable that the time window to conceive is longer in women with high age-specific AMH levels, thus potentially resulting in more child births (3). One small study of 294 women aged 21-22 years paradoxically showed that a lower AMH was associated with a higher parity and a higher age at menarche. However, in this study contraceptive use was not taken into account hindering direct comparison to our study. It is conceivable that females with a higher parity may be more likely to take contraceptive measures which may explain lower AMH values. Furthermore the sample of women who conceived 2 or more children was very small (only 9 women had 3 or more children) which may not be a representative sample to draw such conclusions from. With regard to age at menarche, their finding is paradoxical with most literature as one would expect females with a high AMH to experience menarche later due to the occurrence of PCOS. Without correcting for contraceptive use, however, inferences based on these results may be risky (217).

Of the lifestyle factors investigated, current smoking was most strongly inversely associated with AMH concentration. This is in line with existing research (184;189;190). Although the effect of smoking was dose-dependent in the sense that one had to smoke at least daily, the effect was similar in women who smoked 1-10 cigarettes daily as women who smoked more than 20 cigarettes daily. There was a similar association with the number of pack years smoked, where a significant association was seen after 10-15 years, but after 15 pack years the magnitude of the effect did not change very much for women who smoked more than 20 pack years. Previous smoking was not significantly associated with decreased AMH in both our study as well as a previous paper (189). Translated to a clinical setting, our results thus suggest that if a low age-specific AMH level is measured in a woman of reproductive age, advising her to stop smoking may be beneficial for her AMH levels. Our results suggest that the association between AMH and smoking is similar to the association between age at natural menopause and smoking as both are dose responsive and both effects diminish after quitting smoking (72;197;218). The reversibility of the effect of smoking on AMH and age at natural menopause is contradictory to the theory that the toxic effect of smoking on ovarian follicles results in earlier depletion of the primordial

follicle pool, as such an effect would be irreversible (197). Several other explanations could exist. Firstly, although toxicity to human follicles has been confirmed (219), current smoking may only influence pre-antral and antral follicles but not primordial follicles (189). The alternation of the functionality of the antral follicle pool would explain lower AMH levels and if the primordial pool remained unaltered by smoking, quitting smoking would allow both restoration of the growing follicle pool and normalization of AMH (189). Alternatively, smoking may alter steroidogenesis or the hypothalamic-pituitary-ovarian axis, as suggested by studies that indicate altered levels of sex-hormones amongst smokers (190). This could also lead to lower antral follicle counts and decreased AMH values. The lower age at menopause in smokers could then be explained, not by earlier depletion of the primordial follicle pool, but by the number of antral follicles becoming too small to sustain monthly menstrual cycles. More research is needed to elucidate the exact mechanisms between smoking and measures of ovarian reserve. There was no association between BMI and AMH. This is in line with some (220-222) but not all previous studies (180;186;223). It is important to note that the BMI in our study sample was very homogenous with few obese women (mean BMI= 24.3 +/- 3.9SD).

Strengths and weaknesses

The most important strengths of the present cohort are the large size of the cohort, the fact that it is population-based, and the extensive measurements and inventories on lifestyle factors and menstrual cycle characteristics that were collected starting at a relatively young age of 20 years. Also, the statistical analysis, using the CG-LMS model, provides advantages over classical approaches. Usually, linear regression is used with AMH on a log-scale to model the age related decline of AMH. The effect of determinants can then be tested with multivariate regression analysis while correcting for age. The LMS model has an advantage over this linear regression model as it does not assume that the skewness, the median, and the coefficient of variation are the same at different ages but models them as a function of age. In this way age is more appropriately corrected for when assessing the influence of possible determinants on AMH concentrations. Lastly the AMH Gen II assay (Beckman Coulter Ltd) was used which is believed to be the newest most generalizable measure of AMH. The samples were determined in a single laboratory, by the same experienced lab technician and the precision of assay results was validated with linearity-of-dilution assessment. Samples were frozen

within 12 hours and cryo-preserved within 6 weeks. Efficiency reasons required all samples to go through one temperature controlled freeze-thaw cycle before the AMH assay was performed. These are all factors that argue for the presence of homogeneous specimen processing and supports that we have provided reliable, reproducible AMH measures (224). However, the Gentech II assay remains a manual ELISA. Future automated ELISAs may provide even more stable measures of AMH. We did not time collection our blood samples on the same menstrual cycle day for all participants. However, it has been shown that intra-cycle fluctuation in AMH does not follow classical endocrine patterns, but that it occurs randomly throughout the menstrual cycle, thus supporting our choice of protocol (225). One weakness of the current study might be that there is just one measure of AMH per woman, furthermore the cross-sectional approach used in this study did not allow us to assess all aspects of AMH suppression by certain determinants. However we sought to make the situation in our study most comparable to clinical practice in which a physician is usually confronted with just one AMH measure. In such a situation it is important that the physician is aware of which factors may have had an influence on the AMH level that he/she is confronted with.

Clinical implications

The impact of the studied determinants for clinical interpretation depends on whether a woman has a high or a low AMH for her age to start with. If a woman is above the 75th percentile of AMH for her age, a considerable drop to a lower percentile may not be problematic, but for a woman in the 30th percentile it is conceivable that the effect may be worrisome. Current smoking was associated with a decrease of AMH of 4 percentiles. For a 26 year old woman this means that if she were in the 50th percentile of AMH without smoking she would be in the 46th percentile if she did smoke. This is comparable to a woman of 28 years old in the 50th percentile. Current smoking thus adds two years to her reproductive age in this example. The same calculation can be done for a 36 year old woman in the 50th percentile in which case her AMH would be comparable to a 37-38 year old woman if she were to smoke. For a 40 year old it would thus add 1.5 years to her reproductive age.

AMH is increasingly applied in clinical practice. Several studies have presented nomograms for age specific AMH values in a variety of populations with AMH measured with a variety of assays(178-183).

This study shows that AMH concentrations are influenced by several determinants and may thus aid in the interpretation of AMH results. Future studies on the effects of determinants on AMH would benefit from a prospective longitudinal design to help elucidate the precise mechanisms behind AMH suppression. The future clinical implications of having a high or low AMH for one's age are largely unknown. More large-scale, long term prospective studies are needed to assess the impact of having a low or high age specific AMH. Until this has been done, there is no concrete place for AMH nomograms in a clinical setting.

Conclusion

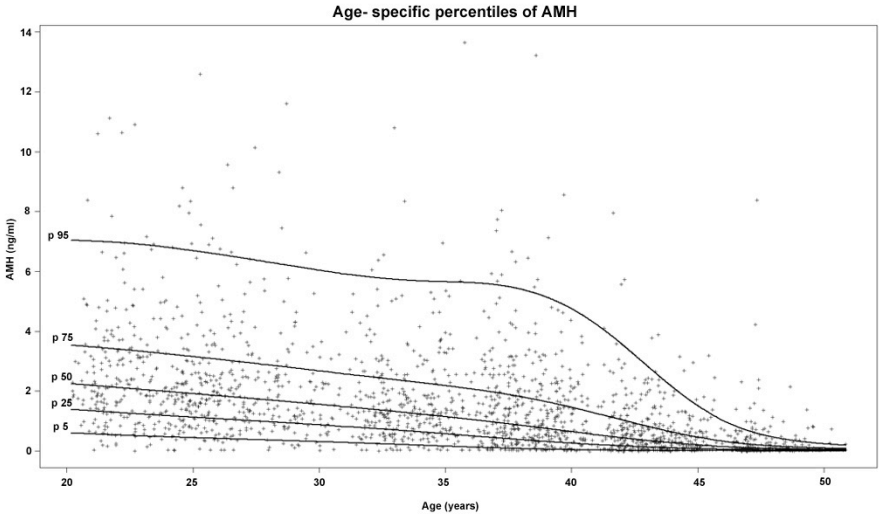
In conclusion, our study adds valuable information to the body of knowledge currently available on AMH in the general population. The lower AMH levels associated with oral contraceptive use and smoking seem reversible, as effects were confined to current use. It is important to give careful consideration to the effect of such determinants when interpreting AMH in a clinical setting.

Acknowledgements

We thank J.W.J.M. Cremers for technical assistance with AMH determinations.

Supplemental Material

Supplemental figure 1: Age specific percentiles of AMH



This figure displays a scatterplot with the distribution of AMH according to age and the solid lines represent the age-specific percentiles (5th,25th,50th,75th and 95th percentiles) of AMH as calculated by the CG-LMS model.

M. Dólleman (1)(3)*, M.J. Faddy (2), J. van Disseldorp (1), Y.T. van der Schouw (3), C.M. Messow (6), B. Leader(5), P.H.M. Peeters(3), A. McConnachie (6), S.M. Nelson (4), F.J.M. Broekmans(1)

1 Department of reproductive medicine and gynaecology, University Medical Centre Utrecht, The Netherlands
2. School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia
3. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
4. School of Medicine, University of Glasgow, McGregor Building, Floor 2, Western Infirmary, Glasgow, UK
5. Clinical Research Division, ReproSource, Inc., Woburn, Massachusetts
6. Robertson Centre for Biostatistics, University of Glasgow, Glasgow G12 8QQ, United Kingdom

Published in J Clin Endocrinol Metab. 2013 May;98(5):1946-53

Chapter 5

The Relationship Between Anti-Müllerian Hormone in Women Receiving Fertility Assessments and Age at Menopause in Subfertile Women: Evidence From Large Population Studies

Abstract

Context: Anti-müllerian hormone (AMH) concentration reflects ovarian ageing and is argued to be a useful predictor of age at menopause (AMP). It is hypothesized that AMH falling below a critical threshold corresponds to follicle depletion which results in menopause. With this threshold, theoretical predictions of AMP can be made. Comparisons of such predictions with observed AMP from population studies support the role for AMH as a forecaster of menopause.

Objective: To investigate whether previous relationships between AMH and AMP, are valid using a much larger dataset.

Setting: AMH was measured in 27,563 women attending fertility clinics.

Study Design: From these data a model of age-related AMH change was constructed using a robust regression analysis. Data on AMP from subfertile women were obtained from the population-based Prospect-EPIC cohort (n=2,249). By constructing a probability distribution of age at which AMH falls below a critical threshold and fitting this to Prospect-EPIC menopausal age data using maximum likelihood, such a threshold was estimated.

Main outcome: Conformity between observed and predicted AMP.

Results: To get a distribution of AMH-predicted AMP that fitted Prospect-EPIC data, we found the critical AMH threshold should vary among women in such a way that women with low age-specific AMH would have lower thresholds while women with high age-specific AMH would have higher thresholds (mean 0.075 ng/ml; IQR 0.038-0.15 ng/ml). Such a varying AMH threshold for menopause is a novel and biologically plausible finding. AMH became undetectable (<0.2 ng/mL) approximately 5 years before the occurrence of menopause, in line with a previous report.

Conclusions: The conformity of the observed and predicted distributions of AMP supports the hypothesis that declining population averages of AMH are associated with menopause, making AMH an excellent candidate biomarker for AMP prediction. Further research will help establish the accuracy of AMH levels to predict AMP within individuals.

Introduction

Female reproductive success and the length of a woman's fertile life-span are considered to be manifestations of the dynamic decline of the primordial follicle pool. Although age is considered to be the key determinant of this decline, in young women it is not a reliable predictor of the duration of their reproductive lifespan, with future menopausal age potentially varying considerably between women of the same age (40). Therefore, an ever-growing body of research has aimed to identify biomarkers that adequately assess the remaining supply of follicles in the ovaries. Anti-müllerian hormone (AMH) is such a quantitative marker of ovarian reserve. AMH is secreted by the cohort of antral follicles up to 8 mm in size (226;227) and has been shown to adequately reflect the gradual decline in follicle numbers associated with increasing age (228). As the onset of menopause is triggered by exhaustion of the follicle pool and considering that AMH is a reflection of the size of the remaining follicle pool, AMH has been used to predict the age at which a woman will become post-menopausal in both retrospective and prospective cohort studies (78;79;181;216). Although the predictive capacity of AMH in these studies was promising, they were based on small numbers of women, justifying further confirmation in larger cohorts.

Several models have been suggested to represent the age-related AMH decline and the age at which menopause occurs (181;183;229). Of these, a quadratic regression function of age has emerged as the preferred model (182;183). However, new insights into the nature of AMH decline, such as the suggestion that AMH becomes undetectable 5 years prior to the final menstrual period (230) have prompted a collaboration of efforts to reassess the relationship between AMH and the onset of menopause, using a previously described approach (181) but with data from a vastly larger cohort of women. The aim of the current study is thus to model age at menopause based on ovarian reserve status derived from age and AMH and to confirm the conformity of predicted and observed AMP distributions.

Materials and Methods

Subjects

To investigate age-dependent changes in AMH we combined information from different sources into one data-set for analysis. AMH and age were

obtained from women attending three centralized AMH testing facilities within the United Kingdom (UK) and one in the United States (US). The UK laboratories were the University of Glasgow (UoG, *n* = 1,407) the Glasgow Centre for Reproductive Medicine (GCRM, *n* = 1,515), and the Glasgow Royal Infirmary (GRI, *n* = 6,783). These AMH values were measured as part of the routine fertility work-up and thus represent all women that would attend these infertility clinics. An additional group of women with normal pelvic ultrasounds and confirmed ovulation but whose partners were known to have severe male factor infertility (<5,000,000 sperm/mL) requiring ICSI was also included (GRI, *n* = 927) (183). A final set of 16,931 AMH measurements was from ReproSource, a clinical reference laboratory that provides centralized AMH testing for US fertility clinics (182). For all women, samples were measured between July 2006 and October 2009. Due to the centralized sources of the AMH values, only age and AMH concentration were known with no clinical characteristics of the women available.

The distribution of age at menopause was estimated from another cohort of women from the Prospect-European Prospective Investigation into Cancer and Nutrition (Prospect-EPIC). This cohort consists of 17,357 women aged between 50 and 70 years who were recruited between 1993 and 1997 for a nationwide breast cancer screening program conducted in The Netherlands. Menopausal status and past reproductive health were derived from extensive questionnaires on reproductive history (231;232). The World Health Organization definition of menopause, namely the absence of spontaneous menstrual bleeding for more than 12 months, was used. For the current study, a cross-sectional cohort of women with a recorded natural menopause was selected from the initial prospective cohort. Only women aged 58 years and over were selected to avoid underrepresentation of women who reached menopause at a late age, and from these women only those women with some indication of subfertility provided data on age at menopause. These women were considered to be more similar (in all parameters except age) to those from whom the AMH concentrations were obtained. Subfertility was assessed via questionnaires and women were considered to have an indication of subfertility if they had one or more of the following criteria: (i) having had an irregular menstrual cycle pattern between 30 and 40 years of age, (ii) having consulted a physician for fertility problems, (iii) nulliparity, (iv) uniparity, (v) having had a miscarriage or (vi) having a long time interval between birth of first and second child (233). After application of these

selection criteria, 2,249 post-menopausal women could be included in this study.

AMH assay

All four centralized AMH testing facilities used the enzyme linked immunosorbent assay (ELISA) provided by DSL (Webster, TX) to measure AMH concentrations in batches. Values were delivered in concentrations of pmol/L (conversion factor to pmol/L = ng/mL × 7.143). At the UoG laboratory the intra-assay and inter-assay coefficients of variation (CVs) were 6.3% and 11.4% respectively; at the GCRM laboratory the CVs were 3.4% and 8.6% respectively; at the GRI laboratory 8.6% and 15.4% respectively and at the ReproSource laboratory the CVs were 5% and 8% respectively (182;183). The limit of detection of these AMH assays was set at 0.2 ng/mL (7).

Analysis

Modelling of age-related AMH decline

The data on AMH and age from the four different centres were analysed using a robust regression methodology, with quadratic functions of age to describe the means (182;183) and skew-*t* distributions to describe the residual variation about these means (234). (For maximum likelihood estimation, see below, those results less than the assay detection limit contribute the information AMH < 0.2 ng/mL to the likelihood.) A natural logarithmic transformation of AMH was applied to stabilise the residual variance in AMH concentrations, thereby creating a more homogenous distribution, but the residual standard deviation was allowed to be age-dependent as a check on the efficacy of this transformation. All these assumptions formed a model for age-related change in AMH concentrations with three components:

1. mean of log(AMH) = $\alpha + \beta \times \text{age} + \gamma \times \text{age}^2$,
2. standard deviation of log(AMH) = $\exp(\sigma + \tau \times \text{age})$, and
3. skew-***t*** distribution of residuals {log(AMH) –mean}⁄standard deviation.

Predicted and observed age at menopause.

Our hypothesis was that variation in age-specific AMH concentrations corresponds to variation in the age at which menopause occurs, through the notion of a *critical threshold* whereby AMH falling below this

threshold represents follicle depletion to the extent that cycles are no longer sustained and menopause follows. From the above (regression based) model for age-related change in AMH, probabilities of AMH at any specified age being below such a threshold can be calculated and related to age at menopause through the equation:

probability that AMH level at age y is below threshold
= probability that menopause has occurred before or at age y.

In this way a probability distribution of age at menopause can be determined for any given threshold and thus provides a model for predicting age at menopause. Recent studies have shown substantial inter- and intra-cycle fluctuations in AMH for individual women (225;235;236) which may be affecting the AMH data but could not be expected to contribute to varying fertility between women. So, to allow for such extraneous variation in AMH, a different standard deviation and skew-*t* residual distribution from those in the regression model were used in constructing this predictive distribution of menopausal age (*i.e.*, only the equation for the mean of log(AMH) from the regression analysis was used here).

These two *linked* models, for AMH and age at menopause, were fitted to both data-sets (AMH and Prospect-EPIC) by maximising the combined likelihood of all these data. This approach gave estimates of all parameters used in the model to describe the: mean (α, β, and γ), standard deviation (σ and τ), skew-*t* residual distribution for the AMH and age regression model (above), the critical AMH threshold, and the (different) standard deviation and skew-*t* residual distribution used in the construction of the predictive distribution of menopausal ages. Agreement between the AMH-based predictive distribution and the observed distribution of age at menopause was assessed by a visual comparison of their cumulative frequencies.

Nomogram

A nomogram was created to show estimated age-specific percentiles for AMH (lower 5%, 10% and 25%, median, and upper 75%, 90% and 95%) from the fitted (regression) model for age-related change in AMH, with the corresponding percentiles of the fitted distribution of menopausal age derived from the AMH threshold modelling. Individual predictions can be made from percentiles of the menopausal age distribution corresponding to those where an individual woman's AMH concentration and age are located in the nomogram. Only women between the ages of 25 and 55

years are represented in the nomogram; this age range was chosen as it represents the most clinically relevant group for prediction of remaining fertile life. Moreover, previous studies have shown age based AMH models to function poorly at the extremes of the age distribution (181-183).

Finally, the estimated distribution of the age at which AMH drops below the detectable limit of the AMH assay (0.2 ng/mL) was compared with the estimated distribution of the age at which AMH drops below the critical menopausal threshold, to assess the time span between AMH becoming undetectable and the predicted onset of menopause. All data were analysed using MATLAB, Version 7.2 software (The MathWorks Inc., Natick, MA).

Results

In the full AMH data-set, there were 3,394 women with AMH values < 0.2 ng/mL and actual values from 24,169 women. This latter group had a mean age of 34.6 years (± 5.3 years) and mean AMH of 2.5 ng/mL (± 2.8 ng/mL). In the cohort of women with any indication of subfertility in the Prospect-EPIC data-set (n = 2,249), the mean age was 63.3 years (± 3.4 years), with a mean age at natural menopause of 49.9 years (± 4.5 years) and median 50 years. ***Table 1*** provides summary statistics of the different cohorts of subjects. The mean age related AMH profiles from individual regression analyses of data from the four different sources are shown in ***supplementary Figure 1***. The fitted quadratic regressions of log(AMH) on age from the different centres are very similar, particularly over the age range of 25 to 55 years, although the differences between these profiles were statistically significant (p-value < 0.001) due to the large amount of data used. At the extremes of age, greater discrepancies were apparent but this can be largely accounted for by the estimated means being less precise there. Moreover, only 0.3% of the residual standard deviation could be attributed to differences between the sources, so these differences were deemed of no practical or clinical significance. There is a clear trend of decreasing AMH with increasing age after about 25 years in all the profiles.
The mean age related AMH profile from a single regression analysis of the combined data from all four sources plotted with the corresponding 95% confidence intervals (CI), and the 90% probability range for observed AMH values are shown in ***supplementary Figure 2***. According to this figure mean AMH starts to decline from approximately 20 years of age

Table 1. Summary statistics of the different cohorts of subjects

	GRI	GCRM	UoG	US	Prospect-EPIC
N=	7710	1515	1407	16931	2249
Age (Years)	34.0 (30.0-38.0)	36.8 (32.9-39.8)	34.9 (31.4-37.5)	35.3 (31.0-39.0)	63.0 (60.0-66.0)
AMH (ng/ml)	2.07 (0.97-3.74)	1.43 (0.74-2.77)	1.67 (0.80-2.95)	1.60 (0.76-3.15)	-
AMH measures below detection (n=)	729	196	134	2335	-
Age at Menopause (years)	-	-	-	-	50.0 (48.0-53.0)

Data are medians (interquartile range) unless otherwise indicated. GRI = Glasgow Royal Infirmary; GCRM= Glasgow Centre for Reproductive Medicine; Uog = University of Glasgow, US=Reprosource United States.

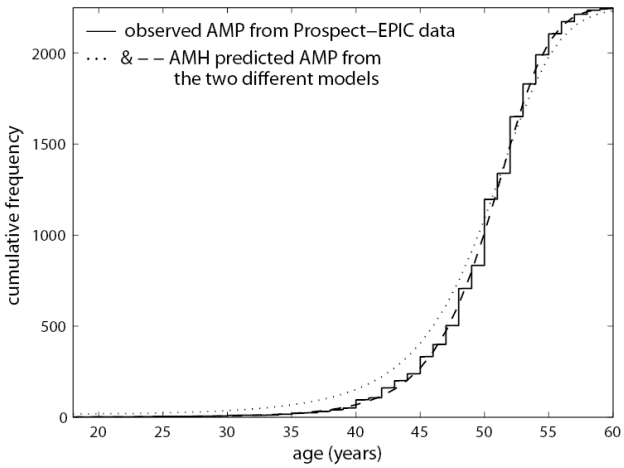


Figure 1. Predicted versus observed distribution of age at menopause.
The distribution of observed age at menopause (AMP) from the Prospect-EPIC cohort of subfertile women (solid line) is shown compared with two predictive distributions constructed from AMH falling below a critical threshold; the dotted line shows relatively poor agreement when the residual distribution of AMH from the regression analysis is used in the predictive model for AMP, while the dashed line shows much better agreement from using a different distribution of AMH.

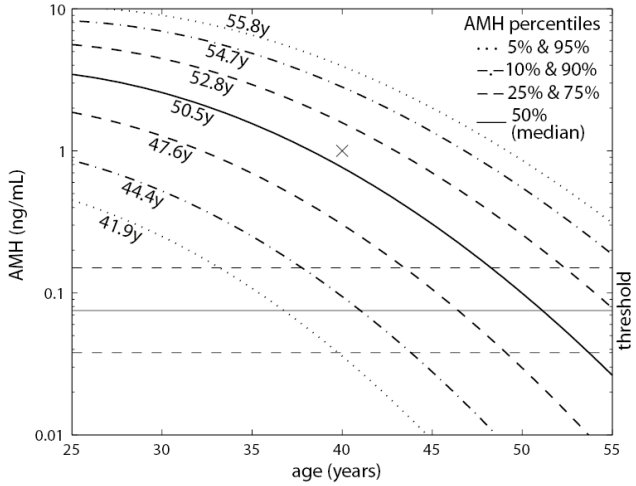


Figure 2. AMH Nomogram and predictions of age at menopause.
Estimated age-specific 5%, 10%, 25%, 50%, 75%, 90% and 95% AMH percentiles from the fitted regression model are plotted. The corresponding percentiles of age at menopause (AMP) modelled by AMH falling below a critical threshold are shown adjacent to these AMH percentiles. The critical threshold (0.075 ng/ml) is depicted by the faint solid horizontal line, while the faint dashed horizontal lines indicate the approximate inter-quartile range when the threshold is allowed to vary. A 40 year old female with an AMH of 1 ng/ml which is between the 50th and 75th percentiles (denoted by x) would thus be expected to have an AMP between 50.5 and 52.8 years.

and continues to decline steadily until it becomes undetectable. The logarithmic transformation of AMH has slightly overcompensated for the heterogeneous residual variation of the raw AMH data as can be seen from the broadening 90% probability range with increasing age.

Figure 1 shows the fit of the distribution of age at menopause predicted by AMH falling below a critical threshold estimated to be 0.075 ng/mL (with standard error 0.004), where good agreement can be seen with the distribution from the Prospect-Epic data on women with any indication of subfertility. Shown for comparison is the very poor fit obtained by assuming the same distribution of log(AMH) in constructing the predictive distribution of menopausal ages as that from the regression analysis of log(AMH) and age. In fact, the residual standard deviation from this regression analysis had to be reduced by multiplying it by an estimated factor of 0.56 (95% CI 0.51 – 0.61) to achieve the improved fit shown in *Figure 1*; in other words, there was significantly (*p*-value < 0.001) more variation in AMH levels than would be needed to explain the variation in age at menopause. All model parameter estimates are shown in *supplementary Table 1*.

In ***Figure 2*** a nomogram is depicted that describes age-related AMH decline for women in terms of the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles of age-specific AMH concentrations. The estimated critical AMH threshold (0.075 ng/ml) is also indicated in this figure. The corresponding percentiles of the predictive distribution of age at menopause (AMP) are shown adjacent to these AMH percentile curves in ***Figure 2***; estimated 5th, 10th and 25th AMP percentiles (with standard errors in brackets) are 41.9(0.27) years, 44.4(0.19) years and 47.6(0.12) years respectively, the estimated 50th AMP percentile (or median) is 50.5(0.09) years, and estimated 75th, 90th and 95th AMP percentiles are 52.8(0.08) years, 54.7(0.10) years and 55.8(0.11) years respectively. Women with AMH concentrations between percentiles can expect menopause to occur between the corresponding AMP percentiles; for example, a 40 year old women with AMH concentration 1 ng/mL which is just above the 50th AMH percentile and just below the 75th percentile (denoted by × in ***Figure 2***) can expect menopause just after 50.5 years and below 52.8 years.

Figure 3 shows that the age at which a woman's AMH drops below the *detection limit* of the assay that was used in this study (0.2 ng/mL) has a distribution which, when compared with the distribution of the age at which AMH falls below the critical threshold for menopause, indicates that menopause occurs about 5 years after AMH becomes undetectable.

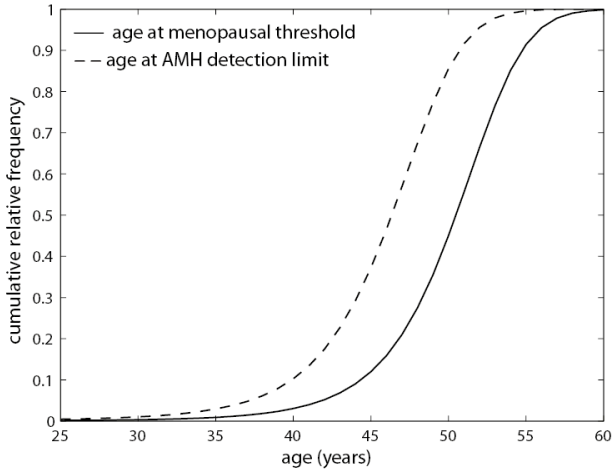


Figure 3. Comparison of the distributions of age at which the AMH threshold and AMH assay detection limit are reached.
The estimated distribution of the age at which AMH drops below the detection limit (dashed line) showing that this occurs about 5 years before AMH drops below the critical threshold for menopause (solid line).

Discussion

This study demonstrates two important things. First, we have shown a close conformity between the distribution of observed age at menopause and a predictive distribution using a robust regression model of changing AMH with increasing age and the assumption that menopause is associated with AMH falling below a critical threshold (where AMH represents follicle depletion to the extent where menopause ensues). Second, we have confirmed earlier reports that AMH becomes undetectable approximately five years prior to menopause. The close conformity of the shape of the observed and AMH-predicted distributions of age at menopause supports the hypothesis that AMH influences the timing of reproductive milestones such as menopause. The findings in this study generally confirm both the dynamics of age-related AMH decline as well as the possibility of prediction of age at menopause as demonstrated in previous studies (181-183). Both findings will have impact on research lines where prospective data are now being obtained to demonstrate the claim that AMH at young age could be a forecaster of reproductive lifespan.

The estimated AMH threshold after which menopause occurs was 0.075 ng/mL, slightly lower than the value (0.086 ng/mL) given in van Disseldorp et al. (2008) (181) which was based on a much smaller data-set of proven fertile women with only a fraction (≈ 0.5%) of the AMH values used in this study and is consequently a much less precise estimate, and in any case different menopausal age data were used. Unfortunately, it would not be possible to confirm this threshold in any prospective study as the AMH assay applied in this study is not considered to be accurate below a detection limit of 0.2 ng/mL(7;237;238). Nevertheless, as it is not uncommon to find 1 or 2 antral follicles by ultrasound in post-menopausal women, this threshold seems plausible.

The finding from our modelling, that the age at which AMH falls below the critical threshold for menopause tends to be about 5 years after the age at which AMH drops below the detection limit of the assay, is in concordance with Sowers et al. who showed that AMH values decline to or past the detection limit at approximately 5 years before the final menstrual period (230).

In achieving the good fit to the Prospect-EPIC data on menopausal ages it was necessary to reduce the standard deviation of log(AMH) used in constructing the AMH-based predictive distribution by multiplying the residual standard deviation from the regression model of log(AMH) and age by an estimated 0.56 (95% CI 0.51 – 0.61). Some excess AMH variation will be due to inter- and intra-cycle variation within women, and

Van Disseldorp et al. have shown that 11% of the (age-adjusted) AMH variance could be due to inter-cycle variation and 13% due to intra-cycle variation for individual women (235). If these two sources of extraneous AMH variation were independent then 76% of the overall AMH variance (or 87% of the standard deviation) would be due to variation between women; but if they were positively correlated then the variation between women could be as low as 52% (corresponding to maximal correlation) of the overall AMH variance (or 72% of the standard deviation). This figure of 0.72 is outside the above 95% confidence interval for the estimated factor by which the AMH residual standard deviation was reduced to achieve a good fit of the predictive distribution of menopausal age. This means that inter-and intra-cycle AMH variation within women cannot explain all of the excess variation in AMH apparent from the regression analysis, and a significant amount remains.

The effect of any excess AMH variation can be reduced by allowing the critical AMH threshold for menopause to vary between women in such a way that it is positively correlated with their actual AMH concentration: higher for women with high AMH for their age and lower for women with low AMH for their age, thereby reducing the variation of the AMH-predicted age at menopause. Using the above figure of 72% of the AMH residual standard deviation in the determination of the predicted distribution of menopausal ages leads to an approximate estimate of the inter-quartile range of the necessary threshold variation of 0.038 – 0.15 ng/mL. This is indicated in ***Figure 2*** with the previous (constant) threshold estimate (0.075 ng/mL) now being the mean. Thresholds between this inter-quartile range would be more likely for women with AMH concentrations between the corresponding (25^{th} and 75^{th}) AMH percentiles than for women with AMH outside these percentiles who would be more likely to have more extreme threshold values.

This discrepancy between variation in age at menopause and residual variation in AMH may not be an artefact of AMH, for a similar discrepancy between variation in age at menopause and residual variation in non-growing follicles (NGF) is apparent in the study by Wallace et al. (239) where menopausal age prediction was based on follicle numbers falling below a critical level of 1000: in that study the 95% prediction interval was 39 – 60 years compared to the observed 40 – 57 years from the Prospect-EPIC data from women with and without any indication of subfertility. Perhaps it can be reasoned as follows, if women with a higher than average AMH for their age are predestined to have a higher age at

menopause, then this would result in an increased time frame for other determinants of the ovarian ageing process to play a role, thus resulting in menopause occurring despite still having some follicles left in the ovaries. The supposed determinants may be somatic or ovarian factors that prevent ovulation and cycling at near, but not complete, depletion of the follicle pool. Alternatively, a compensatory mechanism may exist whereby prematurely aged ovaries (as expressed by lower than average age-specific AMH levels) have a decreased threshold for menopause as a means of extending reproductive life. Evidence for such a varying threshold can be gleaned from recent studies where data show that some women are regularly cycling despite their AMH concentrations being clearly below the detection limit and very close to zero (179;225). Furthermore, a study in 50 post-menopausal women revealed that 36% of women still had an AMH above the assay's limit of detection at their final menstrual period while 64% were below it (230).

There may also be other sources of extraneous AMH variation which could reduce the need for a varying threshold. Some extraneous AMH variation in this study would be from the different sources of the data, but this only amounted to a negligible 0.3% of the residual standard deviation. It could be argued that the *excess* variation in AMH may simply be explained by characteristics of the cohort. As the AMH values are measured in women attending an infertility clinic, one can expect the cohort to include women with either polycystic ovary syndrome (high age-specific AMH) or premature ovarian insufficiency (low age-specific AMH) which could contribute to excess variation in AMH (176;240). Extra variation in AMH may be due to lifestyle factors like smoking which has been associated with a decrease in age-specific AMH and an earlier age at menopause (189;207;241). Assuming similarity between the AMH cohort and the Prospect-EPIC cohort, it is however likely that the level of occurrence of such determinants of AMH would also be similar and hence their effects on AMH-predicted age at menopause might be expected to be similar.

The major strength of this study is that it provides the largest body of information in which the association between AMH and age at menopause has ever been tested. There are however some limitations; for example, data on age at menopause was based on self-reporting which may be prone to recall bias. However, several studies have demonstrated that both the validity and reproducibility of self-reported age at menopause are good (242;243). And from the cohorts of women from whom AMH was measured, only age and AMH is known; other determinants of age

at menopause such as genetic factors, lifestyle factors, and reproductive history could not be compared between this sample of women and the Prospect-EPIC sample (198). A varying AMH threshold might be seen as a bit of a drawback in the model for age at menopause in that although it results in a better fitting model, it cannot be known with any certainty for an individual at what threshold AMH would be predictive of menopause. Some uncertainty in predictions is already apparent according to a prospective study which showed large and overlapping 95% confidence intervals for the predicted ages at menopause (78).

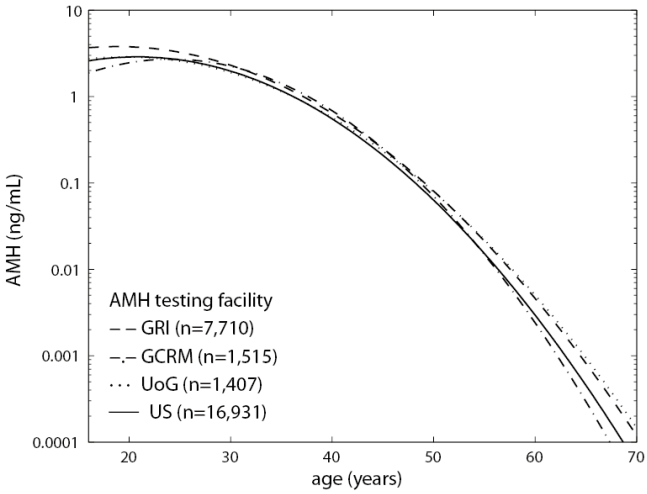
Although the findings of the current study support the notion that AMH does reflect female reproductive status, applying this to get clinically useful information for individual women remains problematic. Firstly the AMH assay is not widely used anymore as it has been replaced by a newer AMH GEN II assay. AMH concentrations measured with the GEN II assay give higher values than the ELISA used in this study (182;183). Although consistent correlations have been found between the newer AMH GEN II assay and the DSL assay (237;244), the results shown here should be interpreted as conceptual evidence that prediction of age at menopause from declining AMH is possible, rather than used for counselling of individual patients. Further research on the relationship between AMH and age at menopause would benefit from using the newer assay.

Using average AMH concentrations over several cycles and phases of these cycles would reduce the effect of natural fluctuations between and within cycles; for example, averaging four measurements could reduce the standard deviation of these effects by 50%. Additional evidence still needs to be obtained from long-term follow up studies in which the occurrence of menopause is prospectively assessed and women are subjected to multiple measurements of AMH over several cycles spaced over a period of one or two decades. Four prospective studies exist that provide evidence that AMH can be used to make more individualised predictions of age at menopause. However, either the follow-up time was not very long resulting in few women reaching menopause or the included women were of late reproductive age when AMH was determined. Furthermore, in three of the studies AMH was only measured once, thus not permitting allowance for natural variation within the individual (78;79;216). In the other study, prediction was based on estimating the rate of change in AMH but it took 3.5 years to get a reliable estimate, and thus 3.5 years before a prediction could be made (216).

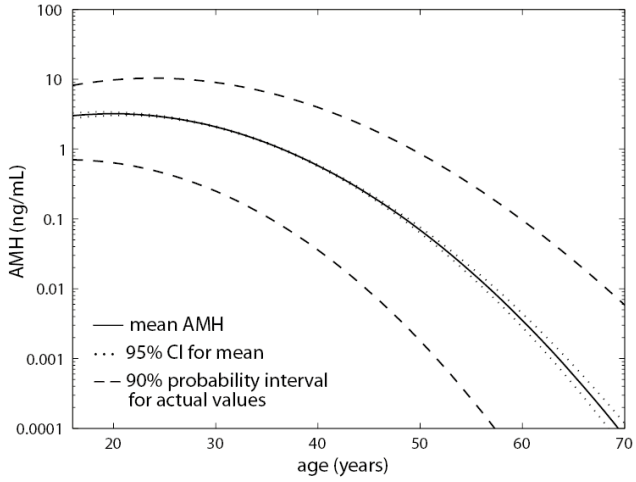
Menopause is the only noticeable marking point for the massive, gradual decline in ovarian follicle numbers over the first 5 decades of a female's life. It can be seen as the end of the reproductive lifespan in women. Age at menopause shows considerable variation between 40 and 60 years, with approximately 10% of women becoming menopausal before the age of 45 years. A relationship between age at menopause and the end of natural fertility is thought to be present, with an interval of approximately 10 years (2). Therefore, from prediction of age at menopause, similar predictions of age at the end of natural fertility can be extrapolated.

In conclusion, this study shows that AMH levels have an association with reproductive events such as age at menopause and reinforces the notion that AMH is capable of predicting the timing of such events more informatively than chronological age alone.

Supplementary data



Supplementary Figure 1. Regression of AMH and age data from the different sources. Fitted quadratic curves of mean log(AMH) for increasing age are shown for the four testing facilities (displayed on the log-scale): GRI= Glasgow Royal Infirmary; GCRM=Glasgow Centre for Reproductive Medicine, UoG=University of Glasgow; US= United States data.



Supplementary Figure 2. Regression of combined AMH and age data from all four sources. The fitted quadratic curve of mean log(AMH) for increasing age is shown by the solid line. The dotted lines indicate 95% confidence intervals for this mean. 90% probability ranges for actual values are shown by the dashed lines. (All on the log-scale).

Supplementary Table 1. Results of the statistical analysis.

Regression model for AMH and age	
mean of log(AMH):	– 0.51(0.17) + 0.17(0.010)age –0.0042(0.00016) age²
standard deviation of log(AMH):	exp{–0.59(0.039) + 0.027(0.0010)age}
skewness of log(AMH) residual distribution:	– 0.37(0.0080)
Predictive model for age at menopause	
critical AMH threshold:	0.075(0.0039)
standard deviation of log(AMH):	exp{–1.17(0.055) + 0.027(0.0010)age}
skewness of distribution of log(AMH):	– 0.39(0.024)

Maximum likelihood parameter estimates (with standard errors in brackets) of components of the models for changing AMH with increasing age and for predicting menopause by AMH falling below a critical threshold.

M. Dólleman[1,2], M. Depmann[1], M.J.C. Eijkemans[2], J. Heimensem[1], S.L. Broer[1], E.M.van der Stroom[3], J.S.E.Laven[4], I.A.J. Van Rooij[1], G.J. Scheffer[1], P.H.M. Peeters[2], Y.T. van der Schouw[2], C.B. Lambalk[3], F.J.M. Broekmans[1]

1. Department of Reproductive Medicine and Gynaecology, University Medical Center Utrecht.
2. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands.
3. Division of Reproductive Medicine, Department of Obstetrics/Gynaecology VU University Medical Center, Amsterdam, The Netherlands.
4. Division of Reproductive Medicine, Department of Obstetrics and Gynecology, Erasmus Medical Center, , Rotterdam, The Netherlands

Accepted for publication in Human Reproduction

Chapter 6

Anti-Müllerian Hormone is a More Accurate Predictor of Individual Time to Menopause Than Mother's Age at Menopause

Abstract

Context: Mother's age at natural menopause (ANM) is considered a proxy for daughter's ANM although studies on its predictive accuracy are lacking. Anti-müllerian hormone (AMH) is a biomarker with known capacity to predict ANM. However, its added value on top of known predictors ,like mother's ANM, is unknown.

Objective: To quantify the predictive value of mother's ANM in the prediction of daughter's time to menopause (TTM), and to assess the added value of AMH when mother's ANM is known.

Design: Cox proportional hazards analysis estimated uni- and multivariate regression coefficients for female age at study entry, mother's ANM and AMH in the prediction of TTM. Discrimination of models was assessed with C-statistics. Clinical added value of AMH was quantified with a Net Reclassification Index (NRI).

Setting: Population-based cohort studies

Participants: To assess additive predictive value of mother's ANM, 165 mother-daughter pairs were used (group 1). To assess the added value of AMH, a second group of 150 women in whom AMH and mother's ANM was recorded prior to a 12-year follow-up period during which daughter's ANM was assessed was used (group 2).

Main outcome: Accuracy of TTM prediction.

Results: A model with female age and mother's ANM had a c-statistic of 79% and 85% in groups 1 and 2 respectively. Both age and mother's ANM were significantly associated with TTM. In group 2, the multivariate model with age, mother's ANM and AMH had a c-statistic of 92%. Only female age and AMH remained significantly associated with TMM. The mean weighted NRI suggests a 47% improvement in predictive accuracy is offered by adding AMH to the model of age and mother's ANM.

Conclusions: AMH and mother's ANM both have added value in forecasting TTM for the daughter based on her age. In comparison, AMH is a more accurate added predictor of TTM.

Introduction

Menopause marks the definite end of the fertile lifespan. The average age at which a woman in the more developed countries enters menopause is 51. However, chronologic age at menopause shows considerable individual variation and ranges between the ages of 40-60 years, with approximately 10% of women becoming menopausal before 45 years of age (2;245).

A fixed temporal relationship between age at menopause, the end of natural fertility and the start of subfertility is thought to be present (2;43). Predicting age at menopause is therefore clinically relevant as it could give women a more accurate idea of the length of their fertile life span which, in turn, may be used during informed decision making about timing of child-bearing. The large variability in menopausal age has prompted researchers to find a more reliable marker than chronological age in predicting age at natural menopause (ANM).

Heritability of age at menopause has been recorded to be substantial, with heritability rates varying between 30-85% (61-63;65;246). Although genetic studies on variation in age at natural menopause have delivered interesting results, currently only 2.5-4.1% of natural variability can be explained by involved common genetic loci. It is expected that in the near future, using more refined genetic techniques, more rare variants will be discovered which might predict more accurately the ANM. (66). For this reason, the age at which a woman's mother reached menopause may be useful as a tool to indicate in what age range a woman herself will become menopausal. Interestingly, no studies have formally assessed the true predictive value of mother's ANM for the forecasting daughter's ANM.

More recently, ovarian reserve tests such as the serum concentration of anti-müllerian hormone (AMH) have been suggested as valuable markers for predicting the size of the primordial follicle pool, i.e. the ovarian reserve (283). As such AMH serves as a proxy for the number of follicles remaining in an individual's ovaries. Since, the exhaustion of the primordial follicle pool coincides with the ANM AMH might constitute a marker for menopausal age as well (78;79;247-249). Interestingly mother's ANM has recently also been found to be a determinant of AMH levels in the daughters (206).

This paper aims to answer two important questions: First, what is the predictive value of mother's ANM in the prediction of daughter's ANM and second, what is the added value of AMH in this prediction when mother's age at menopause is already known?

Methods

Participants and study design

Two study groups of women, from different individual cohorts, contributed information to this study. Firstly, a group of women (cohort 1) in whom both mother's ANM and daughter's ANM were prospectively collected was used to assess the predictive value of mother's ANM in forecasting daughter's ANM. A second, pooled group of women (cohorts 2 and 3) with recorded information on AMH and mother's ANM at baseline who were followed up for more than 10 years, were used to assess the ***added value*** of AMH when mother's ANM is already known. Group 1 was used for two reasons, the large number of mother-daughter pairs
allows calculation of reliable regression coefficients for mother's ANM in the prediction of time to menopause (TTM). Secondly, it was used to verify the magnitude of regression coefficients in group 2 so that the added value of AMH on mother's ANM could be adequately studied without the risk of overestimation due to a smaller study population in group 2.

Study Group 1- Cohort 1

Cohort 1 consisted of female volunteers participating in a prospective follow up study on determinants of the development of breast cancer (250). This study consisted of four birth cohorts, DOM1 1911-1925;DOM2 1926-1931;DOM3 1932-1941 and DOM4 1942-1945. For this study, mothers were selected from the oldest and daughters from younger birth cohorts. Probabilistic linkage was used to identify mother-daughter pairs on the basis of: date of birth of the mother, date of birth of the children, birth order, and part of the (maiden) name, as previously documented and successfully applied (63;65). Information on age at menopause, and whether this was natural or iatrogenic menopause, was collected from questionnaires. Upon first screening the majority of mothers were already post-menopausal. Daughters were either post-menopausal at inclusion or menopausal age was assessed in a follow-up round. In total, 164 mother-daughter pairs were identified in which both females experienced natural menopause (65). Written informed consent was received from all women and the study was approved by the Institutional Review Board of the University Medical Center Utrecht, The Netherlands.

Study Group 2- Cohort 2

Study group 2 is a pooled group from cohorts 2 and 3. Cohort 2 consists of 265 women aged 21-46 years with a regular cycle, who had not taken

contraceptive medication for at least three months and who had no history of infertility or ovarian surgery at inclusion. For more details please refer to Broer *et al.* 2012. At cohort recruitment, AMH was measured and the age at which the participant's mother became menopausal was recorded. At the two follow-up rounds, approximately 11 and 13 years later, women were assessed with questionnaires. The questionnaires pertained to menstrual cycle characteristics, the occurrence of menopause, use of hormones or other medication, as well as reproductive history. Menopause was defined as the absence of menstrual periods for 12 consecutive months. The studies were approved by the Medical Ethical Review Committee of the University Medical Center Utrecht or the Erasmus Medical Center Rotterdam, and written informed consent was received from all women.

Study Group 2- Cohort 3

Cohort 3 was recruited for a study that assessed whether age at menopause was different between women who did and did not have a history of a trisomy-21 pregnancy. Cohort 3 consisted of 220 women aged 25-40 years with regular menstrual cycles. All participants had experienced two or more spontaneous menstrual cycles after discontinuation of breastfeeding or the use of oral contraceptives, no women had gynaecological surgery at inclusion. For more information please refer to van der Stroom *et al.*. 2011(50). AMH and mother's age at menopause was recorded at cohort recruitment. At follow up, approximately 11 years later, the cohort was approached with a questionnaire on general medical and gynaecological/obstetric history with which age at menopause was assessed. Menopause was defined as the absence of menstrual periods for at least 12 consecutive months. Approval for this study was received from the VU Medical Center's scientific review board and ethics committee and written informed consent was obtained from all participants.

Hormone assays

AMH was measured in baseline samples as previously described in the original studies (50;78). In short, in cohort 2 two different AMH assays were used (DSL and Immunotech Coulter) and in cohort 3 the same Immunotech Coulter assay was used as in cohort 2. The AMH measures from cohorts 2 and 3 were determined in the same laboratory in Rotterdam, The Netherlands, during the same time-period. A laboratory-specific conversion factor was calculated with which AMH levels between the two assays were made comparable: AMH levels from the DSL assay were multiplied by a factor of 2 to allow comparison with the Immunotech

Coulter assay as previously successfully described and applied in the original study (78). The Diagnostic Systems Laboratories (DSL) (Webster, TX) had a detection limit of 0.026 ng/ml and inter- and intra-assay coefficients of variation for the were < 5% and <11% respectively. The immunosorbent assay from Immunotech-Coulter (Marseille, France) had a detection limit of 0.05 ng/ml. Intra and interassay coefficients of variation were < 5% and 8%, respectively.

Data analysis

All participants from group 2 had to have complete information on age, AMH and mother's ANM at the start of follow up. ***Figure 1*** shows which 150 of the original 522 women from cohorts 2 and 3 remained eligible for analysis after exclusion of women with missing values. Baselines characteristics were described as medians (IQR) or means (95%CI). To analyze whether the sample of women in group 2 in which we knew mother's ANM was not different from those in which this was missing, baseline characteristics between these women was compared with independent samples T-tests. To justify pooling of AMH values from cohorts 2 and 3, age-specific AMH values were compared. Three AMH values were under the assay's detection limit, these values were included as 0.05 ng/ml.

The age at which a woman enters the cohort has intrinsic predictive value on the TTM prediction. While a young woman at entry will have a low a priori probability of entering menopause in a follow up period of 10-15 years, a 45 year old will have a high probability in the same follow up period. Additional factors may fine-tune these expectations; therefore, the added value of mother's ANM and AMH on top of age at entry (henceforth referred to as "age") was assessed. Cox proportional hazards analysis was used, with follow-up time on the time-axis, to estimate the univariate regression coefficient for age in the prediction of TTM and the multivariate regression coefficients for age, mother's ANM and AMH the prediction of TTM. Follow-up time was described as the number of years until menopause was reached or as the total number of years until the most recent follow-up for women who were premenopausal at the last assessment (at which they were censored). Women who underwent gynaecological surgery were censored at the time of operation, and women taking hormonal medication were censored at the age at treatment initiation. If this information was missing these women were excluded (***Figure 1***). The shape of the associations was assessed and

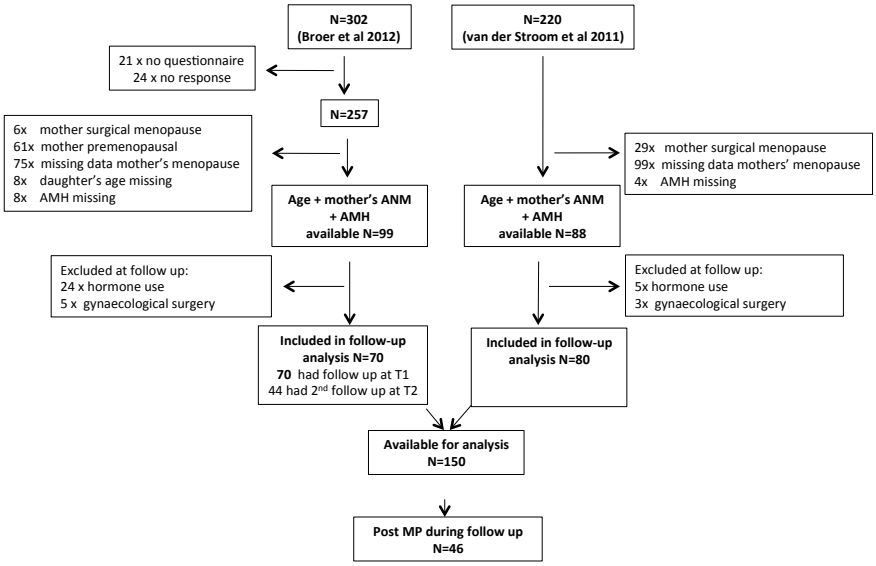


Figure 1. Flowchart of eligible women

where necessary, data were transformed with restricted cubic splines. Regression coefficients with standard errors were transformed to Hazard Ratio's (HR, 95%CI) to simplify interpretation. The discrimination of the univariate and multivariate models was assessed with c-statistics (95% CI). A net reclassification index (NRI) was calculated for the different models with age, mother's ANM and AMH. An NRI quantifies the improvement offered by new markers by examining the extent to which a new marker reclassifies subjects at a higher or lower risk of having an event during follow up (251). A continuous NRI (cNRI) was chosen as no risk categories for the occurrence of menopause exist. The cNRI counts the direction of change per individual instead of counting the percentage that crosses a particular risk threshold. Each patient is counted as +1 or −1 depending on whether the change in calculated risk was in the correct direction (higher for those with events, lower for those without events) (252)**.** The NRI is the sum of the "event NRI" and the "non-event NRI", where the event NRI is the net proportion of patients who did experience menopause during a ten year follow up who had an increase in calculated risk and the non-event NRI is the proportion of women without menopause who had a decrease in calculated risk. The maximum possible cNRI is 200% as, theoretically, all women with an event and all without an event can be correctly reclassified. For ease of interpretation we also reported the average of the two net

percentages. In a sensitivity analysis we excluded daughters with a history of a trisomic pregnancy to see whether this affected the prediction (251).

Results

Baseline characteristics are shown in ***Table 1***. From the women in group 2 in whom no mother's ANM was recorded, 61 indicated that their mothers were still premenopausal and 35 indicated that their mothers experienced surgical menopause. The women in whom mother's ANM was missing were younger than women in whom mother's ANM was known (33.3 versus 35.5 years, p-value <0.001). All other baseline characteristics were comparable.

Table 1. Baseline characteristics.

	Group 1	Group 2		Group 2 Pooled
	Cohort 1	Cohort 2	Cohort 3	Cohorts 2 +3
n=	164	70	80	150
Age at inclusion med(IQR)	39.0 (37.0-41.0)	37.0 (32.8-42.2)	34.7 (33.2-36.6)	35.5 (33.0-38.5)
Age at follow up med(IQR)	49.5 (47.0-52.0)	51.6 (47.0-54.5)	47.2 (45.1-48.6)	48.4 (45-51.4)
Years of Follow up mean (SD)	10.1 (5.6)	12.5 (11.6-14.9)	12.1 (11.6-12.9)	12.4 (11.6-13.2)
AMH med(IQR)	NA	0.6 (0.3-2.4)	1.8 (1.0-3.5)	1.5 (0.5-2.9)
Age at menopause mean (SD)	48.8 (4.2)	50.8 (3.5)*	49.9 (3.0)*	50.7 (4.1)*
Mother's age at menopause mean (SD)	49.8 (4.1)	49.8 (4.3)	49.1 (4.3)	49.4 (4.3)
Menopausal at follow-up (n=)	164	33	13	46

Numbers are medians (med) with interquartile rangers (IQR) or means (SD) as denoted. Means denoted with an * are calculated on the basis of the survival analysis.

Figure 2 shows the comparability of age-adjusted AMH values between cohorts 2 and 3, especially from 30-45 years, thus justifying the pooling of these AMH values.

Accuracy of mother's age at natural menopause

Results of uni- and multivariable analyses are presented in ***Table 2***. In group 1, both age at entry and mother's ANM appeared predictive of TTM

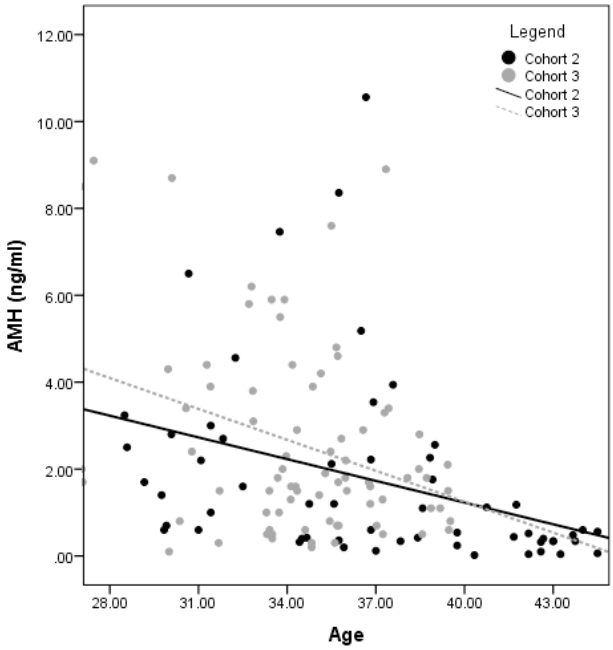


Figure 2. Comparison of age-specific AMH values between Cohorts 2 and 3
Age specific AMH values are shown to justify pooling of AMH values between cohorts 2 and 3.

in the daughter. In the multivariable model, both predictors remained significantly associated with TTM. The HR for age was 1.54 (95% CI 1.42-1.66) meaning that an increase in age at baseline increased the hazard of menopause during follow up by 1.5 times, while an increase in mother's ANM by one year decreased the hazard of menopause by 7% (HR=0.93; 95%CI 0.90-0.96). The c-statistic of this two factor model was 79% (95%CI 76-82%), meaning it can discriminate between women who enter menopause early and women who enter menopause late during follow-up with an accuracy of 79%.

Added value of AMH on mother's age at natural menopause

In a multivariable model with age and mother's ANM, both predictors were significantly associated with TTM: HRs 1.58 (95%CI 1.41-1.78) and 0.91 (95%CI 0.84-0.97), for age and mother's ANM respectively. The c-statistic of this model was 85% (95%CI 79-91%). In a model with all three predictors, female age at entry and AMH remained significant predictors but mother's ANM was no longer significant: HRs for age, mother's ANM

Table 2. Univariate and Multivariate Cox Regression Analysis for TTM Prediction in the daughters

	Group 1 (Cohort 1)					Group 2 (Cohorts 2+3)				
	Regression Analysis			C-statistic		Regression Analysis			C-statistic	
	HR	95% CI	P-value	C-index	95%CI	HR	95% CI	P-value	C-index	95%CI
Univariate Regression										
Daughter's Age	1,54	1.42-1.67	<0.0001	0,77	0.73-0.81	1,59	1.42-1.78	<0.0001	0,84	0.78-0.90
Mother's ANM	0,93	0.90-0.96	<0.0001	0,59	0.51-0.67	0,89	0.75-1.06	0,001	0,63	0.54-0.72
Daughter's AMH						0,02	0.01-0.10	<0.0001	0,86	0.81-0.91
Multivariate Regression										
Daughter's Age + Mother's ANM										
Age	1,54	1.42-1.66	<0.0001	0,79	0.76-0.82	1,58	1.41-1.78	<0.0001	0,85	0.79-0.91
Mother's ANM	0,93	0.90-0.96	<0.0001			0,91	0.84-0.97	0,01		
Daughter's Age + Daughter's AMH										
Age						1,40	1.25-1.57	<0.0001	0,91	0.88-0.94
AMH						0,05	0.01-0.22	<0.0001		
Daughter's Age + Mother's ANM + Daughter's AMH										
Age						1,41	1.26-1.59	<0.0001	0,92	0.88-0.96
Mother's ANM						0,93	0.87-1.01	0,08		
AMH						0,06	0.02-0.24	<0.0001		

This table shows the results of the univariate and multivariate regression analysis for different predictive models. On the left are the results for the study group 1 and on the right for study group 2, from the pooled data from cohorts 2 and 3. Hazard Ratios (HR) are displayed with their 95% CI and the corresponding p-value for the predictor. A value of $p<0.05$ was considered significant. C-statistics and 95%CI are shown to the right of the HRs.

and AMH were 1.41 (95%CI 1.26-1.59), 0.93 (95%CI 0.87-1.01) and 0.06 (95%CI 0.02-0.24) respectively. This model had an accuracy of 92% (95%CI 88-96%), which is similar to a model with only age and AMH (c-statistic 91%; 95%CI 88-94%) and better than a model with age and mother's ANM. In the sensitivity analysis excluding the 44 women who had a history of a trisomy-21 pregnancy (cohort 3), the results were almost identical to the results in the whole group (HR 1.39 for age, 0.93 for mother's ANM and 0.08 for AMH). ***Figure 3*** illustrates the hazard ratios for age, AMH and mother's ANM in the model with all three parameters over the relevant range per parameter.

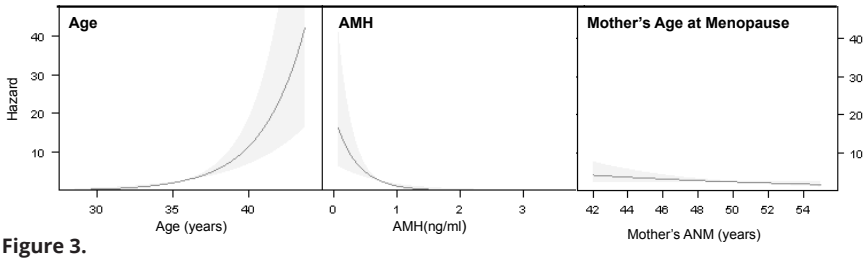


Figure 3.

Table 3. Net Reclassification Index of a model with Daughter's age, mother's ANM and AMH

Model	Event NRI	Nonevent NRI	NRI	Average NRI
Daughter's age		Reference		
Daughter's age + mother's ANM	11%	32%	43%	21,50%
Daughter's age		Reference		
Daughter's age + daughter's AMH	48%	41%	89%	44,5%
Daughter's age + mother's ANM		Reference		
Daughter's age + mother's ANM + daughter's AMH	55%	39%	95%	47%

This table presents the elements of the continuous NRI which show the improvement offered by adding AMH to a model with daughter's age and mother's ANM.

The NRIs per model are shown in ***Table 3***. The model with AMH in addition to age and mother's ANM correctly reclassified an extra 55% of women who did become post-menopausal during follow-up to a higher risk category (event NRI) and correctly reclassified an extra 39% of women who did not become post-menopausal to a lower risk level (non-event

NRI) in comparison to a model with only age and mother's ANM. This corresponds to an average weighted improvement of 47% or a NRI of 95% (***Table 3***). Another way to evaluate improvement offered by AMH can be described in terms of the increase in the accuracy with which a model discriminates between women who enter menopause early or late during follow up. The c-statistic of 85% of the model with age and mother's ANM has 15% to gain in accuracy to attain a perfect c-statistic of 100%. Through addition of AMH to this model the c-statistic rises by 7 of these 15%, which represents almost half of the total amount of accuracy that can possibly by gained by addition of any other marker of TTM.

Discussion

Main findings

The current study demonstrates that mother's ANM provides specific information in forecasting the TTM of the daughter. This information adds to the predictive ability of female age itself in estimating the probability of the occurrence of menopause within the next 10-15 years with an accuracy of ~80%. AMH is shown to independently add value to this prediction, and is suggested to be a more accurate added predictor than mother's ANM.

Findings in view of existing literature

We have shown that mother's ANM has reasonable accuracy in the prediction of daughter's TTM. Although the c-statistics between groups 1 and 2 are not directly comparable due to differences in follow-up duration and the incidence of menopause in the daughters, the HRs are directly comparable. In the multivariate analyses in both groups the HR for age and AMH were very comparable (HR 1.54 versus 1.58 for age and 0.91 versus 0.93 for mother's ANM). This similarity implies two important things: firstly, that despite the small sample size of group 2 and the large number of women that had to be excluded, the estimated hazard ratios are reliable and secondly, that the association between mother's ANM and TTM does not seem to be influenced by whether mother's ANM was recorded by the mothers themselves (group 1) or by the daughters (group 2). Considering that in the clinic, the daughters provide such information, this is an important finding.

It is commonly understood that ANM is a heritable characteristic with a 40-85% heritability (61-63;65;246). Up to now, however, no study has

assessed the predictive value of mother's ANM on daughter's ANM. This is surprising, as clinicians may base their therapeutic approach on this information (e.g. by early initiation of IVF in women whose mothers experienced early menopause) especially when found in combination with slightly lower measures of ovarian reserve.

The best prediction of TTM involved age and AMH, or age, mother's ANM and AMH with c-statistics of 91% and 92%, respectively. The event NRI suggests that in 55% of women who will enter menopause within 10 years, their predicted risk is adequately increased through addition of AMH. This corresponds to an increase in accuracy from 85% with 7% with which women with a short TTM can be discriminated from women with a long TTM. Together, these results advocate AMH as a useful added marker for menopause prediction. The present findings are in line with existing literature that demonstrate interdependency of genetic variants within the AMH molecule as well as in the AMH type II receptor on one hand and variations in ANM on the other hand (253). A recent study has revealed mother's ANM to be a determinant of AMH (206). The stronger role for AMH in predicting TTM compared to mother's ANM, may be explained from several observations. Firstly, it has been shown that AMH is influenced by environmental determinants such as smoking which may also influence menopausal age (254). Information on mother's ANM, on the other hand, will limit itself to the genetic factors shared by mother and daughter. Second, it is likely that reproductive longevity is influenced by both genetic and environmental influences with the genetic component reflecting both a maternal and paternal genetic contribution. Therefore, whilst information from mother's ANM only reflects the maternal half of the genetic influence, AMH may reflect the sum total of genetic and environmental influences.

Recently, retrospective as well prospective studies have emerged that advocate AMH has a prognosticator of ANM (78;79;181;247-249). Although with promising results, none established the added value of AMH on top of patient history information such as mother's ANM. Although our results favour AMH over mothers ANM for forecasting of TTM, considering the relatively small number of women in this study, our findings need confirmation in studies with a long follow-up period allowing improvement of TTM predictions for young women at the beginning of their fertile lifespan.

Strengths and weaknesses

The main strength of this paper lies in the uniqueness of the cohorts. The self-reported ANM from both mothers and daughters in cohort 1 made it an ideal cohort in which to assess the accuracy of mothers ANM, a finding that has not been previously published. It also provided a reliable way to confirm findings from group, 2 thereby verifying that despite the small numbers, that these models do not overestimate the predictive power of the studied predictors.

One limitation of this study is the use of several cohorts, of which one consisted of 3 studies. However, long-term follow up studies on menopause are scarce, especially when information on both AMH and mother's ANM must be known. The biggest difference between cohorts is the recruitment of women on the basis of trisomic pregnancy in the obstetric history in cohort 3, however, a sensitivity analysis without these women did not alter the accuracy of the predictions thus justifying the inclusion of these women. Although the number of women that were available for the analyses on the added value of AMH was small, according to the rule of thumb that one candidate predictor may be assessed per 10 events (natural menopause at follow up), only 30 events had to occur to have enough power to assess the value of the three candidate predictors. We had 46 events in 150 women. The main reason for this small number was due to poor registration of mother's ANM in cohorts 2 and 3. Comparison of the group in which mother's ANM was recorded to the group in which it was missing showed no substantial differences, apart from a younger age in those women where this information was missing. It is possible that mothers were not yet menopausal in these younger women. Nonetheless, these missing values may have led to selection bias about mother's ANM. However, the median ANM of the mothers was 50 years (IQR 47-52 years), which is well within the normal range of age at menopause suggesting that such bias is likely to be minimal. Another possible limitation is that AMH values were measured with two different assays. Recent studies have questioned the reproducibility of AMH values (224). However, all measurements were carried out in the same laboratory by the same experienced lab technicians during the same time-period, with a within-lab developed conversion of the two assay systems, thus

justifying one-on-one comparison of these data. Furthermore, in the original study, a subgroup analysis comparing the performance of AMH measured with different assays revealed no significant difference(78).

Clinical implications

Much variation exists in the rate of reproductive ageing amongst women of the same chronological age, as evidenced by a range of menopausal ages that lies between 40-60 years (2). Considering a fixed temporal relationship this means that natural sterility would ensue between 30-50 years, with the ***start*** of subfertility occurring approximately ten years prior (43). With current trends in delayed childbearing it is conceivable that a considerable proportion of women who delay childbearing would require help to conceive. Therefore, prediction of future ANM may be a better forecaster of reproductive performance than chronological age alone (2) and women could use this information to make informed decisions about the age until which they can delay starting a family. This may reduce the need of assisted reproduction for age-related subfertility, which has a success rate of 50% at most. However, if this information is to be clinically applied, true predictive accuracy must be substantial. Our results describe reasonable accuracy and 95%CIs. However, previous studies that provide individual predictions of ANM have considerably large 95% confidence intervals (78;248). Before achieving clinical applicability, the certainty with which a woman's prediction is made must improve.

Conclusion

This study shows that mother's ANM provides additional information, on top of female age in the prediction of TTM. Furthermore, we found AMH to be a more accurate predictor of individual TTM than mother's ANM. The optimal prediction is made using a combination of female age at and AMH; adding mother's ANM does not further improve this prediction.

F. R. Tehrani,MD1*, M. Dólleman MD2,3, J. van Disseldorp MD, PhD 2, S.L.Broer MD, PhD2, F. Azizi MD1, M. Solaymani-Dodaran MD, PhD, MPH, FFPH, CCT 1, B.C. Fauser MD, PhD2, J.S.E. Laven MD, PhD 4, M.J.C. Eijkemans PhD3 F. Broekmans MD, PhD2

1. Reproductive Endocrinology Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
2. Department of Reproductive Medicine and Gynaecology, University Medical Center Utrecht
3. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
4. Division of Reproductive Medicine, Department of Obstetrics and Gynecology, Erasmus Medical Center, University Medical Center, Rotterdam

Under Review at Climacteric

Chapter 7

Predicting Age at Menopause with Anti-Müllerian Hormone: A Comparison and Crossvalidation Study of Two Existing Models

Abstract

This study crossvalidated two Weibull models of age at natural menopause prediction from two cohorts (SRV and TLGS cohorts). It summarizes advantages and disadvantages of the models and underlines the need for achieving time dependency in dynamic variables like AMH. Models were fitted in original datasets and applied to the crossvalidation datasets. Discriminatory capacity was assessed by calculating c-statistics for models in own data and in the crossvalidation data. Calibration assessed by measuring the slope, intercept and Weibull shape. C-statistics for the SRV model on the SRV data was 0.7 (0.7-0.8), and on TLGS data 0.8 (0.8-0.9). For the TLGS model on the TLGS data 0.9 (0.8-0.9) and in the SRV data 0.7 (0.6-0.8). Calibration of the SRV model on the TLGS data gave a slope, intercept and shape of 1, -0.3 and 1.1 respectively and for the TLGS model on the SRV data of 0.3, 12.7 and 0.6 respectively. Both models discriminate well between women that enter menopause early or late during follow-up. The SRV model showed good agreement between the predicted risk of entering menopause and the observed proportion of women who entered menopause during follow-up (calibration) in the crossvalidation dataset, the TLGS model showed poor calibration.

Introduction

Anti-müllerian hormone (AMH) has emerged as an important biomarker of ovarian ageing. As menopause is the final event that marks depletion of the ovarian reserve, AMH has been suggested to predict the age at which a women will enter menopause (78;79;181;216;247;248). Although the mean age at menopause occurs at around 51 years, a considerable range exists from 40-60 years with approximately 10% of women in the general population becoming menopausal by the age of 45 years (2). Preceding reproductive events such as cycle irregularity in the perimenopause and subfertility are suggested to happen in a fixed temporal fashion with the end of natural fertility occurring approximately 10 years before the final menstrual period. Identifying women with early menopause and thus early subfertility could be used for the primary prevention of infertility by counselling those women to conceive early or to cryopreserve oocytes. Furthermore accurate estimation of time of menopause could identify those at higher risk of cardiovascular disease, osteoporosis, and breast or endometrial cancers due to early or late menopause.
Four models for predicting age at menopause with AMH have been presented in recent literature (78;79;216;247). This study aims to solidify existing evidence that AMH-based predictions of age at natural menopause (ANM) is possible by cross-validating two existing models of menopause prediction from the two most comparable studies (78;79;248). We further attempt to make suggestions towards an optimal predictive model for individualised fertility forecasting based on serum AMH concentration.

Materials and Methods

Patients

Data came from two different studies namely the SRV (Scheffer, van Rooij, de Vet) cohort and the TLGS (Tehran Lipid and Glucose Study) cohort.

The SRV cohort study

The SRV cohort is a combined population from three highly comparable prospective longitudinal studies on ovarian function from the Netherlands that were used for the prediction of menopause in a previous publication from our group (78). It consists of 257 normo-ovulatory women aged between 21 and 46 years. Women were selected if they had regular and predictable cycles, no history of infertility or endocrine disorders, if they had not used contraceptives for at least 3 months, and if they had no

history of ovarian or uterine surgery. In these women baseline (T1) AMH was determined. At follow up (T2), which was approximately 11 years later, cycle status was reassessed by questionnaires in 158 women. From these, 48 women were post-menopausal. The study was approved by the medical ethical review committee and all participants gave written informed consent. For a more detailed description the reader is referred to Broer et al, 2011 (78).

The TLGS cohort study

The TLGS is an ongoing prospective longitudinal study in Tehran that was set up to assess prevalence and risk factors of non-communicable diseases. From the original cohort of 2,412 women, 1,265 women met the eligibility criteria and from these women 266 study participants were randomly selected to model age at menopause using AMH. Eligible women were aged between 20 and 50 years, they had regular and predictable menstrual cycles, they had not taken any form of hormonal contraceptives for at least three months, they were proven fertile defined as having at least one-term pregnancy within 1 year after stopping contraception and they had no history of ovarian or uterine surgery. AMH was measured at baseline (T1) and reproductive status was assessed by questionnaires on average 6 years later (T2). From the 266 original participants, 63 women reached menopause. The study was approved by the institutional review board and all participants gave written informed consent. For a more detailed description the reader is referred to Tehrani et al 2011 (79).

Definitions and endpoints

In the SRV cohort a regular and predictable cycle was defined as a cycle of 19-35 days where consecutive menses were predictable within 7 days. In the TLGS cohort study a regular cycle was defined as a predictable cycle with 21-35 day intervals. The endpoint of this crossvalidation was menopause. Both studies defined menopause as a period of amenorrhea of at least 12 consecutive months, for which no other pathologic or physiologic cause was present. Menopause was assessed at follow up, on average 11 years after inclusion in the SRV cohort and on average 6 years after inclusion in the TLGS cohort.

Predictors

The variables included in the predictive model were age at AMH measurement and serum AMH level. The measurement techniques for AMH were different as described below.

AMH assay

In the SRV cohort, all AMH values were measured in the same laboratory. In one of the three studies AMH was measured using the Active MIS/AMH ELISA from Diagnostics Systems Laboratories (Webster, TX). This system had an interassay and intraassay coefficient of variation of less than 5% at the level of 3.0 ng/ml and less than 11% at the level of 13.0 ng/ml. The limit of detection (LOD) of the assay was 0.026 ng/ml. In the other two studies, contributing to the SRV cohort, AMH was measured with the ultrasensitive immunoenzymometric assay (Immunotech-Coulter, Marseille, France). The limit of detection was 0.05 ng/ml and intra- and interassay coefficients of variation were less than 5% and less than 8%, in the two studies respectively. A correction factor was required to justify comparing and pooling of AMH samples from these three studies. As previously described and successfully applied, serum AMH concentration measured in the Immunotech-Coulter assay was corrected by a factor of 0.5 to be comparable to the DSL values (181) (78).
In the TLGS cohort AMH was measured using the Active MIS/AMH ELISA from Diagnostics Systems Laboratories (DSL-10-14400, Webster, TX). The LOD was 0.006 ng/mL, the intra and interassay coefficients of variation were 5.2 and 9.1 respectively. To compare measures of AMH in the SRV cohort to the TLGS cohort, both AMH values were converted into values representative of the new AMH Generation II assay (AMH Gen II assay) from Immunotech-Coulter by applying a conversion factor. For the SRV cohort a lab-specific conversion factor was determined which amounted to 'DLS x 1.564=Gen II'. For the TLGS cohort we applied a conversion factor of 'DSL x 1.4 – 0.0868= Gen II' as previously described in the literature (7;229).

Predictive models

In the SRV cohort, the predictive model was based on female age and AMH. A Weibull survival model was built with female age on the time axis, with delayed entry at T1 and age-specific percentiles of AMH as a single covariate. Age-specific AMH levels were calculated per participant as follows: first all AMH values were log-transformed, then a flexible spline function was applied to the scatter plot of log(AMH) with age at T1 and with the assumption that there was a normal distribution of residuals around this fitted curve. In a tabular presentation of the model, AMH measurement values, and their corresponding percentiles were shown per age category, as well as the 5^{th}, 25^{th}, 50^{th} and 75^{th} percentile values of

the predicted ANM (78).
The TLGS prediction of ANM was also based on a Weibull survival model with female age on the time axis, and delayed entry at T1. This model contained both Age at T1 and AMH as covariates. The following predictive equation was obtained, where 0.5 is the constant to determine the median ANM and age T1 is the age at inclusion: Age at menopause= $(-\ln(0.5)^{-0.037}\exp(0.10\ AMH + 0.016\ ageT1 + 3.2)$ (248)

Statistical analysis

Data from both studies were converted to contain the same format with regard to variables and coding. Baseline measures were calculated for both studies and compared to the originally published articles. Baselines characteristics were described as means with standard deviation for continuous variables and numbers with percentages for categorical variables. Baseline measures were compared between cohorts using a Mann-Whitney-U test for continuous variables and a chi-squared test for binomial variables. Kaplan Meier curves were created with female age on the time axis and delayed entry at T1, where women were either censored at the end of their follow-up or continued until their age at menopause.
The basic principle in this analysis was to fit the SRV regression model on the SRV data and to then apply this model to the TLGS data. The performance of the prediction rule was studied in terms of discrimination and calibration. Discrimination is quantified by c-statistics and it reflects the capacity of the models to discriminate between women who will enter menopause early or late during follow-up. Calibration reflects the agreement between the predicted risk of entering menopause and the observed proportion of women who entered menopause during follow-up. For a Weibull model, calibration may be assessed through fitting a Weibull model on the validation data using a transformed time axis according to the baseline of the prediction model that is to be validated, and containing the linear predictor of that model, evaluated in the validation data, as a single covariate. A perfect calibration has an intercept of zero, a slope of 1 and the Weibull shape will be equal to 1. Because the models are fitted to their own data (which means that the model will always be perfectly calibrated in its original dataset) the calibration was only done for the model in the crossvalidation data. All analyses were then repeated, but with reversed roles, where the TLGS cohort was used to fit the model and then the performance of that predictive rule was assessed in the SRV data, as described above.

In order to also graphically illustrate differences between the two models, we plotted their Weibull curves for predicted age at menopause for four given ages: 25 years, 30 years, 35 years and 40 years. Within each age category, three AMH values were chosen to represent a low age-specific AMH an average age-specific AMH and a high age-specific AMH. For a 25-year old the selected AMH values were 0.8ng/ml, 3ng/ml and 7 ng/ml. For a 30 year old the selected AMH values were 0.8ng/ml, 2.5ng/ml and 5ng/ml. For a 35 year old the selected AMH values were 0.3 ng/ml, 1.5 ng/ml and 3 ng/ml. Lastly, for a 40 year old the selected AMH values were 0.2 ng/ml, 1 ng/ml and 2 ng/ml. Data were analyzed with SPSS 15.0 (SPSS Inc., Chicago, IL) and R version 2.9.0. (http://www.r-project.org/).

Results

Baseline characteristics of both cohorts are displayed in ***Table 1***. The mean age at inclusion in both cohorts was approximately the same, but the variation in age at inclusion was higher in the TLGS cohort (35.3 years +/-6SD and 37.6 years+/-10 SD for the SRV and TLGS cohorts respectively; p=0.57). This explains how approximately the same proportion of women entered menopause during follow-up (48 women (26%) in SRV versus 63 women (24%) in TLGS) despite a shorter time to follow-up in the TLGS cohort. The Kaplan Meier curves in ***Figure 1*** show the similarity between the two cohorts with regard to the occurrence of menopause during follow up.

Table 1. Baseline characteristics

	SRV	TGLS	P-value
Number of women in cohort	185	266	
Age at AMH measurement (T1)	35.5 (5.9)	37.6 (9.6)	0,57
Mean Age at Follow up (T2)	47.5 (6.11)	42.63 (8.09)	<0.001
AMH (ng/ml)	2.8 (2.8)	4.1 (4.1)	0,02
BMI (T1)	24.0 (4.0)	27.7 (5.0)	<0.001
Number of post-menopausal women at follow up	48 (25.9%)	63 (23.7%)	0.561*

Comparison of baseline characteristics between cohorts. All p-values are calculated with Mann-Whitney-U tests, except those denoted with a * where the p-value is measured with a chi-squared analysis.

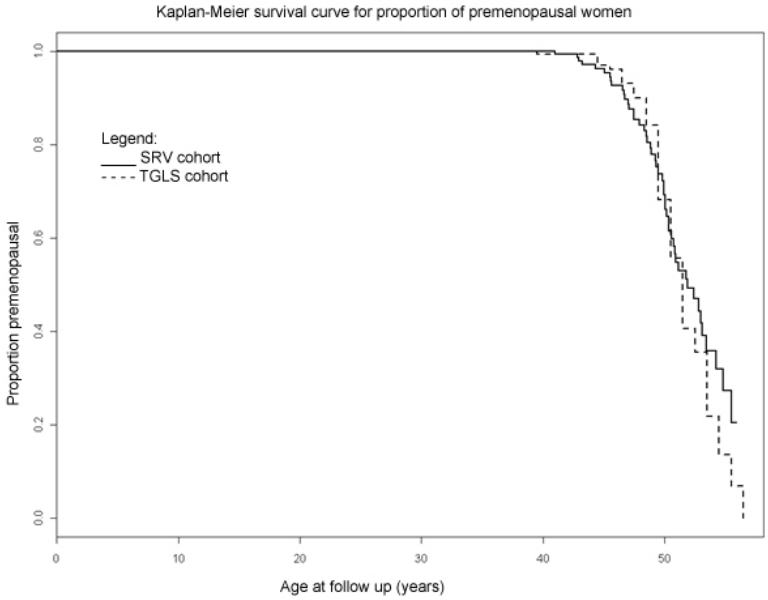


Figure 1. Kaplan Meier curves depicting the decrease in the proportion of premenopausal women with increasing age at follow-up.

The c-statistic for the SRV model in the SRV data was 0.73 (95%CI 0.65-0.82) and for the SRV model in the TLGS data was 0.82 (95%CI 0.75-0.88). When the TLGS model was applied to the TLGS data the c-statistics was 0.88 (95% CI 0.83-0.94), in the SRV data it was 0.72 (95% CI 0.63-0.81) (*Table 2*).

Results of the calibration exercise are displayed in *Table 3*. When the SRV model was calibrated to the TLGS dataset the slope was 0.99, the

Table 2. Accuracy of the models in their original dataset and in the crossvalidation dataset.

	Model applied to:			
	SRV		**TLGS**	
Model sampled from:	C-statistic	95% CI	C-statistic	95% CI
SRV	0,73	(0.65-0.82)	0,82	(0.75-0.88)
TLGS	0,72	(0.63-0.81)	0,88	(0.83-0.94)

This table shows the c-statistics for the two models in their own data and in the crossvalidation data. C-statistics represent the discriminatory capacity of the model, i.e. the accuracy with which the model can distinguish between those women that will enter menopause early during the follow up period and those women that will enter menopause late during the follow up period.

Table 3. Calibration of the models in the crossvalidation dataset.

	Slope (95% CI)	Intercept (95% CI)	Shape (95% CI)
SRV model on TLGS data	0,99 (0.69-1.28)	-0,26 (-0.51- 0.00)	1,13 (1.08-1.39)
TLGS model on SRV data	0,30 (0.14-0.46)	12,66 (12.15-13.18)	0,62 (0.50-0.77)

This table shows the calibration of the models on the crossvalidation data. Calibration consist of the calculation of the slope, intercept and Weibull shape. A perfect calibration has an intercept of zero, a slope of 1 and a Weibull shape of 1. Calibration reflects the agreement between the predicted risk of entering menopause and the observed proportion of women who entered menopause during follow-up.

intercept was -0.26 and the shape 1.13. The slope, which is close to a value of 1, indicates that the predicted risk of entering menopause and the observed proportion of women who entered menopause during follow up are in good agreement. The negative intercept indicates that the SRV model systematically *overestimates* age at menopause in the TLGS data. Conversely, when the TLGS model was calibrated on the SRV data, the slope was 0.30, the intercept was 12.66 and the shape was 0.62. The slope is not close to a value of 1, indicating poor agreement between the predicted risk of entering menopause and the observed proportion of women who entered menopause during follow-up. The intercept is higher than 0, indicating an *underestimation* of ANM, and the shape is also not close to a value of 1.

Figures 2a-2d compare the Weibull curves for the prediction of menopause based on age and AMH in the two predictive models. The solid curves show the predicted ages at menopause for women with a low age-specific AMH, the dotted curves for women with an average age-specific AMH and the dashed curves for women with a high age-specific AMH. The apex of the curves corresponds to the most likely predicted age at menopause with the left and right tails of the curve representing the total range of possible ages at which menopause could occur at a given AMH concentration. The upright Weibull curves represent the predicted age at menopause by the SRV model at the specified AMH values and the upside-down Weibull curves represent predicted age at menopause by the TLGS model at those same specified AMH values. Both models give a Weibull distribution curve that is skewed with a longer left tail. It is evident from these graphs that the predictions according to the SRV model lie closer

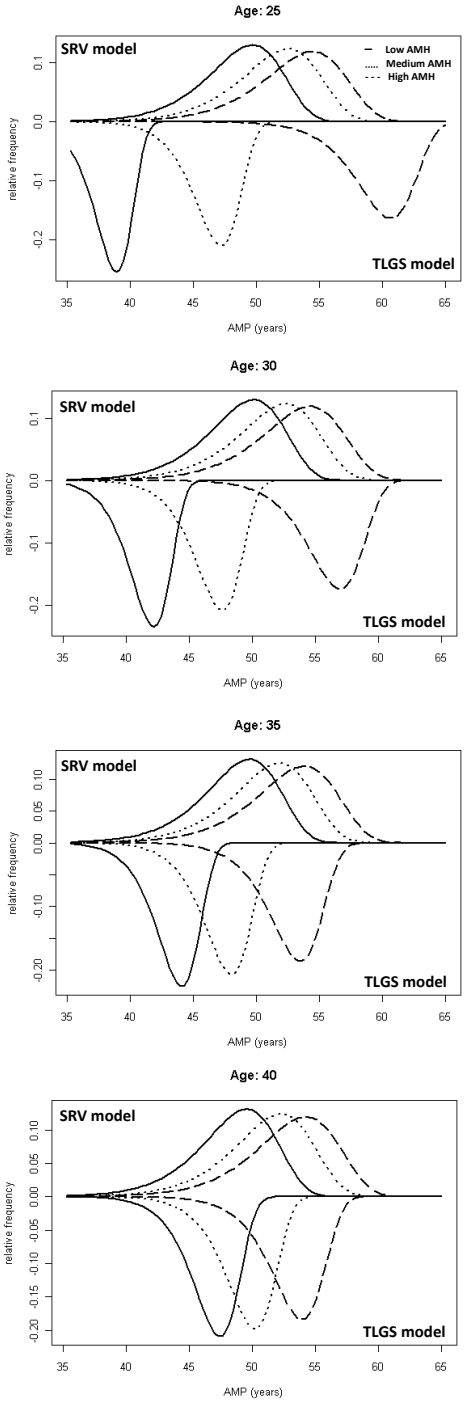


Figure 2. Differential distribution of age at menopause according to age and AMH in the two different models.
This figure shows the distribution of ANM for women in 4 age categories with a low AMH for their age (red curve) an average AMH for their age (green curve) and a high AMH for their age (blue curve). The upright model shows the predicted distributions of ANM in the SRV model and the upside down curve for the predicted distributions according to the same age and same AMH in the TLGS model. The figure displays the increasing conformity of the two models with increasing age.

together and have wide distributions that overlap. Regardless of the age at inclusion, (the age at which AMH was measured), a woman with a low age-specific AMH has a mean ANM of approximately 49 years, a woman with an average age-specific AMH has a mean ANM of approximately 51 years and a woman with a high age-specific AMH has a predicted mean ANM of approximately 54 years.

The Weibull curves for the TLGS model result in predicted ANMs that are spread further apart for women of the same age with different AMH values. The corresponding distributions are narrower than those in the SRV cohort. While a 25-year old with a low age-specific AMH is predicted to reach menopause at 38, a 25 year old with a high age-specific AMH is predicted to reach menopause at 53. Conversely, for a 30 year old the predicted age at menopause is 41 for woman with a low AMH and 49 for a woman with a high AMH. This trend continues so that a 35 year old with a low or a high AMH is predicted to reach menopause at 44 or 50 years respectively and a 40 year old woman with a low or high age-specific AMH has a predicted ANM of 47 and 54 years respectively (***Figures 2a-2d***).

Discussion

This study aimed to crossvalidate and compare two models of menopausal age prediction using female age and AMH. With this study we have provided further evidence that prediction of age at menopause at an individual level is feasible. We have shown that both models are adequately capable of discriminating between those women at risk of having an early menopause and those women with a likely older age at menopause. The discriminatory capacity of all models was above 70%. We further showed that the agreement between the predicted risk of entering menopause and the observed proportion of women who entered menopause during follow-up is better for the SRV model on the TLGS cohort than vice versa. The c-statistics suggest that the TLGS model most accurately discriminates between those women who will enter menopause early or late during menopause. However, a higher c -statistic may be largely explained by a larger amount of variation in the data (e.g., if there are more younger women in the cohort then it is easier to discriminate who will and will not become menopausal during follow up than if all women are of late reproductive age). The TLGS cohort had a larger variation in both the age at which AMH was measured and a larger variation in AMH values (see baselines table).

We have described that in general, the SRV model slightly overestimates age at menopause when applied to the TLGS cohort, and conversely that the TLGS model underestimates age at menopause when applied to the SRV cohort. There are several factors that must be addressed when trying to understand the incongruence of the models on the external data. The median age at menopause, as measured by the Kaplan Meier curves, is approximately the same in the SRV cohort and the TLGS cohort (51.9 is SRV versus 51.5 in TLGS). Therefore, this cannot explain why the one model overestimates ANM and the other underestimates ANM. Furthermore, the Kaplan Meier curves overlap one another reasonably.

There are some discrepancies in the data, however. First of all, from the table with baseline characteristics it is evident that despite the shorter follow up time, and lower age at follow up in the TLGS study, approximately the same proportion of women has become menopausal during the study (26% in TLGS versus 24% in SRV). One would expect this proportion to be smaller. One explanation is that the range of ages at inclusion is much larger in the TLGS cohort with 107 out of the 266 women with an age of 40 or above. Perhaps this discrepancy explains why the SRV model overestimates and the TLGS underestimates age at menopause, the TLGS model expects women to become menopausal sooner during follow up while the SRV model expects women to become menopausal later during follow up. It may also be argued that the relative risk for menopause occurring at a certain age differs between different ethnicities. However, the ***median*** age at menopause discerned from the TLGS cohort was 49.7 years which is comparable to the median age at menopause from large Caucasian cohort studies (49.2 years).

Another issue that remains to be discussed is that the AMH values in the TLGS cohort are higher than in the SRV cohort despite the age at AMH measurement being slightly older in the TLGS cohort. This could be due to differences in lifestyle factors in the cohort, for example smoking has been shown to be an important determinant of AMH (254). In the SRV cohort 16.7% of women smoked, in the TLGS this was 1.2%. Another plausible solution may be that some incongruity in AMH assay system has remained despite application of suitable conversion factors. However, this is probably of more interest to the TLGS model than the SRV model as the SRV model uses percentiles of AMH and not absolute AMH values. A woman's percentile, or rank in a population, does not change after application of a conversion factor, while one's absolute AMH value may

change considerably. Nevertheless, if it were only a problem associated with a conversion factor problem one would expect a systematic incongruence between the observed and predicted ages at menopause, and not a poor performance of all the calibration parameters. The difference in model performance can thus not merely be due to the value of AMH after conversion to the GenII assay, but must also be due to the way in which AMH is incorporated into the predictive model. Both models use age from birth onwards as their time axis with delayed entry of a women at her age at T1. The challenge for AMH is that it is age, and thus follow-up time, dependent. In the SRV model AMH is assigned per woman as a percentile for her age category, not as an absolute value. The time independency is achieved by assuming that a woman will stay in the same percentile over time, which may be considered to be already destined before or at birth. In the TLGS model AMH and age are simultaneously entered as covariates in the model. Both are time (=age) dependent factors. It thus assumes that the combination of age with AMH is a set combination that was already known at birth. The difference in the way in which AMH is incorporated into the models is also evident from the sign of the coefficients of age and AMH in the TLGS model when fitted to the other dataset. In the TLGS model the coefficient for age and AMH are both positive, while you would expect age to be positive (higher chance of menopause at a higher age) and AMH to be negative (lower chance of menopause with a higher AMH). However, in the TLGS model, the way in which AMH is incorporated is such that within a certain age group a higher AMH is associated with a higher age at menopause leading to a positive coefficient for age. In the SRV model, age indeed is positive while AMH is negative. The effect of the different incorporation of age is evident in figures 2a-2d. The SRV model is barely influenced by age but only by age-specific AMH. Regardless of whether a woman is 30 or 40 years old, if she remains in the lower percentiles of AMH her ANM will be lower (around 49) than a woman who remains in the higher percentiles of AMH who will enter menopause at around 54. The resulting predictions of age at menopause widely overlap due to a wide range of predictions. The TLGS model, in contrast, is highly affected by both age and AMH. This leads to mean predictions that are spread wide apart and with narrower ranges. Although the discriminatory capacity of the TLGS model is appealing and it has a higher c-statistic, the question remains whether it is also clinically realistic. In the general population there are very few women that enter menopause before the age of 40 or 45. The TLGS model seems to suggest

otherwise, when looking at their predictions of young reproductive women who would enter menopause at around 43 years with an age-specific AMH level in the average range. With increasing age the models start performing more similarly as shown in Figure 2d. However, the target group, in which predicting age at menopause is most valuable are women that are younger than forty. Therefore, the predicted ages at menopause must also be precise for this group of women. Perhaps the TLGS model extrapolates too extremely for young women. Considering these results it could be suggested that the use of age-specific AMH percentiles yields more realistic prediction of age at menopause than absolute AMH values that are age-corrected in the model where age is already on the time axis. However, it may be more limited in its ability to predict both extreme ends of the spectrum for menopausal age.

The major strength of this study was that the two studies had a very similar set-up with the same variables in the predictive model and with the same endpoint. A challenge of the studies was making the AMH values comparable. To compare the SRV measures of AMH to the TLGS cohort, both AMH values were converted into values representative of the new AMH Generation II assay (AMH Gen II assay) from Immunotech-Coulter by applying a conversion factor. For the SRV cohort an lab-specific conversion factor was determined which amounted to 'DLS x 1.564=Gen II'. For the TLGS cohort the conversion factor of 'DSL x 1.4 – 0.0868 ng/ml= Gen II' as previously described in literature (please not the original conversion in given in pmol/L) (7;237). Determination of a specific conversion factors between the two cohorts used in this study was not possible as the original TLGS serum samples were no longer available. For the same reason, a lab-specific Gen II conversion factor could not be made for the TLGS cohort. Nevertheless, after applying these conversion factors the AMH values were more comparable than before applying the Gen II conversion factor. Although the mean AMH values still differ significantly at baseline, this may also be due to differences in the cohort and not to the AMH conversion factor alone.

Two other prospective studies, both performed in the same cohort of women, have also provided evidence for the notion that AMH can be used to make a more individualised prediction of age at menopause. The cohort consisted of women of later reproductive age (the mean age was 40 years), with a follow up of 14 years. The first study (216) confirmed that AMH is a valuable predictive marker of time to menopause, and the second study elaborates that the rate of change of AMH is a better predictor of time to

menopause than one AMH measure (247). We have confirmed the first finding, but as we only had one measure of AMH we could not study the effect of the rate of change. However, in this study a period of 3 or more years was required to have elapsed between AMH measures before the rate of change could be adequately assessed. Allowing this much time to pass between AMH measures is not practical considering that the value of the test lies in its potential to prevent primary age related infertility in women. More prospective studies with a longer follow-up period and repeated AMH measurements are needed to confirm existing results. Such studies will allow for more precise predictions to be made, also for young women, and will provide smaller confidence intervals.

Conclusions

We have shown that any predictive model must incorporate AMH adequately to make it both age and time dependent, preferably with age-specific AMH values as a time-dependent factor. The current study has provided evidence, in the form of a crossvalidation study, that the prediction of age at menopause with AMH has definite potential. Correct prediction of age at menopause may pave the road to individualised prevention of primary age-related infertility and menopause-related conditions, like cardiovascular disease and breast cancer.

Acknowledgments

Hereby we would like to thank dr. G.J. Scheffer, dr. I.A.J. Van Rooij and dr. A. de Vet for collecting the preliminary data and to all the participants for putting time and effort into this longitudinal study. Thanks to you, research on natural fertility in healthy women is possible.

M. Dólleman (1, 2), W.M.M. Verschuren (3), M.J.C. Eijkemans (2),
F.J.M. Broekmans (1), Y.T. van der Schouw(2)

1. Department of Reproductive Medicine and Gynaecology, University Medical Center Utrecht
2. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
3. National Institute for Public Health and the Environment, Bilthoven, The Netherlands

Manuscript ready for submission

Chapter 8

Anti-Müllerian Hormone & Prediction of Menopause: Results From a Large Prospective Cohort Study

Abstract

Background: Prediction of age at menopause may be useful in predicting end of female fertility. This study aimed to assess the added value of anti-müllerian hormone (AMH) on top of patient characteristics in the prediction of time to menopause (TTM) in a population-based setting.

Methods: 1163 premenopausal women participating in the second follow-up round of the Doetinchem Cohort Study were included. Menopausal status was assessed at follow-up after 5 and 10 years. Multivariable Cox' proportional hazards regression analysis provided associations between potential predictors and TTM. A model without AMH was fit using variables selected through Aikike's information criterion. Performance of the prediction rule was assessed with C-statistics and compared to a model including AMH and one with age only. Added value of AMH was assessed with Net Reclassification Index and change in absolute predicted risk. Performance of these 3 models was compared in subgroups based on age and reproductive characteristics.

Findings: The final model included age, BMI, pack-years of smoking, and menstrual cycle status (regular, irregular, pregnant or taking oral contraceptives). This model had a C-statistic of 0.89, compared to 0.88 of age only. Addition of AMH increased it to 0.91. In a subgroup of 25-43 year olds with regular menstrual cycles, age only had a C-statistic of 0.79 and models without and with AMH of 0.79 and 0.87, respectively. In the entire cohort the risk to enter menopause within 10 years assigned by the model with AMH was on average 3% higher than that assigned by the model without AMH for women who did enter menopause. In the subgroup of young women with regular cycles this increase was 11%.

Interpretation: This study justifies AMH as an additive predictor of TTM. The added value of AMH differs per subgroup of women and is largest in young women with regular menstrual cycles.

Funding: none.

Introduction

Menopause, the only noticeable mark of the end of the female reproductive lifespan, occurs on average at the age of 51 years. However, age at menopause shows considerable individual variation between the ages of 40 and 60 years, with approximately 10% of women becoming menopausal before the age of 45 years. A fixed temporal relationship between the end of natural fertility, and menopause itself, is thought to be present, with the end of natural fertility preceding menopause by approximately 10 years (2;3). This means that a woman with menopause at age 43 may reach the end of their natural fertility at age 3,3 without any noticeable changes such as cycle irregularity. Considering current trends of delaying first child birth, it is conceivable that specifically these women are confronted with infertility and require assisted reproductive techniques (ART) to fulfil their child wish (255) If individualised indications of the remaining fertile life span could be derived from appropriate prediction of age at menopause, steps towards primary prevention of age-related infertility can be made by counselling women towards early child bearing or possibly towards cryopreservation of their oocytes (256).

AMH is a hormone that is secreted solely by ovarian follicles which has consistently been shown to reflect the age-associated depletion of the follicular pool (3;77;105;257;258). As the onset of menopause is incited by exhaustion of the follicle pool and considering that AMH is a reflection of the size of the remaining follicle pool, age-specific AMH values have been used to predict the age at which a woman will become post-menopausal in both retrospective and prospective cohort studies (78;79;181;216;248;249). Lifestyle and environmental factors are also determinants of age at menopause, such as smoking, BMI, oral contraceptive use (196-198). Although studies on AMH unanimously agree that AMH is associated with the timing of menopause, it has never been studied what the *true added value* is of AMH on top of easily obtained information such as female age, and other environmental and lifestyle determinants of menopause. Furthermore, their retrospective design, small cohort sizes or short duration of follow-up, and selected populations of proven fertile women warrant substantiation in a large long-term population-based prospective cohort study. The aim of this study is to assess the added value of AMH in the prediction of time to menopause in a large population-based prospective cohort study.

Materials and Methods

Participants

We used data from the Doetinchem Cohort Study (DCS), an ongoing multipurpose prospective study, initially carried out in a random general population sample of men and women aged 20–59 years (1987–1991) in Doetinchem, the Netherlands(200). Over 95% of the DCS population are Caucasian (199). The aim of the DCS was to study the impact of (changes in) lifestyle factors and biological risk factors on various aspects of health, such as the incidence of chronic diseases, physical and cognitive functioning, and quality of life. The cohort is re-examined every five years with questionnaires and a physical examination at the local health service. Three follow-up examination rounds were completed during 1993–1997, 1998–2002 and 2003–2007. All participants gave written informed consent, and the study was approved according to the guidelines of the Helsinki Declaration by the Medical Ethics Committee of the Netherlands Organization of Applied Scientific Research. Details on the DCS have been extensively described elsewhere (200). For the current study we applied a prospective design with the second examination round as the baseline, and follow-up for ten years at round four.

For the current study 2075women participating in the second examination round of the DCS (1993-1997) were eligible. Women were excluded if they were post-menopausal at the start of the study (n=59), if they had undergone hysterectomy or (uni-or bilateral) oophorectomy (n=196), if information on their reproductive status (n=5) or AMH was missing at baseline (n=41), or if they did not participate in the third and fourth examination round (n=611), leaving 1163 women for analysis.

Outcome

Natural menopause was defined according to the World Health Organization as amenorrhoea for at least 12 consecutive months without other obvious reasons (hysterectomy and/or unilateral or bilateral ovariectomy) from reproductive history questionnaires.

Candidate Predictors

All variables pertaining to patient characteristics and laboratory measures recorded in the Doetinchem Cohort Study were critically reviewed for their potential relationship with menopause using up to date literature and clinical expertise. The following characteristics were considered to be possibly predictive for menopause: AMH, age at inclusion, (pack years of) smoking, BMI, socioeconomic status (SES), age at menarche, parity,

menstrual cycle status (whether the female had a regular cycle, irregular cycle, was pregnant or taking HRT or oral contraceptives), and duration of oral contraceptive (OC) use.

At baseline, blood samples for AMH were collected on a random day of the menstrual or OC cycle. Serum was frozen on the day of vena cubiti puncture and stored in liquid nitrogen for future analysis. Prior to the AMH measurement each sample went through one thaw-freeze cycle on ice for 4 hours; intermitted storage until AMH measurement was at -80° for a maximum of 4 weeks. Serum AMH was measured with the AMH Gen-II ELISA (Beckman-Coulter, Sinsheim, Germany) in a single laboratory, by the same experienced lab technician. The precision of assay results was validated with linearity-of-dilution assessment. The limit of detection for this assay is 0.08 ng/ml, and the limit of quantification is 0.16 ng/ml. The inter-assay and intra-assay coefficients of variation were 3.35% and 4.0% respectively.

Reproductive history was assessed via extensive questionnaires at all examination rounds. The questionnaire included questions on age at first menstrual period, period regularity and length, the number of menstruations in the 12 months prior to questionnaire, date of the last menstruation, current pregnancy, and parity. Additionally, women were asked about current or previous OC use, hormone replacement therapy (HRT), and duration of use. Furthermore the occurrence of, and age at, any gynaecological operations was recorded. A regular cycle was defined as having a regular cycle with a mean cycle length of 24-36 days. The number of years of OC use was stratified into 0 years, <1 year, 1-5, 5-10, 10-15, 15-20 and >20 years.

Body weight and height were measured by trained staff. Body weight was measured to the nearest 100 g on calibrated scales with participants wearing light indoor clothing without shoes, with emptied pockets (200). Ever smokers were identified based on the question 'did you ever smoke regularly'. For ever smokers, information on age at which the respondent started smoking, as well as the total number of years of smoking and average amount of cigarettes smoked was assessed, followed by a question on current smoking ('do you smoke at present'). Pack years of smoking were calculated and divided into 7 strata with 5 year-intervals. Socioeconomic status (SES) was classified into four categories according to the highest level of education that a woman had completed: primary school (SES level 1), lower secondary or vocational school (SES level 2), intermediate vocational or higher secondary school (SES level 3) and higher vocational or university (SES level 4)(200).

Data analyses

Time to menopause was calculated as the time between inclusion and menopause (defined as the absence of menstruations for 12 consecutive months). Cox proportional hazards analysis was used to estimate the univariable and multivariable hazard ratios (HR) with 95% CIs for associations between predictors and time to menopause. For women with induced menopause or hormone replacement therapy time to menopause was censored at the time of the last menses before menopause inducing treatment (surgical or medical) or hormone replacement therapy. Women who remained premenopausal were censored at the time of the most recent follow-up interview. The shape of the association for continuous factors was analysed with restricted cubic splines with three knots to identify those candidate predictors that would need to be added to the model with a simple spline transformation. A total of 1163 women were available for analysis. At follow up, 70 women recalled an age at menopause that was slightly younger than their age at baseline (mean -2.5 years, SD 2.7), while their questionnaires at baseline indicated that they were premenopausal. For these women a random number between 0 and 1 year was generated and entered as their time to menopause.

With time to menopause as the main dependent variable, age is expected to be a strong predictor (i.e. a 45 year old female is more likely to enter menopause during a 10 year follow-up period than a 25 year old female), accordingly age remained in all models. A univariable model with age, transformed with a restricted cubic spline, was fit first. Next, a multivariable model including all candidate predictors (apart from AMH) with appropriate transformations was fit. In this model, the number of candidate predictors in the model was reduced with a backwards selection procedure based on Aikike's Information Criterion (AIC), corresponding to a p-value of 0.157 for predictors with one regression coefficient. The regression coefficients in the final model were adjusted with a shrinkage factor which was estimated with bootstrapping. These coefficients were then transformed to hazard ratios (HR) with 95% confidence intervals (95%CI). Interaction terms between candidate predictors were assessed but not found. In a third step, AMH was added as an extra candidate predictor to this model. The predictive values of the models were assessed and compared with Harrell's C-statistic for time-to-event data (259). The C-statistic indicates how well the model discriminates between women who enter menopause early or late during a 10 year follow up period.

To assess the added clinical value of a model with AMH compared to a

model without AMH or a model based on age alone, a Net Reclassification Index (NRI) was calculated. An NRI quantifies the improvement offered by new markers by examining the extent to which a new marker reclassifies subjects at a higher or lower risk of having an event during follow up (251). A continuous NRI (cNRI) was chosen as no established risk categories for the occurrence of menopause exist. The cNRI counts the number of individuals that change upwards and downwards instead of counting the percentage that crosses a particular risk threshold. Each patient is counted as +1 or −1 depending on whether the change in calculated risk was in the correct direction (higher for those with events, lower for those without events) (252)**.** The NRI is the sum of the "event NRI" and the "non-event NRI", where the event NRI is the net proportion of patients who did experience menopause during a ten year follow up who had an increase in calculated risk and the non-event NRI is the proportion of women without menopause who had a decrease in calculated risk. The maximum possible cNRI is 200% as, theoretically, all women with an event and all without an event can be reclassified in the correct direction. For ease of interpretation we also reported the average of the two net percentages. In addition, we calculated the difference between the estimated probabilities of the models with and without AMH, and present the mean of these differences for women who did and did not experience menopause during follow-up. The above methods were all applied to the entire cohort. Subsequently, performance of the model that was fitted in the entire cohort was assessed in clinically relevant subgroups. The following subgroups were constructed by sequentially removing in an extra group of women:

- Subgroup 1 (n=763): All women aged 20-43 years (excluding women aged >43 years)
- Subgroup 2 (n=704): Women aged 20-43 years with regular cycles or taking OC or with pregnancy at baseline (additionally excluding women taking HRT or with irregular cycles)
- Subgroup 3 (n=677): Women aged 20-43 years with regular cycles or taking OC (additionally excluding women pregnant at baseline)
- Subgroup 4 (n=390): Women aged 20-43 years with regular cycles (additionally excluding taking OC at baseline from subgroup 3)
- Subgroup 5 (n=287): Women aged 20-43 years who were taking OC at baseline (additionally excluding those with a regular cycle at baseline from subgroup 3)

Data was analysed with SPSS version 20.0 (Inc., Chicago, Il, USA) and with R version 2.13 (http://www.r-project.org/).

Results

Baseline characteristics for our total study population are displayed in ***Table 1***. At the time of inclusion, 15 were taking HRT, 161 had an irregular cycle, 27 were pregnant, 345 were taking OC and 615 had a regular cycle. From these women, 169 had become post-menopausal within the first five year and 527 within the ten year period.

In the entire study population, the model with age alone required age to be transformed with a spline as the hazard of menopause increased in a linear fashion up to the age of 38 years after which the slope decreased slightly (***Figure 1B***). The C-statistic of this model was 0.88 (SD 0.01). When all candidate predictors apart from AMH were introduced, backward selection according to AIC resulted in a model containing the following 5 predictors: age, BMI, pack years of smoking, and menstrual cycle status. The C-statistic of this optimal model without AMH was 0.89 (SD 0.01). When AMH was added to this model the HR of AMH was significant ($p <= 0.0001$) and the C-statistic increased to 0.91 (SD 0.03). C-statics are displayed in ***Table 2***, the HRs for each candidate predictor are listed in ***Table 3*** and the shape of the associations displayed in ***Figures 1 A-D***.

Overall, the model with AMH in addition to other predictors correctly reclassified an extra 60.4% of women who did become post-menopausal during follow-up to a higher risk category (event NRI) and correctly reclassified an extra 12.7% of women who did not become post-menopausal to a lower risk level (non-event NRI) in comparison to a model with without AMH. This corresponds to an average improvement of 36.5% (***Table 4***). The mean difference in predicted probabilities between the model with and without AMH was plus 2.5% for women who reached menopause and minus 1.6% in women who did not (***Table 4***).

In the subgroups, excluding women over 43 years of age, age did not have to be transformed with a spline, whereas AMH, did, as the hazard of menopause during follow up decreased until an AMH of 2 ng/ml and then levelled off (***Figure 1A and Table 3***). In each subgroup, the HR of AMH when added to the model was significant (p-value for all <0.0001). Excluding women over 43 years at baseline (subgroup 1) resulted in lower discriminatory accuracy of the model of age alone in (C-statistic 0.84, SD 0.03), whereas additionally excluding women with irregular cycles

Table 1. Baseline Characteristics

Characteristic	Mean (SD) or N (%)
Whole group n=	1163
Age at start of follow up (yrs)	40.8 (7.0)
Age at end of follow up (yrs)	48.0 (5.1)
AMH (ng/ml)	1.1 (1.5)
BMI (kg/m²)	23.8 (3.9)
Menstrual Cycle Status at start follow-up	
Regular	615 (52.9%)
Irregular	161 (13.8%)
OC	345 (29.7%)
Pregnant	27 (2.3%)
HRT	15 (1.3%)
Age at Menarche (yrs)	13.2 (1.4)
Parity (n)	1.8 (1.1)
Years of OC use	
Never	72 (6.2%)
<1 year	52 (4.5%)
1-5 years	230 (19.8%)
5-10 years	349 (30.0%)
10-15 years	253 (21.8%)
15-20 years	149 (12.8%)
>20 years	56 (4.8%)
missing	2 (1.7%)
Smoking	
Packyears of Smoking	7.53 (9.32)
Number of current smokers at start follow-up	585 (49.1%)
Socioeconomic Status	
1 (low)	48 (4.1%)
2	571 (49.1%)
3	316 (27.2%)
4 (high)	228 (19.6%)

Table 2. C-statistics per model in the overall study population and per subgroup

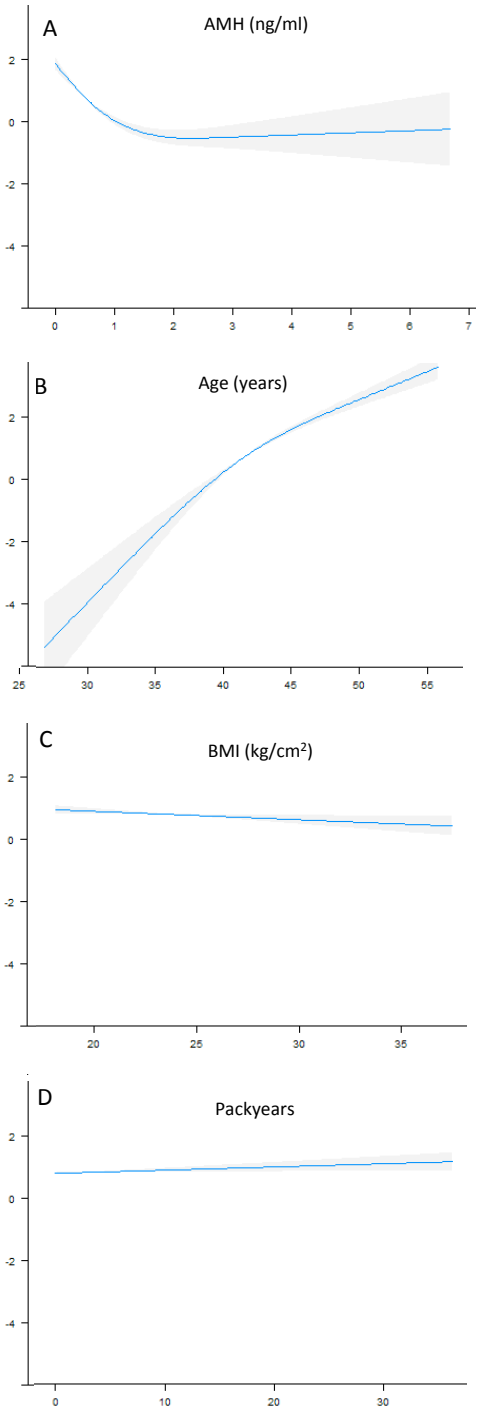
	Whole study			Subgroup1			Subgroup 2		
	C -Stat	**SD**	**n**	**C -Stat**	**SD**	**n**	**C -Stat**	**SD**	**n**
Age only	0,88	0,012		0,84	0,026		0,84	0,027	
Model without AMH*	0,89	0,011	1163	0,85	0,025	763	0,85	0,027	704
Model with AMH**	0,91	0,025		0,89	0,021		0,89	0,021	

Table 2. Continued

	Subgroup 3			Subgroup 4			Sugroup 5		
	C -Stat	**SD**	**n**	**C -Stat**	**SD**	**n**	**C -Stat**	**SD**	**n**
Age only	0,84	0,027		0,79	0,037		0,89	0,05	
Model without AMH*	0,85	0,027	677	0,79	0,038	390	0,90	0,04	287
Model with AMH**	0,89	0,022		0,87	0,03		0,91	0,04	

C-statistics per model and per group are shown. Abbreviations: reg=regular cycle, irreg=irreg cycle, preg=pregnant, OC=oral contraceptices.
*Model with age, BMI, packyears of smoking and menstrual cycle status. **Model with additional AMH.
SG1: All women 20-43yrs;
SG2: 20-43ys with regular cycle, taking OC or pregnant;
SG3: 20-43yrs with regular cycles or taking OC
SG4: 20-43yrs with regular cycles;
SG5: 20-43yrs taking OC

(subgroup 2) and pregnant women (subgroup 4) did not change the C-statistic further. The full model with AMH had a C-statistic of 0.85 (SD 0.03-0.04) in subgroups 1, 2 and 3. Adding AMH increased it to 0.89 (0.02 SD). When additionally excluding OC-users (subgroup 4), the model with age had a C-statistic of 0.79 (SD 0.04), and the models without and with AMH had C-statistics of 0.79 (SD 0.04 SD) and 0.87 (SD 0.03), respectively. In subgroup 5, restricted to OC-users, C-statistics were 0.89 (SD 0.05) for the model with age, 0.90 (SD 0.04) for the full model, and 0.91 (SD 0.04) for the full model plus AMH (***Table 2***).
In subgroups 1, 2, 3, 4, and 5 the mean NRIs were 34.9%, 33.8%, 33.5% , 42.7% and 8.2% respectively. In each subgroup analysis the event NRI was larger than the non-event NRI (***Table 4***), meaning that addition of AMH in younger women helped to increase the predicted risk for those who experienced menopause to a larger degree than it decreased the



Figures 1A-D.
Hazard of time to menopause for the parameters included in the model.
A) AMH
B) Age
C) BMI
D) Packyears of smoking.

Table 3. Hazard ratios per predictor in models with and without AMH

	Whole group n=1163			Subgroup1 n=776			Subgroup 2 n=716		
MODELS without AMH	**HR**	**95%CI**	**p=**	**HR**	**95%CI**	**p=**	**HR**	**95%CI**	**p=**
Age (yrs)	See figure 1A		***	1,53	1.46-1.60	***	1,51	1.44-1.58	***
BMI (kg/m^2)	0,97	0.95-0.99	*	0,99	0.95-1.03	ns	0,99	0.95-1.04	ns
Pack years	1,01	1.01-1.02	**	1,03	1.01-1.04	***	1,03	1.01-1.04	***
Menstrual cycle									
Regularly cycling	Reference category			Reference category			Reference category		
Irregularly cycling	2,36	2.12-2.60	***	2,65	2.14-3.16	***			
Pregnant	1,81	0.67-2.96	ns	1,93	0.77-3.08	ns	1,91	0.75-3.07	ns
Currently taking OC	0,87	0.63-1.11	ns	1,03	0.66-1.40	ns	1,02	0.65-1.39	ns
Currently taking HRT	1,63	1.03-2.22	ns						

	Whole group n=1163			Subgroup1 n=776			Subgroup 2 n=716		
MODELS with AMH	**HR**	**95%CI**	**p=**	**HR**	**95%CI**	**p=**	**HR**	**95%CI**	
Age (yrs)	See figure 1A		***	1,38	1.31-1.45	***	1,36	1.28-1.43	***
AMH (ng/ml)	See figure 1B		***	See figure 1B		***	See figure 1B		***
BMI (kg/m^2)	0,97	0.95-1.00	*	0,98	0.94-1.02	ns	0,98	0.94-1.02	ns
Pack years	1,01	1.00-1.02	*	1,02	1.00-1.03	*	1,02	1.00-1.03	*
Menstrual cycle									
Regularly cycling	Reference category			Reference category			Reference category		
Irregularly cycling	1,91	1.67-2.15	***	2,15	1.63-2.67	**			
Pregnant	1,24	0.09-2.39	ns	1,21	0.05-2.38	ns	1,18	0.02-2.35	ns
Currently taking OC	0,53	0.28-0.77	***	0,63	0.25-1.01	*	0,62	0.23-1.00	*
Currently taking HRT	1,24	0.65-1.83	ns						

Table 3. Continued

	Subgroup 3 n= 687			Subgroup 4 n=396			Subgroup 5 n=291		
MODELS without AMH	**HR**	**95%CI**	**p=**	**HR**	**95%CI**	**p=**	**HR**	**> 95% CI**	**p=**
Age (yrs)	1,52	1.44-1.59	***	1,45	1.37-1.54	***	1,67	1.53-1.81	***
BMI (kg/m^2)	0,99	0.95-1.04	ns	0,99	0.94-1.04	ns	0,97	0.84-1.09	ns
Pack years	1,03	1.01-1.04	***	1,02	1.00-1.04	*	1,05	1.02-1.07	**
Menstrual cycle									
Regularly cycling	Reference category			Reference category			Reference category		
Irregularly cycling									
Pregnant									
Currently taking OC	1,03	0.66-1.40	ns						
Currently taking HRT									

	Subgroup 3 n= 687			Subgroup 4 n=396			Subgroup 4 n=291		
MODELS with AMH	**HR**	**95%CI**	**p=**	**HR**	**95%CI**	**p=**	**HR**	**> 95% CI**	**p=**
Age (yrs)	1,36	1.29-1.44	***	1,29	1.21-1.38	***	1,56	1.40-1.71	
AMH (ng/ml)	See figure 1B		***	See figure 1B		***	See figure 1B		***
BMI (kg/m^2)	0,98	0.94-1.02	ns	0,97	0.92-1.02	ns	0,98	0.86-1.11	ns
Pack years	1,02	1.00-1.03	*	1,01	0.99-1.03	ns	1,04	1.04-1.06	*
Menstrual cycle									
Regularly cycling	Reference category			Reference category			Reference category		
Irregularly cycling									
Pregnant									
Currently taking OC	0,63	0.24-1.01	*						
Currently taking HRT									

Legend : ***= p<0.0001, **= P<0.001, *=p<0.05, ns= not significant

Table 4. Continuous NRIs and the change in predicted risk of menopause for events (women with menopause at follow up) and non-events (women without menopause at follow up) when AMH is added to the model.in the overall study population and per subgroup.

	Event NRI	Non-event NRI	Total NRI	Mean NRI	Δ predicted risk of MP in events (% ± SD)	Δ predicted risk of MP in non-events (% ± SD)
Overall population	60,4%	12,7%	73,1%	36,5%	2.5% (3.7)	-1.7% (1.9)
Subgroup 1	44,1%	25,8%	69,9%	35,0%	6.8% (6.4)	-1.0% (2.7)
Subgroup 2	45,2%	22,4%	67,6%	33,8%	7.2% (5.9)	-0.3% (4.2)
Subgroup 3	45,5%	21,5%	67,0%	33,5%	6.8% (6.3)	-0.3% (4.0)
Subgroup 4	37,7%	18,0%	55,6%	27,8%	11.1% (7.9)	-1.6% (6.8)
Subgroup 5	10,5%	5,8%	16,3%	8,2%	2.6% (6.2)	-2.9% (5.6)

Continuous NRI's per model and per group are shown as well as the change in the predicted risk of menopause for women with and event (menopausal (MP) at follow up) and non-events (not-menopausal at follow up).
SG1: All women 20-43yrs
SG2: 20-43ys with regular cycle, taking OC or pregnant
SG3: 20-43 yrs with regular cycles or taking OC
SG4: 20-43 yrs with regular cycles; SG 5: 20-43yrs taking OC
SG5: 20-43yrs taking OC

predicted risk for those who did not experience menopause during follow up. In subgroups 1, 2, 3 and 4 adding AMH to the full model increased the mean predicted risks for women postmenopausal at follow-up by 6.8%, 7.2%, 6.8% and 11.1%, respectively. For women remaining premenopausal the model with AMH predicted risks that were on average 1%, 0.3%, 0.3% and 1.6% lower that the model without AMH. In subgroup 5 predicted risks were 2.6% higher and 2.9% lower in the models with AMH for women with and without menopause, respectively.

Discussion

Main findings

With this study we have shown that from all relevant lifestyle and reproductive predictors, age and AMH contribute most to the prediction of TTM. Age was the strongest predictor of TTM, which is a very logical finding as older women have a shorter TTM than younger women. Most importantly though, we have provided evidence that AMH not only has additive predictive value for this prediction, even when taking BMI and smoking into account.

The additive value of AMH was most prominent in the group of young women with regular menstrual cycles (subgroup 4). In the group as a whole and in young women taking OCs (subgroup 5), the added effect of AMH was marginal and in subgroups 1, 2 and 3 (young women with/without; regular cycles, OC use or pregnancy) the effect was also smaller in comparison to young, regularly cycling women (subgroup 4). This can be interpreted as a reflection of the fact that the study population as a whole consists for two thirds (n=779) of women of 45 years and above, who have almost 100% probability of entering menopause in the next ten years, leaving little room for improvement by AMH. In the subgroups of women aged 20-43 it is likely that the predictive effect of AMH was influenced by the presence of determinants such as pregnancy, cycle irregularity and oral contraceptive. In these women AMH may not be a pure reflection of ovarian reserve but more a reflection of the amount of suppression induced by the determinants(254).

Clinical value

The accuracy with which the models including AMH could discriminate between women who do and do not enter menopause during follow up was excellent, with an accuracy approaching 90% and small standard

deviations. These C-statistics were similar to those reported in the study of Broer et al. 2012, in which a model with age alone had a C-statistic of 87% and addition of AMH raised this to 90%. The added clinical value as measured by the continuous NRI was considerable with an average of 34-43% improvement added by AMH. Notably, however, this reflects only a marginal improvement of 3-8% as measured in the C-statistics. In terms of change in the risk assigned to women, addition of AMH to the model resulted in the assignment of an average increased risk of menopause by 2.5-11.1% in women who did enter menopause at follow-up and a mean decreased risk assignment of 0.3-2.9% in women who did not enter menopause compared to the model without AMH. The additive effect of AMH was largest for young women with regular cycles, for whom the prediction of time to menopause is ultimately the most interesting as these women will benefit from knowing how many years of their fertile life still remain. This is further exemplified through assessment of the mean increased risk assigned to individuals who enter menopause during follow up by addition of AMH to the predictive model. Whereas for the whole population a mere 2.5% increase in risk is offered by addition of AMH, estimated risk increases by 7-11% in younger women aged 25-43 years (***Table 5***). With information on time to menopause, women with a high chance of an early onset of menopause could be counselled towards not delaying pregnancy to a high age or towards cryopreservation of oocytes. On the individual level, fertility prediction may have implications on the fulfilment of child wish and on the societal level towards creating large enough families for population maintenance and reducing expenditure for ART. A nomogram for menopause prediction using AMH has been proposed, however the added value on top of patient characteristics was not studied and predictions had large 95% CIs making clinical application risky (248). In our view, our findings have to be externally validated before we can proceed to develop nomograms or clinical prediction rules.

Strengths and weaknesses

Several studies have looked at AMH as a predictor of menopause in addition to age (78;216;248;249). However, none have assessed the additive value of AMH on top of readily available patient characteristics with methods in accordance with the current state of art, even though a recently study suggested that smoking and BMI may improve AMH-based prediction (260). A main strength of our study was the population-based design of the study, enabling the study of different subgroups based on

Table 5.

	Menopause at follow up	Mean difference in risk per woman between model with and without AMH	SD
Whole Group	Yes	2,5%	3,7%
	No	-1,6%	1,9%
Subgroup 1	Yes	6,8%	6,4%
	No	-1,0%	2,7%
Subgroup 2	Yes	7,2%	5,9%
	No	-0,3%	4,2%
Subgroup 3	Yes	6,8%	6,3%
	No	-0,3%	4,0%
Subgroup 4	Yes	11,1%	7,9%
	No	-1,6%	6,8%
Subgroup 5	Yes	2,6%	6,2%
	No	-2,9%	5,6%

The mean difference in predicted risk of menopause per woman at follow up are displayed for women who did and did not have menopause at follow up. A model with AMH assigns a higher risk to women with observed menopause at follow up and a lower risk to women who were premenopausal at follow up than a model without AMH.

reproductive characteristics. Also, the size of the population in which other potential determinants of TTM could be assessed next to AMH was much larger than of previous studies. Another strength of the study is that a considerable amount of young women were included for whom the prediction of TTM is the most valuable. This study would have been even stronger if the duration of follow-up was longer so that more of these young women had become menopausal during follow-up as this would have made the predictions for these young women more precise.

Recently, the stability of serum AMH measures have been questioned (224), especially when AMH is stored at room temperature or -20°C. However, it is clear that when appropriate sample processing is done that values are both reproducible, stable and reliable (261). A strength of this study is that AMH Gen II assay (Beckman Coulter Ltd) was used which is currently the most reliable assay of AMH. Furthermore samples were determined by a single experience laboratory technician and assay result precision was validated using linearity of dilution assessment. These are all factors that support homogenous specimen sampling and the provision of both reproducible and reliable AMH measures.

Concluding remarks

This study has, with up to date statistical methods, justified AMH as an additive predictor of both time to menopause and the occurrence of menopause on top of female age and other reproductive and lifestyle factors. However, the added value differs per subgroup of women and is largest in women who are young when AMH is measured and who are still regularly cycling.

Chapter 9

General discussion

Introduction

Female reproductive longevity is the result of a complex interplay between quantitative and qualitative aspects of the gradual disappearance of follicles from the ovaries with increasing female age. Age alone, however, cannot explain the large variability in the extent of ovarian senescence observed between women of the same chronological age. After menarche, in general, the gradual changes in ovarian function across a woman's life are marked by 4 reproductive milestones, namely decreasing natural fertility, sterility, menopausal transition and menopause. These milestones occur consequentially, and there is evidence for fixed time periods between them. The first three steps often go largely unnoticed with menopause being the only obvious hallmark of the definite end of the fertile lifespan. If we would be able to reliably predict on an individual basis at which age menopause is going to occur, this information may be used to derive until which age a women will sustain normal fertility and at which age her natural fertility will have become futile.

The increasing proportion of women delaying the birth of their first child results in an increased demand for assisted reproductive treatment (ART). As the application of ART will not be successful for almost 50% of the couples that rely on this treatment modality, it is becoming more and more interesting to look for ways to identify women at risk of early decrease in their natural fertility, thereby opening ways for preventive management.

Individual fertility forecasting, through the use of reproductive biomarkers such as anti-müllerian hormone (AMH), may be useful in predicting the outcome of ART. Furthermore, by making individual predictions of reproductive longevity, it may become possible to identify women at risk of early menopause at stages of their lives at which measures can be taken to prevent future infertility. Such measures could for instance comprise early family building or oocyte banking. At present AMH is the reproductive biomarker with the most potential for making such predictions, though some challenges still exist with AMH as a reproductive biomarker.

This thesis aimed to examine the role of AMH as a reproductive biomarker in both fertile and subfertile women. It did so by focusing on different aspects of fertility forecasting, and the original studies presented in this thesis center around 3 important aims:

1. To examine factors that influence age-specific AMH values in the general population.
2. To investigate the role for AMH in predicting response to, and pregnancy, after ART.
3. To examine the role of AMH in predicting age at natural menopause.

In this thesis we showed that one measure of AMH is not generalizable to all women of the same age but that it is significantly influenced by both reproductive and lifestyle factors. We have shown that AMH can adequately predict response to IVF, but insufficiently predicts the occurrence of an ongoing pregnancy in women with a poor response to IVF treatment. And lastly, we have presented evidence that AMH has additive value in the prediction of female age at menopause in both subfertile women and in the general population, but that the accuracy with which predictions are made depend on the subgroup of women in which AMH is determined.
The ultimate question that remains to be answered at the end of this thesis is: *"Is AMH truly a forecaster of current and future fertility?"* The answer here would have to be: *"Yes, AMH has definite potential as a fertility forecaster, both for short term and long term management options."* However, AMH is still an experimental marker and current application should only be performed in the context of reliable clinical research as important challenges remain to be addressed with regard to its clinical utility.

AMH measures in the general population

AMH has long been thought to be a very stable and generalizable measure of ovarian reserve as early reports showed low inter and intra-cycle variability, as well as stability throughout the menstrual cycle (237;262-264). Furthermore it was reported that AMH levels were independent of life-style factors or menstrual cycle suppression through oral contraceptive use or pregnancy, however, recent reports show considerable intra-individual fluctuation of AMH, especially in women below the age of 25 years, and significant suppression of AMH through life-style and reproductive factors (184;185;185-188;192-195;229). Lastly, the assay seems to not yet be fully developed to always yield stable results that are comparable between laboratories (224;265). This is one of the main reasons why measuring AMH for the moment should be research-bound, until a stable assaying system has been implemented.

Another reason why AMH measurements must be performed solely in the context of – preferably long term follow up - research is that we do not exactly know what is a 'normal' AMH level per age-category of women. In chapter 4 we have provided undeniable evidence for the influence of AMH by common lifestyle and reproductive determinants. Interestingly, the suppressive effect of factors such as oral contraceptive (OC) use and smoking seemed to be reversible. As replenishment of the ovarian follicle pool in adult life has mostly been denied this suggests that AMH does not truthfully reflect the ovarian reserve in these women but that it is a transient reflection of the amount of suppression of the pool of growing follicles induced by smoking or OC-use. AMH must be interpreted in the context of the patient and not as the 'holy grail' of reproductive potential when applied to whole populations.
Until now the population in which AMH seems to most adequately reflect ovarian reserve is in non-smoking, non-OC taking women. An extremely high AMH may be associated with polycystic ovary syndrome (PCOS) (177;266). It has further been suggested that women with an extremely low AMH are at risk of early depletion of their ovarian reserve, however no reliable reference values of normal exist as of yet (chapter 5)(78;176) . There are several reasons for this, firstly the assay has undergone many changes (as will be discussed later) and the consequences tied to all the "shades of normal" are not yet known thus hindering us from being able to say what ***is*** normal. Many studies dichotomize outcomes according to high/low AMH, such as a shorter duration of conception in women with AMH levels above 0.7ng/mL (267). The assignment of such cut-offs is arbitrary, static with regard to the influence of age and risky to generalize to other populations. If adequate reference values are created it will help us understand what is normal and how much deviation from the norm results in measurable clinical consequences such as longer duration to pregnancy, subfertility or early menopause.

AMH and assisted reproduction

Currently, the most common application of AMH, as well as for other ovarian reserve markers like the antral follicle count (AFC), in clinics around the world is to predict response to, and outcome after, ovarian hyperstimulation in ART. The idea behind such predictions is that it could be useful in counselling patients on their prognosis, to optimise responses to stimulation dosages in ART and thereby to maximise success

after ART and possibly improve cost efficiency. In fact, AMH is commonly used to determine the starting dosage for IVF treatment, despite the fact that this practice is not yet adequately evidence-based. Chapter 2 has demonstrated that AMH is a valuable predictor of an excessive response to IVF treatment, and previous research by our group has shown that also for prediction of a poor response AMH is useful (162). Excessive response prevention is an important measure in reducing ART-related complications and patient discomfort. In theory, dose adjustments based on AMH predictions of response category would result in less poor and/or excessive responders. However, this is supported by a single study, while other studies fail to demonstrate such effect or are hampered by methodological flaws (135;150;268). Especially in women with a poor response, there is insufficient evidence that increasing the stimulation dose also increases the number of retrieved oocytes or pregnancy rates (269). Ongoing pregnancy in this scenario is a considerably more relevant outcome. A reason why pregnancy rates have been observed to remain the same despite higher dosing (149) may be that the quantitative and qualitative aspects of follicular decline do not occur in a one-to-one ratio. By increasing the absolute number of oocytes retrieved, one does not necessarily increase the proportion of good *quality* oocytes nor embryos (29). In women with a good prognosis in ART, the significant association between AMH and ongoing pregnancy was similarly found to be mediated by the total *number* of oocytes that could be fertilised and not *quality* of these extra oocytes (161). In line with this is the finding that relatively good pregnancy prospects are also found in young women with an unexpected low ovarian reserve as expressed by a poor response to ART (270). Together these results support our theory that the relationship between quantity and quality is highly female age-modulated, where low quantity leads to poor number of good quality oocytes and embryo's. but with a large protective or augmenting effect of female age. This is in line with recent publications and is further supported by the chance categories of pregnancy prediction in chapter 3 where young age and high AMH were both related to increases in pregnancy but the highest pregnancy rates were seen where both AMH was high and age was young. (57;271). These results emphasize that AMH alone can under no circumstance be used to identify null-prognosis patients whom should be advised abstain from fertility treatment. Before deciding whether use of AMH-adjusted stimulation dosing is warranted, the results of the OPTIMIST and the CONSORT trial must be awaited, which will clarify what both the clinical

consequences and the cost-efficacy are of individualised dosing on the basis of ovarian reserve testing (147;268). Until then, clinical use of AMH outside formalized clinical study protocols should be abandoned.

AMH and prediction of menopause

Chapters 5-8 in this thesis present a robust relationship between individual AMH level and age at natural menopause (ANM). In fact, AMH seems to be a better predictor of individual age at menopause that mother's age at natural menopause (chapter 6). Heritability of age at menopause has been recorded to be substantial, with heritability rates varying between 30% and 85% (61) (62) (63;65;246). Nevertheless until now only 2.5-4.1% of natural variability in age at menopause has been explained by involved genetic loci (66). So while, information on mother's age at natural menopause may limit itself to the genetic factors shared by mother and daughter, AMH may reflect the sum total of genetic (maternal and paternal) and environmental influences. Chapter 8 shows that the prediction of ANM is largely guided by individual age. This influences the predictive capacity where the ***added*** value of AMH on top of age is larger in young women than in older women. Interestingly, this is also the most clinically relevant group as these women can use such predictions to guide their decision making with regard to the age at which they plan to start building a family. Both our studies as well as previous studies that describe individual predictions of age at menopause have considerably large 95% confidence intervals (78;248). Therefore, before such predictions can be applied in the clinic, the certainty with which a woman's reproductive lifespan prediction is made, must improve. As a recent genome wide association study identified 13 new genetic loci associated with age at menopause (66), there may even be room for predictive models that take both AMH, environmental and lifestyle factors as well as genetic information into account to make more precise predictions.

With regard to predicting future fertility, another important question that remains to be answered is whether the rate of decline in AMH is the same in all women across all ages. Most studies have performed just one measure of AMH, however, it may be possible that a woman in the 15th percentile of AMH for her age after one measurement may be in the 50th percentile of AMH for her age 5 years later. An early follow up study demonstrated that in many cases the AMH levels over time remain in the same percentile band, suggesting that the decline rates per individual may not be highly different,

However, one recent study shows that the rate of AMH decline can differ between women of the same age and that this rate is an independent predictor of time to menopause after correction for female age, AMH and smoking status (247). Although the definition of rate of change in this study was ambiguous (by basing the AMH rate of change on only two time points one does not take into account inter-individual variation) it does illuminate that the rate of change may not per se be the same for women of similar age with the same AMH. Differing rates of AMH decline may also explain the phenomenon observed in chapter 5 in which we found evidence that the threshold of AMH after which menopause is triggered varies between women. More long term follow up studies are needed which assess the reproductive decline in women over a period of more than 20 years. Such studies would have to have repeated measures of AMH to investigate the consistency across individuals in the rate of AMH decline. Steps are currently being undertaken to realize such a scientific endeavour at our research center.

The AMH assay

The challenges with AMH that remain to be overcome before AMH can be reliably applied in clinics, should comprise the challenge of the AMH assay. Recently, the sample stability and reproducibility of AMH with current assay systems have been called into question (224). Although the claims of AMH instability in stored samples and different assay performance at different concentrations have also been adequately refuted (272), it does underline the importance of AMH measurement by consistent, trained personnel and protocoled sampling and storage of serum. Both chapter 5 (supplementary figure 1) and another study (237) are evidence for the potential robustness of the AMH test and the comparability between centres with similar populations and the same assay. However, now that international consensus has been reached about which assay to use, the next step is to ensure that all laboratories store samples and perform the test under the same conditions. Rustamov et al describe a 2 fold increase in AMH between fresh samples and those stored at room temperature for 48 hours, and a significant difference between samples stored at -20° and -80°C (224). If all samples are processed in the same way and center-specific reference values are constructed this may not be so problematic. Yet, in clinics that compare their fresh samples to reference values from studies with stored samples this can give misleading results. Recent studies

have shown promising results with pre-mixing of samples with assay buffer to potentiate AMH stability at all temperatures (265) or with centrifugation of samples within 5 hours (273). Another option in the future is AMH measurement by automated electrochemiluminesce platforms. Perhaps these measures will solve the current problem. However more research is needed, and in the meanwhile AMH must thus only be performed by experts in the context of clinical studies. Moreover, if a reliable golden standard of AMH measurement is found, all results found thus far might need to be re-established.

Future prospects

Challenges of AMH aside, if we were to theorize that 5 years from now a perfect prediction of age at menopause can be made using female age and AMH, what would we do with this knowledge? With information on time to menopause women with a high chance of an early onset of menopause could be counselled towards trying to conceive at a young age or towards cryopreservation of oocytes. Such primary preventative measures are extremely important as the success rate of ART in subfertile couples is only 50% (14;15). In several countries, cryopreservation of oocytes for women with "anticipated gamete exhaustion" or "social freezing" has become increasingly popular and it has been shown that numerous women have a positive attitude towards such "safeguarding" of their reproductive potential (256). The main reasons for cryopreservation in these women were to create an insurance against infertility, to buy more time to find a partner or to take away the pressure of finding a partner and to have already tried everything to optimize future reproductive chances (256). Although it would be reasonable to defer from Stoop's study that women who would consider social freezing would also want to know their the length of their reproductive lifespan, it would be interesting to perform a study in which women are asked if they would want to know at what age they would become menopausal (and thus subfertile), and if they would change their behaviour according to the results of the menopause prediction.
Initially cryopreservation of oocytes was restricted to women who must undergo gonadotoxic treatment. In 2012 the European Society for Human Reproduction and Embryology task force released a statement in which they concluded that the arguments *against* social freezing were not convincing (274). However they emphasized that more research must be done to prevent women from being deceived that it is a guarantee for a

healthy child. Considering that there are many factors besides oocyte quantity and quality that determine ongoing pregnancy such as uterine factors, sperm factors and comorbidity, the most important message in this statement remained that women who conceive at a young age through natural reproduction have the best chance of a healthy child (274). It could, for example, be very interesting to compare live birth rates, cost-efficiency, and psychological impact between women who have and have not undergone screening and/or treatment in the context of anticipated gamete exhaustion. Another application of menopause prediction may be through the association between reproductive ageing, menopause and cardiovascular disease (CVD). It has been shown that early menopause is associated with increased risks for CVD, osteoporosis, colorectal cancer and cognitive demise (275-277) (278). Therefore, through timely identification of women at risk of early menopause or cardiovascular disease, preventative measures to avoid such menopause related morbidity may improve outcome. Interestingly, a greater ovarian reserve as measured by AMH has been associated with having a healthier cardiometabolic risk factor profile: women with low and medium AMH levels and had a 27% and 31% increase in the number of CVD risk factors after correction for age, smoking, ethnicity, OC-use and parity (279). Perhaps AMH could serve as an independent marker of CVD in premenopausal women if future research focuses on this. Remarkably, age at menopause is not in the most recent Framingham General Cardiovascular Risk Profile Score(280), even though it was previously suggested to make independent risk scores for pre- and post-menopausal women (281). Nevertheless, it would be worthwhile studying the value of AMH in this setting, especially when profiling young pre-menopausal women with regard to their future cardiovascular profile.

In conclusion, while appropriate reference values are being created per age category and until the consequences of having a low of high AMH for one's age are being established, AMH should only be determined in the context of clinical studies. At present, the most important clinical role of AMH at this stage is to serve as a red-flag for reduced ovarian reserve in women of reproductive age who must undergo further diagnostics. This thesis has established a definite role for AMH as a forecaster for both current and future individual fertility. The potential of AMH is definitely in line with our medical paradigm which has been progressively evolving from reactive medicine to personalised, predictive, preventive, and participatory medicine (282). However, for AMH to realize its potential, much research remains to be done.

Chapter 10

English Summary
Nederlandse Samenvatting
References

Summary

The overall aim of this thesis was to investigate the role of Anti-müllerian hormone (AMH) as a predictor of current and future fertility.

Chapter 1 sketches the theoretical background of female reproductive ageing upon which the successive chapter are based. Anti-müllerian hormone (AMH) is a hormone that can be measured in the circulating blood of women. It has been found to be a useful marker of the ovarian reserve, or the number of follicles ("eggs") remaining in a woman's ovaries. In every woman this number of follicles declines with increasing age. However, the age at which the reserve is exhausted leading first to subfertility then to infertility and last to menopause differs a lot between women even if they have the same age. Because the level of AMH may be related to the age at which a women becomes menopausal this hormone may also be used to predict the age until which a woman remains fertile. Furthermore, through its reflection of how many follicles a woman may still have AMH may be useful in predicting how women will respond to assisted reproductive treatment (ART) such as in vitro fertilisation (IVF). The response to ART and the prediction of menopause is what "fertility forecasting" refers to in this thesis.

Chapter 2 presents the results from an individual patient data meta-analysis, in which the results from 33 original studies were pooled so that data from 6,825 patients could be included in the analysis of whether an excessive response to IVF treatment can be predicted using ovarian reserve tests such as AMH. In IVF the goal is to retrieve between 4-15 oocytes ("eggs"), partly because an excessive response (>15 oocytes) leads to discomfort and potentially dangerous complications for patients. This chapter showed that both AMH and another ovarian reserve marker AFC (antral follicle count) add value to female age in the prediction of an excessive response in IVF. We further investigated whether patient characteristics such as age, body mass index and the duration of subfertility influenced the predictive performance of AMH and AFC and we found that the accuracy of AMH was not influenced by patient characteristics but that the accuracy of AFC was influenced by female age.

Chapter 3 is a study which used the same dataset as in Chapter 2 but selected only those 430 women with a poor response to IVF treatment

(<4 oocytes retrieved). This group of women represents a challenging group in IVF clinics because their pregnancy rates are lower. The aim was to investigate whether ovarian reserve tests such as AMH and AFC are useful in predicting which women with a poor response will or will not become pregnant after IVF treatment. We found that any single variable or combination of variables achieved only limited accuracy in the prediction of ongoing pregnancy for the individual.

Chapter 4 investigated which lifestyle factors and reproductive factors influence AMH levels in the general population. The study included 2,320 women. It showed that women who are pregnant, who are taking the oral contraceptive pill, and who smoke at the time of AMH measurement all have significantly lower AMH levels. The effect seemed to be reversible. BMI, waist circumference, alcohol consumption, physical exercise and socioeconomic status were not significantly associated with AMH values.

Chapter 5 we compared the distribution of AMH in 27,563 subfertile women to the distribution of age at menopause in another cohort of 2,249 women with subfertility and used these distributions to make predictions of age at menopause using AMH. The similarity between the observed ages at menopause and the predicted ages at menopause supported the hypothesis that age-related AMH decline is associated with menopause, thus making AMH an excellent candidate biomarker for age at menopause prediction.

When addressing probable future age at menopause women are often asked at what age their mother's became menopausal. In Chapter 6 we compared the predictive performance of AMH with the predictive performance of mother's age at menopause. We found that AMH and mother's age at menopause are both predictive age at menopause, but AMH is a more accurate predictor of age at menopause. Perhaps this is because AMH reflects the environmental influences as well as the genetic influences of ovarian ageing.

In Chapter 7 two models (the SRV and TLGS models) of menopausal age prediction were compared and the validity of these models was tested by applying each model, first to their own derivation dataset, and then to the other. The results showed that both models discriminate well between women that enter menopause early or late during follow up. The models

were also tested to see if the proportion of women who were predicted to enter menopause during follow-up coincided with the actual observed proportion of women who entered menopause during follow up. For the SRV model on the TLGS data this correlation was strong, for the TLGS model in the SRV data this correlation was poor. This difference is thought to be the result of a difference in which AMH is modeled to be an age and time dependent predictor in the models.

In Chapter 8 we looked at how well age at menopause can be predicted in a population-based sample of 1,163 women. At the start of the study AMH was measured, 10 years later menstrual status was assessed. A model including female age, BMI, pack-years of smoking, and menstrual cycle status (regular, irregular, pregnant or taking oral contraceptives) discriminated rather well between women who entered menopause early or late during follow up. Adding AMH to this model increased the accuracy of the prediction of menopause occurring within 10 years, but the added value varied by age and reproductive characteristics of women.

In Chapter 9**,** the results and conclusions from the previous chapters are discussed. In this thesis we showed that one measure of AMH is not generalizable to all women of the same age but that it is significantly influenced by both reproductive and lifestyle factors. We have shown that AMH can adequately predict response to IVF, but insufficiently predicts the occurrence of an ongoing pregnancy in women with a poor response to IVF treatment. And lastly, we have presented evidence that AMH has additive value in the prediction of female age at menopause in both subfertile women and in the general population, but that the accuracy with which predictions are made depend on the subgroup of women in which AMH is determined. Chapter 9 further discusses what the clinical implications are of these results. Perhaps, in the future AMH can be used to predict until which age women remain fertile. In women who are at risk of early infertility this information can be used to guide them to start having children earlier or to freeze their oocytes which can be used when they do have an active child-wish. At present, the precision of the prediction of age at menopause is no way precise enough to be generally applied. Before AMH can be widely applied, important issues must be resolved around the laboratory technicalities of AMH measurements.

Nederlandse samenvatting

In dit proefschrift wordt de rol van AMH onderzocht als individuele voorspeller van de actuele en toekomstige fertiliteit (vruchtbaarheid) van de vrouw.

In Hoofdstuk 1, de introductie, wordt het concept reproductieve veroudering uitgelegd. AMH is een hormoon dat meetbaar is in het bloed van vrouwen. Uit voorafgaand onderzoek is gebleken dat AMH een waardevolle marker is voor ovariële reserve, ofwel het aantal resterende follikels in de eierstokken van de vrouw. Met het ouder worden, neemt bij iedere vrouw het aantal eicellen in de ovaria (eierstokken) geleidelijk af, tegelijkertijd verslechterd ook de kwaliteit van de eicellen en dit leidt tot vier achtereenvolgende mijlpalen: verminderde vruchtbaarheid, onvruchtbaarheid, de menopauzale transitie en uiteindelijk de menopauze (overgang). Het is aannemelijk dat er een vaste tijdsrelatie bestaat tussen deze reproductieve mijlpalen, waarbij de menopauze ongeveer 10 jaar na de onvruchtbaarheid optreedt. Er is een aanzienlijk verschil in het reproductieve verouderingsproces tussen vrouwen van dezelfde leeftijd, dit blijkt uit de grote spreiding in menopauzeleeftijd van 40 tot 60 jaar. Dit zou vervolgens inhouden dat sommige vrouwen al op hun dertigste leeftijd onvruchtbaar zijn. Omdat AMH de ovariële reserve weergeeft, werd verondersteld dat AMH ook gebruikt kan worden om te voorspellen op welke leeftijd een vrouw de menopauze bereikt. Hieruit zou dan weer kunnen worden afgeleid tot wanneer zij vruchtbaar is. Tevens zou AMH ook bruikbaar kunnen zijn om de respons van een vrouw te voorspellen op een IVF behandeling (het aantal eitjes dat verkregen wordt na een IVF behandeling). Het voorspellen van de menopauze en de respons op IVF behandelingen wordt in dit proefschrift "fertility forecasting" genoemd.

Hoofdstuk 2 en 3 beschrijven twee meta-analyses op individueel patiënt niveau over de voorspellende waarde van ovariële reserve testen zoals AMH. In Hoofdstuk 2 werd beschreven wat de waarde is van AMH in het voorspellen van een hyperrespons op IVF. En in Hoofdstuk 3 werd vervolgens gekeken naar de voorspelling van een zwangerschap na een slechte respons op een IVF behandeling. De originele data van eerder gepubliceerde studies werden verzameld en samengevoegd. Uiteindelijk was er informatie verkrijgbaar uit 33 verschillende databases van 6852 patiënten die een IVF behandeling hadden ondergaan. Met deze data werd aangetoond dat AMH een goede voorspeller is voor de respons op een IVF behandeling (het aantal eitjes dat verkregen wordt; Hoofdstuk 2), maar niet

voor het voorspellen van een doorgaande zwangerschap bij vrouwen met een slecht respons op IVF (Hoofdstuk 3)

In Hoofdstuk 4 werd aangetoond dat verschillende leefstijlfactoren van invloed zijn op AMH waarden, zoals het gebruik van orale anticonceptiva, zwangerschap en roken, terwijl andere factoren zoals body mass index, buikomvang, alcohol consumptie, lichamelijke beweging en sociaal economische status geen invloed hadden op leeftijdsgebonden AMH waarden.

In Hoofdstuk 5 werden twee cohorten van subfertiele vrouwen met elkaar vergeleken. Door de distributie van AMH in cohort 1, bestaande uit 27,536 vrouwen, te vergelijken met de distributie van de leeftijd van de menopauze in cohort 2, bestaande uit 2,249 vrouwen konden we voorspellingen maken voor menopauze leeftijd, gebaseerd op de leeftijdsgebonden AMH waarden. De gelijkenis tussen de geobserveerde en de voorspelde menopauze leeftijden was aanzienlijk, hiermee werd onze hypothese dat leeftijdsgebonden AMH geassocieerd is met de menopauze leeftijd onderbouwd. Aan de hand hiervan kunnen we stellen dat AMH een goede biomarker is voor het voorspellen van de menopauze leeftijd.

Vaak wordt er gevraagd naar de leeftijd waarop de moeder van een vrouw menopauzaal werd om een indicatie te geven van de leeftijd waarop zij zelf (als dochter) menopauzaal zal worden. In Hoofdstuk 6 vergeleken we de predictieve accuratesse van moeders menopauzeleeftijd met dat van AMH in het voorspellen van leeftijd menopauze van de dochter. Uit dit onderzoek bleek AMH een betere voorspelling te geven van de menopauze leeftijd van de dochter dan moeders menopauze leeftijd. Wellicht is AMH een weerspiegeling van zowel genetische determinanten van reproductieve veroudering als leefstijl en/of omgevingsfactoren.

In Hoofdstuk 7 werd de accuratesse en de validiteit van twee gepubliceerde modellen van menopauze leeftijd predictie vergeleken, een Nederlands model (model 1) en een Iranees model (model 2). Eerst werden beiden modellen toegepast op de eigen data en daarna op de validatie dataset. De resultaten lieten zien dat beide modellen goed konden discrimineren tussen wie vroeg of laat in de menopauze kwam. Daarna werden de modellen getoetst om te kijken of het aantal vrouwen waarbij voorspeld was dat zij menopauzaal zouden worden tijdens de follow-up, ook

daadwerkelijk menopauzaal waren geworden tijdens follow-up. Voor het SRV model was deze correlatie sterk terwijl deze correlatie zwak was voor het TLGS model. Dit komt waarschijnlijk door het verschil waarop AMH als leeftijd- en tijdsgebonden variabel is gemodelleerd.

In Hoofdstuk 8 werd een model gebouwd voor menopauze leeftijd voorspelling met informatie van 1,163 vrouwen uit het Doetinchem cohort. Aan het begin van de studie werd AMH gemeten in alle premenopauzale vrouwen, en 10 jaar later werd beoordeeld of vrouwen wel of niet menopauzaal waren geworden. Een model met leeftijd, BMI, roken en de menstruele cyclus (regulair, irregulair, zwanger of orale anticonceptiva gebruik) kon goed discrimineren tussen vrouwen die tijdens follow up vroeg of juist laat menopauzaal werden. Door AMH toe te voegen werd de predictieve accuratesse hoger, maar de toegevoegde waarde van AMH verschilde per leeftijd en de reproductieve karakteristieken van de vrouwen.

Hoofdstuk 9 bespreekt de resultaten en conclusies die uit dit proefschrift getrokken kunnen worden. In dit proefschrift wordt aangetoond dat AMH potentie heeft voor het voorspellen van de menopauze leeftijd van een vrouw. Dit is interessant, gezien het feit dat we de leeftijd tot wanneer een vrouw vruchtbaar blijft kunnen afleiden van de menopauze leeftijd. Om die reden kan AMH mogelijk in de toekomst gebruikt worden om vrouwen met een verhoogd risico op vroege infertiliteit te adviseren om eerder een gezin te stichten ofwel eicellen in te laten vriezen om zo een toekomstige kinderwens in vervulling te laten gaan. Echter, deze gevonden resultaten zijn nog niet geschikt voor een klinische toepassing en studies met lange termijn follow up zijn daarom noodzakelijk om te onderzoeken of het klinische beleid aangepast kan of moet worden op basis van zulke individuele voorspellingen. Daarnaast dienen de laboratorium procedures rondom de AMH metingen gestandaardiseerd te worden om op internationaal niveau betrouwbare AMH vergelijkingen mogelijk te maken.

References

(1) Faddy MJ, Gosden RG. A mathematical model of follicle dynamics in the human ovary. Hum Reprod 1995 Apr;10(4):770-5.

(2) te Velde ER, Pearson PL. The variability of female reproductive ageing. Hum Reprod Update 2002 Mar;8(2):141-54.

(3) Broekmans FJ, Knauff EA, te Velde ER, Macklon NS, Fauser BC. Female reproductive ageing: current knowledge and future trends. Trends Endocrinol Metab 2007 Mar;18(2):58-65.

(4) Treloar AE, Boynton RE, Behn BG, Brown BW. Variation of the human menstrual cycle through reproductive life. Int J Fertil 1967 Jan;12(1 Pt 2):77-126.

(5) den Tonkelaar I, te Velde ER, Looman CW. Menstrual cycle length preceding menopause in relation to age at menopause. Am J Human Biol 1998 Jun 3;29(2):115-23.

(6) Visser JA, Themmen AP. Anti-Mullerian hormone and folliculogenesis. Mol Cell Endocrinol 2005 Apr 29;234(1-2):81-6.

(7) Nelson SM, La MA. The journey from the old to the new AMH assay: how to avoid getting lost in the values. Reprod Biomed Online 2011 Oct;23(4):411-20.

(8) White YA, Woods DC, Takai Y, Ishihara O, Seki H, Tilly JL. Oocyte formation by mitotically active germ cells purified from ovaries of reproductive-age women. Nat Med 2012;18(3):413-21.

(9) Nybo Andersen AM, Wohlfahrt J, Christens P, Olsen J, Melbye M. Maternal age and fetal loss: population based register linkage study. BMJ 2000 Jun 24;320(7251):1708-12.

(10) Nicolaidis P, Petersen MB. Origin and mechanisms of non-disjunction in human autosomal trisomies. Hum Reprod 1998 Feb;13(2):313-9.

(11) van Noord-Zaadstra BM, Looman CW, Alsbach H, Habbema JD, te Velde ER, Karbaat J. Delaying childbearing: effect of age on fecundity and outcome of pregnancy. BMJ 1991 Jun 8;302(6789):1361-5.

(12) Leridon H. Demographic effects of the introduction of steroid contraception in developed countries. Hum Reprod Update 2006 Sep;12(5):603-16.

(13) OECD. OECD Family Database. 2011. Ref Type: Generic

(14) Leridon H. Can assisted reproduction technology compensate for the natural decline in fertility with age? A model assessment. Hum Reprod 2004 Jul;19(7):1548-53.

(15) Habbema JD, Eijkemans MJ, Nargund G, Beets G, Leridon H, te Velde ER. The effect of in vitro fertilization on birth rates in western countries. Hum Reprod 2009 Jun;24(6):1414-9.

(16) Fauser BC, Devroey P, Macklon NS. Multiple birth resulting from ovarian stimulation for subfertility treatment. Lancet 2005 May 21;365(9473):1807-16.

(17) Lutz W, O'Neill BC, Scherbov S. Demographics. Europe's population at a turning point. Science 2003 Mar 28;299(5615):1991-2.

(18) Block E. A quantitative morphological investigation of the follicular system in newborn female infants. Acta Anat (Basel) 1953;17(3):201-6.

(19) Block E. Quantitative morphological investigations of the follicular system in women; variations at different ages. Acta Anat (Basel) 1952;14(1-2):108-23.

(20) Baker TG. A quantitative and cytological study of germ cells in human ovaries. Proc R Soc Lond B Biol Sci 1963 Oct 22;158:417-33.

(21) Markstrom E, Svensson EC, Shao R, Svanberg B, Billig H. Survival factors regulating ovarian apoptosis -- dependence on follicle differentiation. Reproduction 2002 Jan;123(1):23-30.

(22) Faddy MJ, Gosden RG. A model conforming the decline in follicle numbers to the age of menopause in women. Hum Reprod 1996 Jul;11(7):1484-6.

(23) Santoro N, Randolph JF, Jr. Reproductive hormones and the menopause transition. Obstet Gynecol Clin North Am 2011 Sep;38(3):455-66.

(24) Richardson SJ, Senikas V, Nelson JF. Follicular depletion during the menopausal transition: evidence for accelerated loss and ultimate exhaustion. J Clin Endocrinol Metab 1987 Dec;65(6):1231-7.

(25) Leidy LE, Godfrey LR, Sutherland MR. Is follicular atresia biphasic? Fertil Steril 1998;70(5):851-9.

(26) Hansen KR, Knowlton NS, Thyer AC, Charleston JS, Soules MR, Klein NA. A new model of reproductive aging: the decline in ovarian non-growing follicle number from birth to menopause. Hum Reprod 2008 Mar;23(3):699-708.

(27) Faddy MJ, Gosden RG, Gougeon A, Richardson SJ, Nelson JF. Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. Hum Reprod 1992 Nov;7(10):1342-6.

(28) Lisabeth L, Harlow S, Qaqish B. A new statistical approach demonstrated menstrual patterns during the menopausal transition did not vary by age at menopause. J Clin Epidemiol 2004 May;57(5):484-96.

(29) Ata B, Kaplan B, Danzer H, Glassner M, Opsahl M, Tan SL, et al. Array CGH analysis shows that aneuploidy is not related to the number of embryos generated. Reprod Biomed Online 2012 Jun;24(6):614-20.

(30) Gianaroli L, Magli MC, Ferraretti AP, Munne S. Preimplantation diagnosis for aneuploidies in patients undergoing in vitro fertilization with a poor prognosis: identification of the categories for which it should be proposed. Fertil Steril 1999 Nov;72(5):837-44.

(31) Wells D, Delhanty JD. Comprehensive chromosomal analysis of human preimplantation embryos using whole genome amplification and single cell comparative genomic hybridization. Mol Hum Reprod 2000 Nov;6(11):1055-62.

(32) Handyside AH, Montag M, Magli MC, Repping S, Harper J, Schmutzler A, et al. Multiple meiotic errors caused by predivision of chromatids in women of advanced maternal age undergoing in vitro fertilisation. Eur J Hum Genet 2012 Feb 8.

(33) Warburton D. The effect of maternal age on the frequency of trisomy: change in meiosis or in utero selection? Prog Clin Biol Res 1989;311:165-81.

(34) Lamb NE, Yu K, Shaffer J, Feingold E, Sherman SL. Association between maternal age and meiotic recombination for trisomy 21. Am J Hum Genet 2005 Jan;76(1):91-9.

(35) Tarin JJ. Aetiology of age-associated aneuploidy: a mechanism based on the 'free radical theory of ageing'. Hum Reprod 1995 Jun;10(6):1563-5.

(36) Wilding M, De PG, De ML, Marino M, Alviggi C, Dale B. Chaotic mosaicism in human preimplantation embryos is correlated with a low mitochondrial membrane potential. Fertil Steril 2003 Feb;79(2):340-6.

(37) Schwartz D, 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 Feb 18;306(7):404-6.

(38) Bouchard G. Population studies and genetic epidemiology in northeast Quebec. Can Stud Popul 1989;16(1):61-86.

(39) de Boer EJ, den T, I, te Velde ER, Burger CW, Klip H, van Leeuwen FE. A low number of retrieved oocytes at in vitro fertilization treatment is predictive of early menopause. Fertil Steril 2002 May;77(5):978-85.

(40) te Velde ER, Eijkemans R, Habbema HD. Variation in couple fecundity and time to pregnancy, an essential concept in human reproduction. Lancet 2000 Jun 3;355(9219):1928-9.

(41) O'Connor KA, Holman DJ, Wood JW. Declining fecundity and ovarian ageing in natural fertility populations. Am J Human Biol 1998 Oct 12;30(2):127-36.

(42) Sauer MV, Kavic SM. Oocyte and embryo donation 2006: reviewing two decades of innovation and controversy. Reprod Biomed Online 2006 Feb;12(2):153-62.

(43) Broekmans FJ, Soules MR, Fauser BC. Ovarian aging: mechanisms and clinical consequences. Endocr Rev 2009 Aug;30(5):465-93.

(44) Brook JD, Gosden RG, Chandley AC. Maternal ageing and aneuploid embryos--evidence from the mouse that biological and not chronological age is the important influence. Hum Genet 1984;66(1):41-5.

(45) Eichenlaub-Ritter U, Chandley AC, Gosden RG. The CBA mouse as a model for age-related aneuploidy in man: studies of oocyte maturation, spindle formation and chromosome alignment during meiosis. Chromosoma 1988;96(3):220-6.

(46) Freeman SB, Yang Q, Allran K, Taft LF, Sherman SL. Women with a reduced ovarian complement may have an increased risk for a child with Down syndrome. Am J Hum Genet 2000 May;66(5):1680-3.

(47) Haadsma ML, Mooij TM, Groen H, Burger CW, Lambalk CB, Broekmans FJ, 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 Feb;25(2):552-8.

(48) Nasseri A, Mukherjee T, Grifo JA, Noyes N, Krey L, Copperman AB. Elevated day 3 serum follicle stimulating hormone and/or estradiol may predict fetal aneuploidy. Fertil Steril 1999;71(4):715-8.

(49) van Montfrans JM, Dorland M, Oosterhuis GJ, van Vugt JM, Rekers-Mombarg LT, Lambalk CB. Increased concentrations of follicle-stimulating hormone in mothers of children with Down's syndrome. Lancet 1999 May 29;353(9167):1853-4.

(50) van der Stroom EM, Konig TE, van Dulmen-den BE, Elzinga WS, van Montfrans JM, Haadsma ML, et al. Early menopause in mothers of children with Down syndrome? Fertil Steril 2011 Oct;96(4):985-90.

(51) Kline JK, Kinney AM, Levin B, Kelly AC, Ferin M, Warburton D. Trisomic pregnancy and elevated FSH: implications for the oocyte pool hypothesis. Hum Reprod 2011 Jun;26(6):1537-50.

(52) Thum MY, Kalu E, Abdalla H. Elevated basal FSH and embryo quality: lessons from extended culture embryos: raised FSH and blastocyst quality. J Assist Reprod Genet 2009 Jun;26(6):313-8.

(53) Biljan MM, Buckett WM, Dean N, Phillips SJ, Tan SL. The outcome of IVF-embryo transfer treatment in patients who develop three follicles or less. Hum Reprod 2000 Oct;15(10):2140-4.

(54) Hanoch J, Lavy Y, Holzer H, Hurwitz A, Simon A, Revel A, et al. Young low responders protected from untoward effects of reduced ovarian response [see comments]. Fertil Steril 1998 Jun;69(6):1001-4.

(55) Oudendijk JF, Yarde F, Eijkemans MJ, Broekmans FJ, Broer SL. The poor responder in IVF: is the prognosis always poor?: a systematic review. Hum Reprod Update 2012 Jan;18(1):1-11.

(56) Sunkara SK, Rittenberg V, Raine-Fenning N, Bhattacharya S, Zamora J, Coomarasamy A. Association between the number of eggs and live birth in IVF treatment: an analysis of 400 135 treatment cycles. Hum Reprod 2011 Jul;26(7):1768-74.

(57) Weghofer A, Dietrich W, Barad DH, Gleicher N. Live birth chances in women with extremely low-serum anti-Mullerian hormone levels. Hum Reprod 2011 Jul;26(7):1905-9.

(58) Haadsma ML, Groen H, Mooij TM, Burger CW, Broekmans FJ, Lambalk CB, et al. Miscarriage risk for IVF pregnancies in poor responders to ovarian hyperstimulation. Reprod Biomed Online 2010 Feb;20(2):191-200.

(59) Voorhuis M, Onland-Moret NC, van der Schouw YT, Fauser BC, Broekmans FJ. Human studies on genetics of the age at natural menopause: a systematic review. Hum Reprod Update 2010 Jul;16(4):364-77.

(60) Morris DH, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ. Familial concordance for age at natural menopause: results from the Breakthrough Generations Study. Menopause 2011 Sep;18(9):956-61.

(61) Torgerson DJ, Thomas RE, Reid DM. Mothers and daughters menopausal ages: is there a link? Eur J Obstet Gynecol Reprod Biol 1997 Jul;74(1):63-6.

(62) Snieder H, MacGregor AJ, 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 Jun;83(6):1875-80.

(63) de Bruin JP, Bovenhuis H, van Noord PA, Pearson PL, van Arendonk JA, te Velde ER, et al. The role of genetic factors in age at natural menopause. Hum Reprod 2001 Sep;16(9):2014-8.

(64) Murabito JM, Yang Q, Fox C, Wilson PW, Cupples LA. Heritability of age at natural menopause in the Framingham Heart Study. J Clin Endocrinol Metab 2005 Jun;90(6):3427-30; 42480.

(65) van Asselt KM, Kok HS, Pearson PL, Dubas JS, Peeters PH, te Velde ER, et al. Heritability of menopausal age in mothers and daughters. Fertil Steril 2004 Nov;82(5):1348-51.

(66) Stolk L, Perry JR, Chasman DI, He C, Mangino M, Sulem P, et al. Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. Nat Genet 2012;44(3):260-8.

(67) Kok HS, van Asselt KM, van der Schouw YT, van der Tweel I, Peeters PH, Wilson PW, et al. Heart disease risk determines menopausal age rather than the reverse. J Am Coll Cardiol 2006 May 16;47(10):1976-83.

(68) Koochmeshgi J, Hosseini-Mazinani SM, Morteza SS, Hosein-Pur-Nobari N, Teimoori-Toolabi L. Apolipoprotein E genotype and age at menopause. Ann N Y Acad Sci 2004 Jun;1019:564-7.

(69) Bonomini F, Filippini F, Hayek T, Aviram M, Keidar S, Rodella LF, et al. Apolipoprotein E and its role in aging and survival. Exp Gerontol 2010 Feb;45(2):149-57.

(70) van Disseldorp J., Broekmans FJ, Peeters PH, Fauser BC, van der Schouw YT. The association between vascular function-related genes and age at natural menopause. Menopause 2008 May;15(3):511-6.

(71) van Asselt KM, Kok HS, Peeters PH, Roest M, Pearson PL, te Velde ER, et al. Factor V Leiden mutation accelerates the onset of natural menopause. Menopause 2003 Sep;10(5):477-81.

(72) Midgette AS, Baron JA. Cigarette smoking and the risk of natural menopause. Epidemiology 1990 Nov;1(6):474-80.

(73) Yang Q, Sherman SL, Hassold TJ, Allran K, Taft L, Pettay D, et al. Risk factors for trisomy 21: maternal cigarette smoking and oral contraceptive use in a population-based case-control study. Genet Med 1999 Mar;1(3):80-8.

(74) Matzuk MM. Eggs in the balance. Nat Genet 2001 Aug;28(4):300-1.

(75) Broer SL, Mol B, Dolleman M, Fauser BC, Broekmans FJ. The role of anti-Mullerian hormone assessment in assisted reproductive technology outcome. Curr Opin Obstet Gynecol 2010 Jun;22(3):193-201.

(76) Gougeon A, Ecochard R, 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 Mar;50(3):653-63.

(77) Broekmans FJ, Kwee J, Hendriks DJ, Mol BW, Lambalk CB. A systematic review of tests predicting ovarian reserve and IVF outcome. Hum Reprod Update 2006 Nov;12(6):685-718.

(78) Broer SL, Eijkemans MJ, Scheffer GJ, van Rooij I, de Vet A., Themmen AP, et al. Anti-mullerian hormone predicts menopause: a long-term follow-up study in normoovulatory women. J Clin Endocrinol Metab 2011 Aug;96(8):2532-9.

(79) Tehrani FR, Shakeri N, Solaymani-Dodaran M, Azizi F. Predicting age at menopause from serum antimullerian hormone concentration. Menopause 2011 Jul;18(7):766-70.

(80) Hendriks DJ, te Velde E.R., Looman CW, et al. The role of a poor response in the prediction of the cumulative ongoing pregancy rate in in vitro fertilisation.In: Dynamic and basal ovarian reserve tests for outcome prediction in IVF: comparisons and meta-analyses. 2005. Utrecht, Academic Thesis.

(81) Hendriks DJ, Kwee J, Mol BW, te Velde ER, Broekmans FJ. 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 Apr;87(4):764-75.

(82) Scott RT, Jr., Elkind-Hirsch KE, Styne-Gross A, Miller KA, Frattarelli JL. 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 Apr;89(4):868-78.

(83) Patrizio P, Fragouli E, Bianchi V, Borini A, Wells D. Molecular methods for selection of the ideal oocyte. Reprod Biomed Online 2007 Sep;15(3):346-53.

(84) Tanghe S, Van Soom A., Nauwynck H, Coryn M, de KA. Minireview: Functions of the cumulus oophorus during oocyte maturation, ovulation, and fertilization. Mol Reprod Dev 2002 Mar;61(3):414-24.

(85) Adriaenssens T, Segers I, Wathlet S, Smitz J. The cumulus cell gene expression profile of oocytes with different nuclear maturity and potential for blastocyst formation. J Assist Reprod Genet 2011 Jan;28(1):31-40.

(86) Boomsma CM, Macklon NS. What can the clinician do to improve implantation? Reprod Biomed Online 2006 Dec;13(6):845-55.

(87) Broer SL. Assessment of Current and Future Ovarian Reserve Status. 2011. Enschede, The Netherlands, Gildeprint Drukkerijen.

(88) Eijkemans MJ, Habbema J.D.F., 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.

(89) Broekmans FJ, Faddy MJ, Scheffer G, te Velde ER. Antral follicle counts are related to age at natural fertility loss and age at menopause. Menopause 2004 Nov;11(6 Pt 1):607-14.

(90) Delvigne A, Rozenberg S. Epidemiology and prevention of ovarian hyperstimulation syndrome (OHSS): a review. Hum Reprod Update 2002 Nov;8(6):559-77.

(91) Bancsi LF, Huijs AM, den Ouden CT, Broekmans FJ, Looman CW, Blankenstein MA, et al. Basal follicle-stimulating hormone levels are of limited value in predicting ongoing pregnancy rates after in vitro fertilization. Fertil Steril 2000 Mar;73(3):552-7.

(92) Baart EB, Martini E, van den Berg I, Macklon NS, Galjaard RJ, Fauser BC, et al. Preimplantation genetic screening reveals a high incidence of aneuploidy and mosaicism in embryos from young women undergoing IVF. Hum Reprod 2006 Jan;21(1):223-33.

(93) Luna M, Grunfeld L, Mukherjee T, Sandler B, Copperman AB. 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 Apr;87(4):782-7.

(94) Heijnen EM, Eijkemans MJ, De Klerk C., Polinder S, Beckers NG, Klinkert ER, et al. A mild treatment strategy for in-vitro fertilisation: a randomised non-inferiority trial. Lancet 2007 Mar 3;369(9563):743-9.

(95) Eldar-Geva T, Ben-Chetrit A, Spitz IM, Rabinowitz R, Markowitz E, Mimoni T, 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 Nov;20(11):3178-83.

(96) Verberg MF, Eijkemans MJ, Macklon NS, Heijnen EM, Baart EB, Hohmann FP, et al. The clinical significance of the retrieval of a low number of oocytes following mild ovarian stimulation for IVF: a meta-analysis. Hum Reprod Update 2009 Jan;15(1):5-12.

(97) van der Gaast MH, Eijkemans MJ, van der Net JB, de Boer EJ, Burger CW, van Leeuwen FE, et al. Optimum number of oocytes for a successful first IVF treatment cycle. Reprod Biomed Online 2006 Oct;13(4):476-80.

(98) Erdem M, Erdem A, Gursoy R, Biberoglu K. 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 Feb;21(2):37-45.

(99) Fauser BC, Diedrich K, Devroey P. Predictors of ovarian response: progress towards individualized treatment in ovulation induction and ovarian stimulation. Hum Reprod Update 2008 Jan;14(1):1-14.

(100) Freour T, Mirallie S, Bach-Ngohou K, Denis M, Barriere P, Masson D. 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 Jan;375(1-2):162-4.

(101) Gnoth C, Schuring AN, Friol K, Tigges J, Mallmann P, Godehardt E. Relevance of anti-Mullerian hormone measurement in a routine IVF program. Hum Reprod 2008 Jun;23(6):1359-65.

(102) Ho HY, Lee RK, Lin MH, Hwu YM. Estradiol level on day 9 as a predictor of risk for ovarian hyperresponse during controlled ovarian hyperstimulation. J Assist Reprod Genet 2003 Jun;20(6):222-6.

(103) Jayaprakasan K, Hilwah N, Kendall NR, Hopkisson JF, Campbell BK, Johnson IR, 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 Jul;22(7):1932-41.

(104) Klinkert ER, Broekmans FJ, Looman CW, Habbema JD, te Velde ER. 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 Mar;83(3):811-4.

(105) van Rooij I, Broekmans FJ, te Velde ER, Fauser BC, Bancsi LF, de Jong FH, et al. Serum anti-Mullerian hormone levels: a novel measure of ovarian reserve. Hum Reprod 2002 Dec;17(12):3065-71.

(106) Kwee J, Elting MW, Schats R, Bezemer PD, Lambalk CB, Schoemaker J. 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 Jul;18(7):1422-7.

(107) La Marca A., Giulini S, Tirelli A, Bertucci E, Marsella T, Xella S, et al. Anti-Mullerian hormone measurement on any day of the menstrual cycle strongly predicts ovarian response in assisted reproductive technology. Hum Reprod 2007 Mar;22(3):766-71.

(108) Nakhuda GS, Chu MC, Wang JG, Sauer MV, Lobo RA. Elevated serum mullerian-inhibiting substance may be a marker for ovarian hyperstimulation syndrome in normal women undergoing in vitro fertilization. Fertil Steril 2006 May;85(5):1541-3.

(109) Liu KE, 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 Aug;90(2):297-301.

(110) Smeenk JM, Stolwijk AM, Kremer JA, Braat DD. External validation of the templeton model for predicting success after IVF. Hum Reprod 2000 May;15(5):1065-8.

(111) Riggs RM, Duran EH, Baker MW, Kimble TD, Hobeika E, Yin L, 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 Aug;199(2):202-8.

(112) Smeenk JM, Sweep FC, Zielhuis GA, Kremer JA, Thomas CM, Braat DD. Antimullerian hormone predicts ovarian responsiveness, but not embryo quality or pregnancy, after in vitro fertilization or intracyoplasmic sperm injection. Fertil Steril 2007 Jan;87(1):223-6.

(113) Nardo LG, Gelbaya TA, Wilkinson H, Roberts SA, Yates A, Pemberton P, 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 Nov;92(5):1586-93.

(114) McIlveen M, Skull JD, 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 Mar;22(3):778-85.

(115) Merce LT, Barco MJ, Bau S, Troyano JM. Prediction of ovarian response and IVF/ICSI outcome by three-dimensional ultrasonography and power Doppler angiography. Eur J Obstet Gynecol Reprod Biol 2007 May;132(1):93-100.

(116) Steinkampf MP, Hammond KR, Nichols JE, Slayden SH. Effect of obesity on recombinant follicle-stimulating hormone absorption: subcutaneous versus intramuscular administration. Fertil Steril 2003 Jul;80(1):99-102.

(117) Broer SL, Mol BW, Hendriks D, Broekmans FJ. The role of antimullerian hormone in prediction of outcome after IVF: comparison with the antral follicle count. Fertil Steril 2009 Mar;91(3):705-14.

(118) Broeze KA, Opmeer BC, Bachmann LM, Broekmans FJ, Bossuyt PM, Coppus SF, 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.

(119) Broeze KA, Opmeer BC, Van Geloven N., Coppus SF, Collins JA, Den Hartog JE, et al. Are patient characteristics associated with the accuracy of hysterosalpingography in diagnosing tubal pathology? An individual patient data meta-analysis. Hum Reprod Update 2011 May;17(3):293-300.

(120) Broer SL, Dolleman M, Opmeer BC, Fauser BC, Mol BW, Broekmans FJ. AMH and AFC as predictors of excessive response in controlled ovarian hyperstimulation: a meta-analysis. Hum Reprod Update 2011 Jan;17(1):46-54.

(121) Ebner T, Sommergruber M, Moser M, Shebl O, Schreier-Lechner E, Tews G. Basal level of anti-Mullerian hormone is associated with oocyte quality in stimulated cycles. Hum Reprod 2006 Aug;21(8):2022-6.

(122) Aflatoonian A, Oskouian H, Ahmadi S, Oskouian L. 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 Jun;26(6):319-25.

(123) Ng EH, Yeung WS, Ho PC. The significance of antral follicle count in controlled ovarian stimulation and intrauterine insemination. J Assist Reprod Genet 2005 Oct;22(9-10):323-8.

(124) Sterrenburg MD, Veltman-Verhulst SM, Eijkemans MJ, Hughes EG, Macklon NS, Broekmans FJ, et al. Clinical outcomes in relation to the daily dose of recombinant follicle-stimulating hormone for ovarian stimulation in in vitro fertilization in presumed normal responders younger than 39 years: a meta-analysis. Hum Reprod Update 2011 Mar;17(2):184-96.

(125) Janes H, Longton G, Pepe M. Accommodating Covariates in ROC Analysis. Stata J 2009 Jan 1;9(1):17-39.

(126) Pepe M, Longton G, Janes H. Estimation and Comparison of Receiver Operating Characteristic Curves. Stata J 2009 Mar 1;9(1):1.

(127) Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-36.

(128) Ashrafi M, Madani T, Tehranian AS, Malekzadeh F. Follicle stimulating hormone as a predictor of ovarian response in women undergoing controlled ovarian hyperstimulation for IVF. Int J Gynaecol Obstet 2005 Oct;91(1):53-7.

(129) Caroppo E, Matteo M, Schonauer LM, Vizziello G, Pasquadibisceglie A, Vitti A, et al. Basal FSH concentration as a predictor of IVF outcome in older women undergoing stimulation with GnRH antagonist. Reprod Biomed Online 2006 Dec;13(6):815-20.

(130) Muttukrishna S, Suharjono H, McGarrigle H, Sathanandan M. Inhibin B and anti-Mullerian hormone: markers of ovarian response in IVF/ICSI patients? BJOG 2004 Nov;111(11):1248-53.

(131) Muttukrishna S, McGarrigle H, Wakim R, Khadum I, Ranieri DM, Serhal P. Antral follicle count, anti-mullerian hormone and inhibin B: predictors of ovarian response in assisted reproductive technology? BJOG 2005 Oct;112(10):1384-90.

(132) Nelson SM, Yates RW, 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 Sep;22(9):2414-21.

(133) Ng EH, Chan CC, Tang OS, Ho PC. Antral follicle count and FSH concentration after clomiphene citrate challenge test in the prediction of ovarian response during IVF treatment. Hum Reprod 2005 Jun;20(6):1647-54.

(134) Popovic-Todorovic B, Loft A, Lindhard A, Bangsboll S, Andersson AM, Andersen AN. 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 Apr;18(4):781-7.

(135) Popovic-Todorovic B, Loft A, Bredkjaeer HE, Bangsboll S, Nielsen IK, Andersen AN. 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 Nov;18(11):2275-82.

(136) Tomas C, Nuojua-Huttunen S, Martikainen H. Pretreatment transvaginal ultrasound examination predicts ovarian responsiveness to gonadotrophins in in-vitro fertilization. Hum Reprod 1997 Feb;12(2):220-3.

(137) van Swieten EC, van der Leeuw-Harmsen, Badings EA, van der Linden PJ. Obesity and Clomiphene Challenge Test as predictors of outcome of in vitro fertilization and intracytoplasmic sperm injection. Gynecol Obstet Invest 2005;59(4):220-4.

(138) Vladimirov IK, Tacheva DM, Kalinov KB, Ivanova AV, Blagoeva VD. Prognostic value of some ovarian reserve tests in poor responders. Arch Gynecol Obstet 2005 Jun;272(1):74-9.

(139) Vladimirov I, Tacheva D, Blagoeva V. [Prognostic value of some hormonal and ultrasound ovarian reserve tests]. Akush Ginekol (Sofiia) 2003;42(5):14-20.

(140) Yong PY, Baird DT, Thong KJ, McNeilly AS, Anderson RA. 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 Jan;18(1):35-44.

(141) Anckaert E, Smitz J, Schiettecatte J, Klein BM, Arce JC. The value of anti-Mullerian hormone measurement in the long GnRH agonist protocol: association with ovarian response and gonadotrophin-dose adjustments. Hum Reprod 2012 Jun;27(6):1829-39.

(142) Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Stat Med 2008 Dec 20;27(29):6111-36.

(143) Andersen AN, Witjes H, Gordon K, Mannaerts B. Predictive factors of ovarian response and clinical outcome after IVF/ICSI following a rFSH/GnRH antagonist protocol with or without oral contraceptive pre-treatment. Hum Reprod 2011 Dec;26(12):3413-23.

(144) Lee S, Ozkavukcu S, Heytens E, Moy F, Alappat RM, Oktay K. Anti-Mullerian hormone and antral follicle count as predictors for embryo/oocyte cryopreservation cycle outcomes in breast cancer patients stimulated with letrozole and follicle stimulating hormone. J Assist Reprod Genet 2011 Jul;28(7):651-6.

(145) Nakhuda GS, Douglas NC, Thornton MH, Guarnaccia MM, Lobo R, Sauer MV. Anti-Mullerian hormone testing is useful for individualization of stimulation protocols in oocyte donors. Reprod Biomed Online 2011 Feb;22 Suppl 1:S88-S93.

(146) Riggs R, Kimble T, Oehninger S, Bocca S, Zhao Y, Leader B, et al. Anti-Mullerian hormone serum levels predict response to controlled ovarian hyperstimulation but not embryo quality or pregnancy outcome in oocyte donation. Fertil Steril 2011 Jan;95(1):410-2.

(147) van Tilborg TC, Eijkemans MJ, Laven JS, Koks CA, de Bruin JP, Scheffer GJ, et al. The OPTIMIST study: optimisation of cost effectiveness through individualised FSH stimulation dosages for IVF treatment. A randomised controlled trial. BMC Womens Health 2012 Sep 18;12(1):29.

(148) Popovic-Todorovic B, Loft A, Lindhard A, Bangsboll S, Andersson AM, Andersen AN. 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 Apr;18(4):781-7.

(149) Olivennes F, Howies CM, Borini A, Germond M, Trew G, Wikland M, et al. Individualizing FSH dose for assisted reproduction using a novel algorithm: the CONSORT study. Reprod Biomed Online 2011 Feb;22 Suppl 1:S73-S82.

(150) Jayaprakasan K, Hopkisson J, Campbell B, Johnson I, Thornton J, Raine-Fenning N. A randomised controlled trial of 300 versus 225 IU recombinant FSH for ovarian stimulation in predicted normal responders by antral follicle count. BJOG 2010 Jun;117(7):853-62.

(151) Moolenaar LM, Broekmans FJ, van Disseldorp J., Fauser BC, Eijkemans MJ, Hompes PG, et al. Cost effectiveness of ovarian reserve testing in in vitro fertilization: a Markov decision-analytic model. Fertil Steril 2011 Oct;96(4):889-94.

(152) Van der Meer M, Hompes PG, De Boer JA, Schats R, Schoemaker J. Cohort size rather than follicle-stimulating hormone threshold level determines ovarian sensitivity in polycystic ovary syndrome. J Clin Endocrinol Metab 1998 Feb;83(2):423-6.

(153) Saldeen P, Kallen K, Sundstrom P. The probability of successful IVF outcome after poor ovarian response. Acta Obstet Gynecol Scand 2007;86(4):457-61.

(154) Ulug U, Ben-Shlomo I, Turan E, Erden HF, Akman MA, Bahceci M. Conception rates following assisted reproduction in poor responder patients: a retrospective study in 300 consecutive cycles. Reprod Biomed Online 2003 Jun;6(4):439-43.

(155) van Loendersloot LL, van Wely M., Limpens J, Bossuyt PM, Repping S, van der Veen F. Predictive factors in in vitro fertilization (IVF): a systematic review and meta-analysis. Hum Reprod Update 2010 Nov;16(6):577-89.

(156) Hendriks DJ, te Velde ER, Looman CW, Bancsi LF, Broekmans FJ. Expected poor ovarian response in predicting cumulative pregnancy rates: a powerful tool. Reprod Biomed Online 2008 Nov;17(5):727-36.

(157) Lie Fong S., Baart EB, Martini E, Schipper I, Visser JA, Themmen AP, et al. Anti-Mullerian hormone: a marker for oocyte quantity, oocyte quality and embryo quality? Reprod Biomed Online 2008 May;16(5):664-70.

(158) Hazout A, Bouchard P, Seifer DB, Aussage P, Junca AM, Cohen-Bacrie P. 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 Nov;82(5):1323-9.

(159) Silberstein T, MacLaughlin DT, Shai I, Trimarchi JR, Lambert-Messerlian G, Seifer DB, et al. Mullerian inhibiting substance levels at the time of HCG administration in IVF cycles predict both ovarian reserve and embryo morphology. Hum Reprod 2006 Jan;21(1):159-63.

(160) Cupisti S, Dittrich R, Mueller A, Strick R, Stiegler E, Binder H, et al. Correlations between anti-mullerian hormone, inhibin B, and activin A in follicular fluid in IVF/ICSI patients for assessing the maturation and developmental potential of oocytes. Eur J Med Res 2007 Dec 14;12(12):604-8.

(161) Arce JC, La MA, Mirner KB, Nyboe AA, Fleming R. Antimullerian hormone in gonadotropin releasing-hormone antagonist cycles: prediction of ovarian response and cumulative treatment outcome in good-prognosis patients. Fertil Steril 2013 Feb 5.

(162) Broer SL, van Disseldorp J., Broeze KA, Dolleman M, Opmeer BC, Bossuyt P, 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. Hum Reprod Update 2013 Jan;19(1):26-36.

(163) La Marca A., Nelson SM, Sighinolfi G, Manno M, Baraldi E, Roli L, et al. Anti-Mullerian hormone-based prediction model for a live birth in assisted reproduction. Reprod Biomed Online 2011 Apr;22(4):341-9.

(164) Tremellen KP, Kolo M, Gilmore A, Lekamge DN. Anti-mullerian hormone as a marker of ovarian reserve. Aust N Z J Obstet Gynaecol 2005 Feb;45(1):20-4.

(165) Kini S, Li HW, Morrell D, Pickering S, Thong KJ. Anti-mullerian hormone and cumulative pregnancy outcome in in-vitro fertilization. J Assist Reprod Genet 2010 Aug;27(8):449-56.

(166) Irez T, Ocal P, Guralp O, Cetin M, Aydogan B, Sahmay S. Different serum anti-Mullerian hormone concentrations are associated with oocyte quality, embryo development parameters and IVF-ICSI outcomes. Arch Gynecol Obstet 2011 Nov;284(5):1295-301.

(167) Majumder K, Gelbaya TA, Laing I, Nardo LG. The use of anti-Mullerian hormone and antral follicle count to predict the potential of oocytes and embryos. Eur J Obstet Gynecol Reprod Biol 2010 Jun;150(2):166-70.

(168) Lekamge DN, Barry M, Kolo M, Lane M, Gilchrist RB, Tremellen KP. Anti-Mullerian hormone as a predictor of IVF outcome. Reprod Biomed Online 2007 May;14(5):602-10.

(169) Blazar AS, Lambert-Messerlian G, Hackett R, Krotz S, Carson SA, Robins JC. Use of in-cycle antimullerian hormone levels to predict cycle outcome. Am J Obstet Gynecol 2011 Sep;205(3):223-5.

(170) Broer SL, Dolleman M, van DJ, Broeze KA, Opmeer BC, Bossuyt PM, et al. Prediction of an excessive response in in vitro fertilization from patient characteristics and ovarian reserve tests and comparison in subgroups: an individual patient data meta-analysis. Fertil Steril 2013 May 28.

(171) Ferraretti AP, La MA, Fauser BC, Tarlatzis B, Nargund G, Gianaroli L. ESHRE consensus on the definition of 'poor response' to ovarian stimulation for in vitro fertilization: the Bologna criteria. Hum Reprod 2011 Jul;26(7):1616-24.

(172) Bhide P, Gudi A, Shah A, Timms P, Grayson K, Homburg R. Anti-Mullerian hormone as a predictor of pregnancy following IVF. Reprod Biomed Online 2012 Dec 8.

(173) Ubaldi F, Rienzi L, Baroni E, Ferrero S, Iacobelli M, Minasi MG, et al. Cumulative pregnancy rates after transfer of fresh and thawed embryos. Eur J Obstet Gynecol Reprod Biol 2004 Jul 1;115 Suppl 1:S106-S109.

(174) Tiitinen A, Hyden-Granskog C, Gissler M. What is the most relevant standard of success in assisted reproduction?: The value of cryopreservation on cumulative pregnancy rates per single oocyte retrieval should not be forgotten. Hum Reprod 2004 Nov;19(11):2439-41.

(175) Broekmans FJ, Visser JA, Laven JS, Broer SL, Themmen AP, Fauser BC. Anti-Mullerian hormone and ovarian dysfunction. Trends Endocrinol Metab 2008 Nov;19(9):340-7.

(176) Knauff EA, Eijkemans MJ, Lambalk CB, ten Kate-Booij MJ, Hoek A, Beerendonk CC, et al. Anti-Mullerian hormone, inhibin B, and antral follicle count in young women with ovarian failure. J Clin Endocrinol Metab 2009 Mar;94(3):786-92.

(177) Pigny P, Jonard S, Robert Y, Dewailly D. Serum anti-Mullerian hormone as a surrogate for antral follicle count for definition of the polycystic ovary syndrome. J Clin Endocrinol Metab 2006 Mar;91(3):941-5.

(178) Lie Fong S., Visser JA, Welt CK, de Rijke YB, Eijkemans MJ, Broekmans FJ, et al. Serum Anti-Mullerian Hormone Levels in Healthy Females: A Nomogram Ranging from Infancy to Adulthood. J Clin Endocrinol Metab 2012 Sep 19.

(179) La Marca A., Spada E, Grisendi V, Argento C, Papaleo E, Milani S, et al. Normal serum anti-Mullerian hormone levels in the general female population and the relationship with reproductive history. Eur J Obstet Gynecol Reprod Biol 2012 May 10.

(180) La Marca A., Sighinolfi G, Giulini S, Traglia M, Argento C, Sala C, et al. Normal serum concentrations of anti-Mullerian hormone in women with regular menstrual cycles. Reprod Biomed Online 2010 Oct;21(4):463-9.

(181) van Disseldorp J., Faddy MJ, Themmen AP, de Jong FH, Peeters PH, van der Schouw YT, et al. Relationship of serum antimullerian hormone concentration to age at menopause. J Clin Endocrinol Metab 2008 Jun;93(6):2129-34.

(182) Nelson SM, Messow MC, McConnachie A, Wallace H, Kelsey T, Fleming R, et al. External validation of nomogram for the decline in serum anti-Mullerian hormone in women: a population study of 15,834 infertility patients. Reprod Biomed Online 2011 Aug;23(2):204-6.

(183) Nelson SM, Messow MC, Wallace AM, Fleming R, McConnachie A. Nomogram for the decline in serum antimullerian hormone: a population study of 9,601 infertility patients. Fertil Steril 2011 Feb;95(2):736-41.

(184) Shaw CM, Stanczyk FZ, Egleston BL, Kahle LL, Spittle CS, Godwin AK, et al. Serum antimullerian hormone in healthy premenopausal women. Fertil Steril 2011 Jun 30;95(8):2718-21.

(185) Arbo E, Vetori DV, Jimenez MF, Freitas FM, Lemos N, Cunha-Filho JS. Serum antimullerian hormone levels and follicular cohort characteristics after pituitary suppression in the late luteal phase with oral contraceptive pills. Hum Reprod 2007 Dec;22(12):3192-6.

(186) Steiner AZ, Stanczyk FZ, Patel S, Edelman A. Antimullerian hormone and obesity: insights in oral contraceptive users. Contraception 2010 Mar;81(3):245-8.

(187) Kristensen SL, Ramlau-Hansen CH, Andersen CY, Ernst E, Olsen SF, Bonde JP, et al. The association between circulating levels of antimullerian hormone and follicle number, androgens, and menstrual cycle characteristics in young women. Fertil Steril 2012 Mar;97(3):779-85.

(188) van den Berg MH, van Dulmen-den Broeder E., Overbeek A, Twisk JW, Schats R, van Leeuwen FE, et al. Comparison of ovarian function markers in users of hormonal contraceptives during the hormone-free interval and subsequent natural early follicular phases. Hum Reprod 2010 Jun;25(6):1520-7.

(189) Plante BJ, Cooper GS, Baird DD, Steiner AZ. The impact of smoking on antimullerian hormone levels in women aged 38 to 50 years. Menopause 2010 May;17(3):571-6.

(190) Schuh-Huerta SM, Johnson NA, Rosen MP, Sternfeld B, Cedars MI, Reijo Pera RA. Genetic variants and environmental factors associated with hormonal markers of ovarian reserve in Caucasian and African American women. Hum Reprod 2012 Feb;27(2):594-608.

(191) Dafopoulos A, Dafopoulos K, Georgoulias P, Galazios G, Limberis V, Tsikouras P, et al. Smoking and AMH levels in women with normal reproductive history. Arch Gynecol Obstet 2010 Aug;282(2):215-9.

(192) Streuli I, Fraisse T, Pillet C, Ibecheole V, Bischof P, de ZD. Serum antimullerian hormone levels remain stable throughout the menstrual cycle and after oral or vaginal administration of synthetic sex steroids. Fertil Steril 2008 Aug;90(2):395-400.

(193) Deb S, Campbell BK, Pincott-Allen C, Clewes JS, Cumberpatch G, Raine-Fenning NJ. Quantifying effect of combined oral contraceptive pill on functional ovarian reserve as measured by serum anti-Mullerian hormone and small antral follicle count using three-dimensional ultrasound. Ultrasound Obstet Gynecol 2012 May;39(5):574-80.

(194) Somunkiran A, Yavuz T, Yucel O, Ozdemir I. Anti-Mullerian hormone levels during hormonal contraception in women with polycystic ovary syndrome. Eur J Obstet Gynecol Reprod Biol 2007 Oct;134(2):196-201.

(195) Li HW, Wong CY, Yeung WS, Ho PC, Ng EH. Serum anti-mullerian hormone level is not altered in women using hormonal contraceptives. Contraception 2011 Jun;83(6):582-5.

(196) Dossus L, Kvaskoff M, Bijon A, Fervers B, Boutron-Ruault MC, Mesrine S, et al. Determinants of age at menarche and time to menstrual cycle regularity in the French E3N cohort. Ann Epidemiol 2012 Oct;22(10):723-30.

(197) Gold EB. The timing of the age at which natural menopause occurs. Obstet Gynecol Clin North Am 2011 Sep;38(3):425-40.

(198) Morris DH, Jones ME, Schoemaker MJ, McFadden E, Ashworth A, Swerdlow AJ. Body mass index, exercise, and other lifestyle factors in relation to age at natural menopause: analyses from the breakthrough generations study. Am J Epidemiol 2012 May 15;175(10):998-1005.

(199) Ganesh SK, Tragante V, Guo W, Guo Y, Lanktree MB, Smith EN, et al. Loci influencing blood pressure identified using a cardiovascular gene-centric array. Hum Mol Genet 2013 Jan 8.

(200) Verschuren WM, Blokstra A, Picavet HS, Smit HA. Cohort profile: the Doetinchem Cohort Study. Int J Epidemiol 2008 Dec;37(6):1236-41.

(201) WHO. Waist Circumference and Waist–Hip Ratio: Report of a WHO Expert Consultation. 8-12-2008. Geneva, World Health Organization (WHO).

(202) Pols MA, Peeters PH, Ocke MC, Slimani N, Bueno-De-Mesquita HB, Collette HJ. Estimation of reproducibility and relative validity of the questions included in the EPIC Physical Activity Questionnaire. Int J Epidemiol 1997;26 Suppl 1:S181-S189.

(203) Wareham NJ, Jakes RW, Rennie KL, Schuit J, Mitchell J, Hennings S, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. Public Health Nutr 2003 Jun;6(4):407-13.

(204) Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992 Jul;11(10):1305-19.

(205) Stasinopoulos DM, Rigby RA. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. Journal of Statistical Software 23[7]. 28-11-2007. 19-07-2007, American Statistical Association.

(206) Bentzen JG, Forman JL, Larsen EC, Pinborg A, Johannsen TH, Schmidt L, et al. Maternal menopause as a predictor of anti-Mullerian hormone level and antral follicle count in daughters during reproductive age. Hum Reprod 2012 Nov 6.

(207) van Noord PA, Dubas JS, Dorland M, Boersma H, te VE. Age at natural menopause in a population-based screening cohort: the role of menarche, fecundity, and lifestyle factors. Fertil Steril 1997 Jul;68(1):95-102.

(208) Gold EB, Bromberger J, Crawford S, Samuels S, Greendale GA, Harlow SD, et al. Factors associated with age at natural menopause in a multiethnic sample of midlife women. Am J Epidemiol 2001 May 1;153(9):865-74.

(209) Pache TD, Wladimiroff JW, de Jong FH, Hop WC, Fauser BC. Growth patterns of nondominant ovarian follicles during the normal menstrual cycle. Fertil Steril 1990 Oct;54(4):638-42.

(210) Taieb J, Grynberg M, Pierre A, Arouche N, Massart P, Belville C, et al. FSH and its second messenger cAMP stimulate the transcription of human anti-Mullerian hormone in cultured granulosa cells. Mol Endocrinol 2011 Apr;25(4):645-55.

(211) Hagen CP, Sorensen K, Anderson RA, Juul A. Serum levels of antimullerian hormone in early maturing girls before, during, and after suppression with GnRH agonist. Fertil Steril 2012 Nov;98(5):1326-30.

(212) Jensen AM, Brocks V, Holm K, Laursen EM, Muller J. Central precocious puberty in girls: internal genitalia before, during, and after treatment with long-acting gonadotropin-releasing hormone analogues. J Pediatr 1998 Jan;132(1):105-8.

(213) Gougeon A. Regulation of ovarian follicular development in primates: facts and hypotheses. Endocr Rev 1996 Apr;17(2):121-55.

(214) Nelson SM, Stewart F, Fleming R, Freeman DJ. Longitudinal assessment of antimullerian hormone during pregnancy-relationship with maternal adiposity, insulin, and adiponectin. Fertil Steril 2010 Mar 1;93(4):1356-8.

(215) Hagen CP, Vestergaard S, Juul A, Skakkebaek NE, Andersson AM, Main KM, et al. Low concentration of circulating antimullerian hormone is not predictive of reduced fecundability in young healthy women: a prospective cohort study. Fertil Steril 2012 Dec;98(6):1602-8.

(216) Freeman EW, Sammel MD, Lin H, Gracia CR. Anti-Mullerian Hormone as a Predictor of Time to Menopause in Late Reproductive Age Women. J Clin Endocrinol Metab 2012 Feb 29.

(217) Bragg JM, Kuzawa CW, Agustin SS, Banerjee MN, McDade TW. Age at menarche and parity are independently associated with Anti-Mullerian hormone, a marker of ovarian reserve, in Filipino young adult women. Am J Hum Biol 2012 Nov;24(6):739-45.

(218) Adena MA, Gallagher HG. Cigarette smoking and the age at menopause. Ann Hum Biol 1982 Mar;9(2):121-30.

(219) Mattison DR, Thorgeirsson SS. Smoking and industrial pollution, and their effects on menopause and ovarian cancer. Lancet 1978 Jan 28;1(8057):187-8.

(220) Park HT, Cho GJ, Ahn KH, Shin JH, Kim YT, Hur JY, et al. Association of insulin resistance with anti-Mullerian hormone levels in women without polycystic ovary syndrome (PCOS). Clin Endocrinol (Oxf) 2010 Jan;72(1):26-31.

(221) Skalba P, Cygal A, Madej P, bkowska-Huc A, Sikora J, Martirosian G, et al. Is the plasma anti-Mullerian hormone (AMH) level associated with body weight and metabolic, and hormonal disturbances in women with and without polycystic ovary syndrome? Eur J Obstet Gynecol Reprod Biol 2011 Oct;158(2):254-9.

(222) Halawaty S, ElKattan E, Azab H, ElGhamry N, Al-Inany H. Effect of obesity on parameters of ovarian reserve in premenopausal women. J Obstet Gynaecol Can 2010 Jul;32(7):687-90.

(223) Freeman EW, Gracia CR, Sammel MD, Lin H, Lim LC, Strauss JF, III. Association of anti-mullerian hormone levels with obesity in late reproductive-age women. Fertil Steril 2007 Jan;87(1):101-6.

(224) Rustamov O, Smith A, Roberts SA, Yates AP, Fitzgerald C, Krishnan M, et al. Anti-Mullerian hormone: poor assay reproducibility in a large cohort of subjects suggests sample instability. Hum Reprod 2012 Jul 9.

(225) Overbeek A, Broekmans FJ, Hehenkamp WJ, Wijdeveld ME, van Disseldorp J., van Dulmen-den Broeder E., et al. Intra-cycle fluctuations of anti-Mullerian hormone in normal women with a regular cycle: a re-analysis. Reprod Biomed Online 2012 Jun;24(6):664-9.

(226) La Marca A., Malmusi S, Giulini S, Tamaro LF, Orvieto R, Levratti P, et al. Anti-Mullerian hormone plasma levels in spontaneous menstrual cycle and during treatment with FSH to induce ovulation. Hum Reprod 2004 Dec;19(12):2738-41.

(227) Weenen C, Laven JS, Von Bergh AR, Cranfield M, Groome NP, Visser JA, et al. Anti-Mullerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. Mol Hum Reprod 2004 Feb;10(2):77-83.

(228) Kelsey TW, Anderson RA, Wright P, Nelson SM, Wallace WH. Data-driven assessment of the human ovarian reserve. Mol Hum Reprod 2012 Feb;18(2):79-87.

(229) Kelsey TW, Wright P, Nelson SM, Anderson RA, Wallace WH. A validated model of serum anti-mullerian hormone from conception to menopause. PLoS One 2011;6(7):e22024.

(230) Sowers MR, Eyvazzadeh AD, McConnell D, Yosef M, Jannausch ML, Zhang D, et al. Anti-mullerian hormone and inhibin B in the definition of ovarian aging and the menopause transition. J Clin Endocrinol Metab 2008 Sep;93(9):3478-83.

(231) Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 2002 Dec;5(6B):1113-24.

(232) Boker LK, van Noord PA, van der Schouw YT, Koot NV, Bueno de Mesquita HB, Riboli E, et al. Prospect-EPIC Utrecht: study design and characteristics of the cohort population. European Prospective Investigation into Cancer and Nutrition. Eur J Epidemiol 2001;17(11):1047-53.

(233) Kok HS, van Asselt KM, van der Schouw YT, Grobbee DE, te Velde ER, Pearson PL, et al. Subfertility reflects accelerated ovarian ageing. Hum Reprod 2003 Mar;18(3):644-8.

(234) Jones MC, Faddy MJ. A skew extension of the t-distribution, with applications. Journal of the Royal Statistical Society: Series B 65, 159-174. 2003.

(235) van Disseldorp J., Lambalk CB, Kwee J, Looman CW, Eijkemans MJ, Fauser BC, et al. Comparison of inter- and intra-cycle variability of anti-Mullerian hormone and antral follicle counts. Hum Reprod 2010 Jan;25(1):221-7.

(236) Sowers M, McConnell D, Gast K, Zheng H, Nan B, McCarthy JD, et al. Anti-Mullerian hormone and inhibin B variability during normal menstrual cycles. Fertil Steril 2010 Sep;94(4):1482-6.

(237) Wallace AM, Faye SA, Fleming R, Nelson SM. A multicentre evaluation of the new Beckman Coulter anti-Mullerian hormone immunoassay (AMH Gen II). Ann Clin Biochem 2011 Jul;48(Pt 4):370-3.

(238) Kumar A, Kalra B, Patel A, McDavid L, Roudebush WE. Development of a second generation anti-Mullerian hormone (AMH) ELISA. J Immunol Methods 2010 Oct 31;362(1-2):51-9.

(239) Wallace WH, Kelsey TW. Human ovarian reserve from conception to the menopause. PLoS One 2010;5(1):e8772.

(240) Laven JS, Mulders AG, Visser JA, Themmen AP, de Jong FH, Fauser BC. Anti-Mullerian hormone serum concentrations in normoovulatory and anovulatory women of reproductive age. J Clin Endocrinol Metab 2004 Jan;89(1):318-23.

(241) Tehrani FR, Solaymani-Dodaran M, Hedayati M, Azizi F. Is polycystic ovary syndrome an exception for reproductive aging? Hum Reprod 2010 Jul;25(7):1775-81.

(242) den Tonkelaar I. Validity and reproducibility of self-reported age at menopause in women participating in the DOM-project. Am J Human Biol 1997 Jun;27(2):117-23.

(243) Colditz GA, Stampfer MJ, Willett WC, Stason WB, Rosner B, Hennekens CH, et al. Reproducibility and validity of self-reported menopausal status in a prospective cohort study. Am J Epidemiol 1987 Aug;126(2):319-25.

(244) Li HW, Ng EH, Wong BP, Anderson RA, Ho PC, Yeung WS. Correlation between three assay systems for anti-Mullerian hormone (AMH) determination. J Assist Reprod Genet 2012 Dec;29(12):1443-6.

(245) Powell CM, Taggart RT, Drumheller TC, Wangsa D, Qian C, Nelson LM, et al. Molecular and cytogenetic studies of an X;autosome translocation in a patient with premature ovarian failure and review of the literature. Am J Med Genet 1994 Aug 1;52(1):19-26.

(246) Murabito JM, Yang Q, Fox CS, Cupples LA. Genome-wide linkage analysis to age at natural menopause in a community-based sample: the Framingham Heart Study. Fertil Steril 2005 Dec;84(6):1674-9.

(247) Freeman EW, Sammel MD, Lin H, Boorman DW, Gracia CR. Contribution of the rate of change of antimullerian hormone in estimating time to menopause for late reproductive-age women. Fertil Steril 2012 Nov;98(5):1254-9.

(248) Tehrani FR, Solaymani-Dodaran M, Tohidi M, Gohari MR, Azizi F. Modeling age at menopause using serum concentration of anti-mullerian hormone. J Clin Endocrinol Metab 2013 Feb;98(2):729-35.

(249) Dolleman M, Faddy MJ, van DJ, van der Schouw YT, Messow CM, Leader B, et al. The Relationship Between Anti-Mullerian Hormone in Women Receiving Fertility Assessments and Age at Menopause in Subfertile Women: Evidence From Large Population Studies. J Clin Endocrinol Metab 2013 Mar 18.

(250) Miltenburg GA, Peeters PH, Fracheboud J, Collette HJ. Seventeen-year evaluation of breast cancer screening: the DOM project, The Netherlands. Diagnostisch Onderzoek (investigation) Mammacarcinoom. Br J Cancer 1998 Oct;78(7):962-5.

(251) Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011 Jan 15;30(1):11-21.

(252) Pickering JW, Endre ZH. New metrics for assessing diagnostic potential of candidate biomarkers. Clin J Am Soc Nephrol 2012 Aug;7(8):1355-64.

(253) Kevenaar ME, Themmen AP, Rivadeneira F, Uitterlinden AG, Laven JS, van Schoor NM, et al. A polymorphism in the AMH type II receptor gene is associated with age at menopause in interaction with parity. Hum Reprod 2007 Sep;22(9):2382-8.

(254) Dolleman M, Verschuren WM, Eijkemans MJ, Dolle ME, Jansen EH, Broekmans FJ, et al. Reproductive and Lifestyle Determinants of Anti-Mullerian Hormone in a Large Population-based Study. J Clin Endocrinol Metab 2013 Mar 26.

(255) de Graaff AA, Land JA, Kessels AG, Evers JL. Demographic age shift toward later conception results in an increased age in the subfertile population and an increased demand for medical care. Fertil Steril 2011 Jan;95(1):61-3.

(256) Stoop D, Nekkebroeck J, Devroey P. A survey on the intentions and attitudes towards oocyte cryopreservation for non-medical reasons among women of reproductive age. Hum Reprod 2011 Mar;26(3):655-61.

(257) Lee MM, Donahoe PK, Hasegawa T, Silverman B, Crist GB, Best S, et al. Mullerian inhibiting substance in humans: normal levels from infancy to adulthood. J Clin Endocrinol Metab 1996 Feb;81(2):571-6.

(258) Gruijters MJ, Visser JA, Durlinger AL, Themmen AP. Anti-Mullerian hormone and its role in ovarian function. Mol Cell Endocrinol 2003 Dec 15;211(1-2):85-90.

(259) Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. Stat Med 2006 Oct 30;25(20):3474-86.

(260) La MA, Sighinolfi G, Papaleo E, Cagnacci A, Volpe A, Faddy MJ. Prediction of age at menopause from assessment of ovarian reserve may be improved by using body mass index and smoking status. PLoS One 2013;8(3):e57005.

(261) Fleming R, Nelson SM. Reproducibility of AMH. Hum Reprod 2012 Dec;27(12):3639-41.

(262) La MA, Stabile G, Artenisio AC, Volpe A. Serum anti-Mullerian hormone throughout the human menstrual cycle. Hum Reprod 2006 Dec;21(12):3103-7.

(263) Tsepelidis S, Devreker F, Demeestere I, Flahaut A, Gervy C, Englert Y. Stable serum levels of anti-Mullerian hormone during the menstrual cycle: a prospective study in normo-ovulatory women. Hum Reprod 2007 Jul;22(7):1837-40.

(264) Hehenkamp WJ, Looman CW, Themmen AP, de Jong FH, te Velde ER, Broekmans FJ. Anti-Mullerian hormone levels in the spontaneous menstrual cycle do not show substantial fluctuation. J Clin Endocrinol Metab 2006 Oct;91(10):4057-63.

(265) Han X. Pre-mixing samples with assay buffer is an essential pre-requisite for reproducible Antimullerian Hormone (AMH) measurement using the Beckman Coulter Gen II assay (Gen II). McShane M, Sahertian R, White C, Ledger WL, editors. 2013. Ref Type: Generic

(266) Iliodromiti S, Kelsey TW, Anderson RA, Nelson SM. Can Anti-Mullerian hormone predict the diagnosis of polycystic ovary syndrome? A systematic review and meta-analysis of extracted data. J Clin Endocrinol Metab 2013 Jun 17.

(267) Steiner AZ, Herring AH, Kesner JS, Meadows JW, Stanczyk FZ, Hoberman S, et al. Antimullerian hormone as a predictor of natural fecundability in women aged 30-42 years. Obstet Gynecol 2011 Apr;117(4):798-804.

(268) Olivennes F, Howles CM, Borini A, Germond M, Trew G, Wikland M, et al. Individualizing FSH dose for assisted reproduction using a novel algorithm: the CONSORT study. Reprod Biomed Online 2009 Feb;18(2):195-204.

(269) Lekamge DN, Lane M, Gilchrist RB, Tremellen KP. Increased gonadotrophin stimulation does not improve IVF outcomes in patients with predicted poor ovarian reserve. J Assist Reprod Genet 2008 Nov;25(11-12):515-21.

(270) Yarde F, Voorhuis M, Dolleman M, Knauff EA, Eijkemans MJ, Broekmans FJ. Antimullerian hormone as predictor of reproductive outcome in subfertile women with elevated basal follicle-stimulating hormone levels: a follow-up study. Fertil Steril 2013 Jun 10.

(271) Brodin T, Hadziosmanovic N, Berglund L, Olovsson M, Holte J. Antimullerian hormone levels are strongly associated with live-birth rates after assisted reproduction. J Clin Endocrinol Metab 2013 Mar;98(3):1107-14.

(272) Fleming R, Nelson SM. Reproducibility of AMH. Hum Reprod 2012 Dec;27(12):3639-41.

(273) Fleming R, Fairbairn C, Blaney C, Lucas D, Gaudoin M. Stability of AMH measurement in blood and avoidance of proteolytic changes. Reprod Biomed Online 2013 Feb;26(2):130-2.

(274) Dondorp W, de WG, Pennings G, Shenfield F, Devroey P, Tarlatzis B, et al. Oocyte cryopreservation for age-related fertility loss. Hum Reprod 2012 May;27(5):1231-7.

(275) van Wayenburg CA, van der Schouw YT, van Noord PA, Peeters PH. Age at menopause, body mass index, and the risk of colorectal cancer mortality in the Dutch Diagnostisch Onderzoek Mammacarcinoom (DOM) cohort. Epidemiology 2000 May;11(3):304-8.

(276) Kritz-Silverstein D, Barrett-Connor E. Early menopause, number of reproductive years, and bone mineral density in postmenopausal women. Am J Public Health 1993 Jul;83(7):983-8.

(277) van der Schouw YT, van der GY, Steyerberg EW, Eijkemans JC, Banga JD. Age at menopause as a risk factor for cardiovascular mortality. Lancet 1996 Mar 16;347(9003):714-8.

(278) Kok HS, Kuh D, Cooper R, van der Schouw YT, Grobbee DE, Wadsworth ME, et al. Cognitive function across the life course and the menopausal transition in a British birth cohort. Menopause 2006 Jan;13(1):19-27.

(279) Bleil ME, Gregorich SE, McConnell D, Rosen MP, Cedars MI. Does accelerated reproductive aging underlie premenopausal risk for cardiovascular disease? Menopause 2013 Apr 23.

(280) D'Agostino RB, Sr., Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008 Feb 12;117(6):743-53.

(281) D'Agostino RB, Russell MW, Huse DM, Ellison RC, Silbershatz H, Wilson PW, et al. Primary and subsequent coronary risk appraisal: new results from the Framingham study. Am Heart J 2000 Feb;139(2 Pt 1):272-81.

(282) Sobradillo P, Pozo F, Agusti A. P4 medicine: the future around the corner. Arch Bronconeumol 2011 Jan;47(1):35-40.

(283) de Vet A, Laven JS, de Jong FH, Themmen AP, Fauser BC Antimüllerian hormone serum levels: a putative marker for ovarian aging. Fertil Steril 2002 : 77(2):357-62

Chapter 11

List of Publications
Acknowledgements
About the Author

List of Publications

1. Broer SL, Mol B, **Dolleman M**, Fauser BC, Broekmans FJ. The role of anti-Mullerian hormone assessment in assisted reproductive technology outcome. Curr Opin Obstet Gynecol 2010 Jun;22(3):193-201.

2. Broer SL, **Dolleman M**, Opmeer BC, Fauser BC, Mol BW, Broekmans FJ. AMH and AFC as predictors of excessive response in controlled ovarian hyperstimulation: a meta-analysis. Hum Reprod Update 2011 Jan;17(1):46-54.

3. Broer SL, **Dolleman M**, van DJ, Broeze KA, Opmeer BC, Bossuyt PM, et al. Prediction of an excessive response in in vitro fertilization from patient characteristics and ovarian reserve tests and comparison in subgroups: an individual patient data meta-analysis. Fertil Steril 2013 Aug;100(2):420-9.

4. Broer SL, van DJ, Broeze KA, **Dolleman M**, Opmeer BC, Bossuyt P, 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. Hum Reprod Update 2013 Jan;19(1):26-36.

5. **Dolleman M**, Verschuren WM, Eijkemans MJ, Dolle ME, Jansen EH, Broekmans FJ, et al. Reproductive and lifestyle determinants of anti-Mullerian hormone in a large population-based study. J Clin Endocrinol Metab 2013 May;98(5):2106-15.

6. **Dolleman M**, Faddy MJ, van DJ, van der Schouw YT, Messow CM, Leader B, et al. The relationship between anti-Mullerian hormone in women receiving fertility assessments and age at menopause in subfertile women: evidence from large population studies. J Clin Endocrinol Metab 2013 May;98(5):1946-53.

7. Yarde F, Voorhuis M, **Dolleman M**, Knauff EA, Eijkemans MJ, Broekmans FJ. Antimullerian hormone as predictor of reproductive outcome in subfertile women with elevated basal follicle-stimulating hormone levels: a follow-up study. Fertil Steril 2013 Sep;100(3):831-8.

Acknowledgements

Whoop-there it is!! This thesis would not have been possible without the cooperation, feedback and support from everyone that was involved, both directly and indirectly, in the writing of this book. First of all, a big thank you to all the women that participated in the many studies that were used in writing this thesis.

Professor Broekmans, dear Frank. Little did I know what was in store for me when I came to you as a medical student looking for an interesting project...I definitely did not think that it would ultimately result in this book that lies before you now. Frank, you put me up to quite a challenge when we agreed upon the short timeframe in which I would have to finish my thesis, but we managed! Thank you for sharing your knowledge, your experience and your enthusiasm with me, and thanks for helping me get to where I really wanted to be- in training to become a gynaecologist. It has been great working with you!

Professor Mol, dear Ben Willem. You recognized the researcher in me before I knew there was one. Thank you for encouraging me to continue my academic career as a PhD researcher, for creating time in which I could round off some of my projects but mostly for your always critical view of results and the conclusions drawn from them.

Professor van der Schouw, dear Yvonne. I have learnt so much from you in such a short time! You reminded me to go back to the basics when I had already started somewhere half way. You always knew exactly what the weak spots were in my papers, and came with concrete ways to improve them. Most importantly, your meticulousness and the speed in which you gave feedback enabled me to be done within the short timeframe that I had to finish my thesis. Thank you.

Dr. Disseldorp, dear Jeroen. Eight years ago, as I enthusiastically served birthday drinks to my fellow students at the University College Bar I would have never thought that the inventor of this notorious 'Dissel' would be my co-promotor years later! Due to the efforts with which you set up the IPD-studies you created a path that first Simone and then I could further pave. Thank you for both the basis upon which I could go further and your help with the studies presented in this book.

Dr. Eijkemans, dear René. Quite frankly, without your statistical help this book would not exist. Thank you for all that you have taught me and the patience with which you did so. I really appreciate the time which you managed to create for me in your already busy agenda. You made statistics less scary and more understandable and always in a fun way.

Dr. Broer, dear Simone. You took me under your wing as a medical student and taught me the basis upon which I could independently continue with scientific search. You encouraged me to continue with research outside of the designated time during med-school which inevitably helped in getting me a spot as a PhD researcher. Thanks a million.

Dear Professor Bossuyt and dr. Opmeer. Thank you for all your help in setting up the IPD-analyses.

Dear EXPORT collaborators, thank you for sharing your data which made the IPD-analyses possible.

Dear Professor Nelson, Professor Faddy, dr. Leader and Professor Peeters, it was great working with such motivated, friendly and knowledgeable people. Thank you.

Dear Professor Lambalk, prof Laven and dr. Tehrani thank you for sharing your data with me which made chapters 6 and 7 possible.

Dear dr. Verschuren, thank you for allowing me to work with the Doetinchem Cohort Study data and your help with chapters 5 and 8.

To all the other co-authors that I have not yet mentioned, thank you for your participation in the various articles presented in this thesis.

Dear Ellis, Ingrid and Tessa. Behind every great professor hides a great secretary! You three are always helpful and always up to speed. Thanks.

To my fellow researchers and ex-PhD students from the AZU and WKZ. Thanks for all the good times, the advice, the fun Monday lunches, and the eventful international conferences. It has been fun! Before you know it, it will be your turn to defend your thesis, so enjoy the PhD lifestyle while you can and I will be there to toast to your success in the Academiegebouw in the coming years!

To my gynaecology residency colleagues. Two assistenten-weekenden down and six to go! I can't wait to start in January!

To my colleagues in Gelre Apeldoorn, what a great hospital to start my gynaecology career in! I am already looking forward to the next two years although I might have to start my own little wind-mill farm to justify my carbon footprint from driving up and down from Amsterdam every day. Thanks for your support and understanding these last months in the final sprint to finishing this thesis.

Dear SUMMAatjes, the embarrassing intern moments, Friday offwards and evenings in the dilapidated "zusterflat" seem like a lifetime ago now that we are all off to a flying start in all different areas of medicine. I hope to see you all lots in the near future and wish you all the best.

Dearest Fe, Yv and Q, the relaxed evenings in our "chillpaleisje" make the commute to and from work every day totally worth it. Thanks for all the fun times and your well-needed pep-talks these past months!

Dear Soha and Shamim, I owe the smashing cover of this book to you! Thank you so much. You are two of the most humble and kind-hearted people I know and I can't wait to see you in December!

Dear Marieke, Isabel, Laurien, and Eva, my "sista's from different mista's". You really are all like sisters to me and have played a huge part in me feeling more at home in Holland. Whether it's a shoulder to cry on or a buddy to laugh with I know I can count on you. Thanks for always being there for me.

To all my UCU friends, thanks for all the fun times, the good chats and the epic trips these past years. I'm hanging around for a while so you haven't gotten rid of me yet!

To all my friends from Kenya, the fact that our friendship can overcome both distance and time says enough. Asanteni & see you soon!

A big thanks to the people who gave me a home away from home, tante Maart, Johan and Aat and Jaap and Marlis.

Lieve paranimfen, Felicia and Blanka, you two are absolute rockstars! I never thought two colleagues could turn into such good friends, although in retrospect it seems inevitable considering our mutual love for Africa,

sunshine and summer festivals and our mutual loathing of winter. Thanks for all your help, your pep-talks, the amazing surprise 'stellingen diner' and also for reminding me of the importance of not only working hard but playing hard as well.

Felicia, m'Africa Kabisa! You are a great flatmate an awesome travelbuddy and an amazing friend. I will forever miss the train rides to work sharing stories about life and love, reading horoscopes or brainstorming about our various research projects. Let's add many more destinations to the already groovy list of Stockholm, Istanbul, Nairobi, London, Verona and Paris. Blanka, Mzungu! I still can't believe we were born in the same hospital in Nairobi, a mere week apart. You are a super loyal friend and I value that about you. I know I can always count on you for a sentimental chat about Africa, a tasty meal and a partner in crime for Valtifest (or any other dress up party in town)! Ladies.....I'm so lucky to be backed by two such intelligent, hard-working and fun-loving women like you today. I couldn't have done it without you!

Lieve "Dollemannen clan", a family tie like ours is something truly special and something we should all work at to keep strong in the future. Thanks for being there for me.

Dearest Nicolette & Sanne en Willem & Francien, even though my life here couldn't be more different than your lives in Gambella and Nairobi I know I can count on you for anything. I love you lots and lots and lots and wish you all good things in life and more!

Allerliefste papa en mama. Your endless unconditional love and support is always felt despite the 6662 km distance between us. Although my life has been here and will be here for the unforeseeable future, Kenya is and will always be home. Here's to the wind, your health and your happiness!

About the Author

Madeleine Dólleman was born in Nairobi, Kenya on the 1st of November 1985. She grew up in Nairobi with her parents Willem and Marie-Jose, her sister Nicolette and her brother Willem. She attended the International School of Kenya from which she graduated in 2004. Afterwards she moved to the Netherlands where she attended University College Utrecht and received her bachelor's degree in medical sciences. After graduating she was accepted into SUMMA, a selective medical masters program which prepared her for both clinical work and research in an academic setting. During her university years Madeleine took the opportunity to spend a lot of time abroad amongst which a 6 month university exchange-program in Sydney, Australia, an internship in Kenya's 2nd largest refugee-camp Kakuma and an inspiring 4 month journey through northern Kenya with a mobile clinic team that focused on primary health care and family planning. During medical school Madeleine got involved in research at the department of reproductive medicine with prof. dr. F.J. Broekmans and she was offered a spot as PhD researcher after receiving her medical degree in December 2011. In May 2013 she started working as a gynaecology resident at the department of obstetrics and gynaecology at Gelre Ziekenhuizen Apeldoorn where she will also start her specialization in January 2014.