

The predictive accuracy of anti-Müllerian hormone for live birth after assisted conception: a systematic review and meta-analysis of the literature

Stamatina Iliodromiti^{1,*}, Thomas W. Kelsey², Olivia Wu³,
Richard A. Anderson^{4,†}, and Scott M. Nelson^{1,†}

¹Reproductive and Maternal Medicine, School of Medicine, University of Glasgow, G12 8QQ Glasgow, UK ²School of Computer Science, University of St Andrews, KY16 9AJ St Andrews, UK ³Health Economics and Health Technology, Institute of Health and Wellbeing, University of Glasgow, G12 8RZ Glasgow, UK ⁴Medical Research Council Centre for Reproductive Health, Queens Medical Research Institute, University of Edinburgh, EH16 4TJ Edinburgh, UK

*Correspondence address. Reproductive and Maternal Medicine, 3rd Floor McGregor Building, University of Glasgow, G12 8QQ Glasgow, UK. E-mail: stamatina.iliodromiti@glasgow.ac.uk

Submitted on October 1, 2013; resubmitted on December 12, 2013; accepted on January 14, 2014

TABLE OF CONTENTS

- Introduction
 - Methods
 - Eligibility criteria
 - Search
 - Study selection and data collection
 - Quality appraisal of the selected studies
 - Data synthesis and analysis
 - Ethics
 - Results
 - Search results
 - Pooled estimates
 - Study quality assessment and publication bias
 - Discussion
 - Clinical application
 - Strengths and limitations
 - Conclusion
-

BACKGROUND: Anti-Müllerian hormone (AMH) is an established marker of ovarian reserve and a good predictor of poor or excessive ovarian response after controlled hyperstimulation. However, it is unclear whether it can predict the ultimate outcome of assisted conception, live birth. We undertook a systematic review and meta-analysis to examine whether AMH is a predictor of live birth in women undergoing assisted conception.

METHODS: The study was conducted according to the PRISMA guidelines. PubMed, Embase, Medline, Web of Knowledge and the Cochrane trial register and unpublished literature were searched. Studies fulfilling the eligibility criteria were included in the systematic review and those with extractable data were included in the meta-analysis. Quality assessment was performed with the QUADAS 2 checklist. A summary estimate of diagnostic odds ratio (DOR) was derived using the random effects model for binary data. A hierarchical summary receiver operating characteristic model provided pooled estimates before and after adjusting for age and AMH assay as covariates.

† Joint senior authors.

RESULTS: Out of 361 non-duplicate studies, 47 were selected; 17 met the eligibility criteria and 13 had extractable data and thus were included in the meta-analysis. Three out of the 13 studies included only women with expected low ovarian reserve and were analysed individually from the remaining 10 to minimize heterogeneity. The DOR for women with unknown ovarian reserve ($n = 5764$ women) was 2.39 (95% confidence interval (CI): 1.85–3.08). After adjustment for age the DOR was little changed at 2.48 (95% CI: 1.81–3.22) and the DOR adjusted for AMH assay was almost identical at 2.42 (95% CI: 1.86–3.14). For women with expected low ovarian reserve ($n = 542$ women) the DOR was 4.63 (95% CI: 2.75–7.81).

CONCLUSIONS: AMH, independently of age, has some association with predicting live birth after assisted conception and may be helpful when counselling couples before undergoing fertility treatment. However, its predictive accuracy is poor.

Key words: anti-Müllerian hormone / IVF / live birth / diagnostic odds ratio / hierarchical summary receiver operator characteristic curve

Introduction

The age-related decline in oocyte quantity and quality (Nelson *et al.*, 2013b) underpins the decline in success rates and prospect of live birth after assisted conception with advancing maternal age (Oudendijk *et al.*, 2012). Age alone however is of limited accuracy in predicting live birth; thus there is a need for improved prediction for individualization of counselling. The substantial heterogeneity in the size of the ovarian reserve at any given age (Wallace and Kelsey, 2010) results in marked inter-individual variation in ovarian response despite optimal ovarian stimulation. Analysis of this heterogeneity may provide insights into understanding individual fertility and how it changes with age, and it is also a likely source of clinically useful biomarkers. A variety of ovarian reserve tests have been developed and their predictive capacity for ovarian response examined. In recent systematic reviews, individual patient data meta-analysis and international multicentre trials, anti-Müllerian hormone (AMH) has been confirmed as the current best biomarker for prediction of poor and excessive ovarian response (Broekmans *et al.*, 2006; Broer *et al.*, 2009, 2011, 2013; Nelson *et al.*, 2009; Anckaert *et al.*, 2012).

Given the strength of the relationships with oocyte yield, the association of AMH with pregnancy after assisted conception has been examined, but results were inconclusive. Some studies have concluded that AMH is not associated with pregnancy (Broekmans *et al.*, 2006; van Rooij *et al.*, 2006) while others have found a positive association (Nelson *et al.*, 2007; Honnma *et al.*, 2013). A recent individual patient data meta-analysis in 1008 patients undergoing fertility treatment demonstrated a weak association of AMH with ongoing pregnancy (Broer *et al.*, 2013). As live birth is the ultimate outcome of assisted conception, clarification of whether AMH is predictive of this outcome is warranted.

The observational studies that have examined the association of AMH and live birth have either been small (Nelson *et al.*, 2007; Lee *et al.*, 2009; Majumder *et al.*, 2010) or restricted to specific subpopulations of infertility patients (Gleicher *et al.*, 2010; Arce *et al.*, 2013). To provide an accurate estimate of the effect size we undertook a systematic review and meta-analysis of all eligible studies to examine whether AMH can predict live birth in women undergoing assisted conception.

Methods

This study was conducted according to the PRISMA guidelines (Liberati *et al.*, 2009) and followed a structured protocol established among the authors prior to the start of the literature search.

Eligibility criteria

Studies were included if they met the following criteria: (i) the study population included women of reproductive age undergoing IVF or ICSI, with any stimulation protocol; (ii) serum AMH was measured in all study participants before ovarian stimulation; (iii) the clinical outcome of live birth was recorded for all participants; and (iv) any study design except case reports. Thus, studies referring to intrauterine insemination were excluded as were studies referring to on-going pregnancy rate but not live birth rate. Studies referring to follicular fluid AMH or oocyte donation programmes were excluded. Our search included studies published up to August 2013 and there was no language restriction.

Search

The following electronic databases were searched: PubMed, Embase, Medline, Web of Knowledge and the Cochrane trial register. Search terms for live birth (MeSH; live birth, ongoing pregnancy, pregnancy) and key words ‘anti-müllerian hormone’, ‘müllerian-inhibiting substance’, or ‘müllerian-inhibiting factor’ were combined with a search filter for studies related to humans. The abstracts of all studies identified were screened by two researchers (S.I. and S.M.N.). Any studies including data on AMH and live birth or other assisted conception outcomes were read in full. The reference lists of the selected papers were hand-searched in order to identify potentially relevant papers. Grey literature was also searched via the open-grey website.

Study selection and data collection

Subsequently, two researchers (S.I. and S.M.N.) carefully read and independently judged all selected articles. If a study fulfilled the eligibility criteria, it was included in the systematic review. If the study provided extractable information about a cut-off of AMH which was associated with a higher or lower chance of live birth or the sensitivity, specificity or area under the curve (AUC) of the receiver operating characteristic curve of AMH in predicting the chance of live birth, it was selected for the meta-analysis. Any disagreement between the two researchers was resolved with discussion. If a study was selected for the systematic review but did not provide data that could be included in the meta-analysis, the authors were contacted via email. If the authors did not reply or the relevant information was not available, the studies were not included in the meta-analysis. If the authors did not provide the information asked, but the relevant information was extractable using a reverse engineering technique through Plot digitizer, a computer software programme which can extract data from published plots, the articles and data therein were used in the meta-analysis.

For each study the first author, year of publication, number of cycles, number of patients, stimulation protocol, mean/median age of the patients, suggested cut-off point of AMH (converted to ng/ml using the conversion formula $\text{ng/ml} = 7.14 \text{ pmol/l}$), AMH assay, number of live births among

the patients with low or high serum AMH (below or above the cut-off point), study design and patient selection were extracted.

Quality appraisal of the selected studies

Each selected study was further assessed according to the QUADAS 2 checklist to assess the risk of bias and the applicability of primary diagnostic accuracy studies (Whiting et al., 2011). It consists of four main domains testing the risk of bias (low, high or unclear) in patient selection, index test, reference test and flow of each study.

We tried to minimize the risk of publication bias by a comprehensive search strategy which included unpublished (grey) literature. The risk of publication bias and potential small study effect was visually assessed by constructing a funnel plot which plots estimates of diagnostic accuracy against statistical precision (Sterne et al., 2011). In addition, we performed a linear regression of log diagnostic ratios on the inverse root of effective sample sizes as a test for funnel plot asymmetry, where a non-zero slope coefficient ($P < 0.10$) is suggestive of significant asymmetry and small study bias (Deeks et al., 2005).

Data synthesis and analysis

The statistical analysis was undertaken using the Stata/SE (version 12.1, Stata Corp, USA) and SAS/STAT software. Pooling of the data related to the number of live births among the study participants with AMH below and above a cut-off point used the random effects model for binary data and provided a summary estimate of diagnostic odds ratio (DOR) and 95% confidence intervals (CI). The DOR summarizes the diagnostic accuracy of the AMH tests and can take values from 0 to infinity. It expresses the odds of detecting AMH above a cut-off point (positive test result) among women with live births relative to the odds of detecting AMH above the cut-off among women without live births (Glas et al., 2003).

Heterogeneity resulting from true diagnostic accuracy not being identical in each study was quantified by the I-squared measure (Higgins et al., 2003). Sensitivity analysis was performed for studies with similar study populations (women with unknown ovarian reserve versus women with reduced ovarian reserve) to assess whether the DOR varies according to patients' characteristics. In addition, the summary ROC curve, sensitivity, specificity, positive and negative likelihood ratio of AMH for predicting live birth was generated. This was conducted by fitting a two-level mixed logistic regression model, with independent binomial distributions for the true positives and true negatives restricted to the sensitivity and specificity in each study, and a bivariate normal model for the logit transforms of sensitivity and specificity between studies (Rutter and Gatsonis 2001; Reitsma et al., 2005; Harbord et al., 2007). In addition, the hierarchical model estimated the characteristics of the ROC curve and the adjusted DOR after including age and AMH assay as covariates (Takwoingi and Deeks, 2010).

Ethics

Formal ethics approval was not required because this analysis consists of pooling of published studies.

Results

Search results

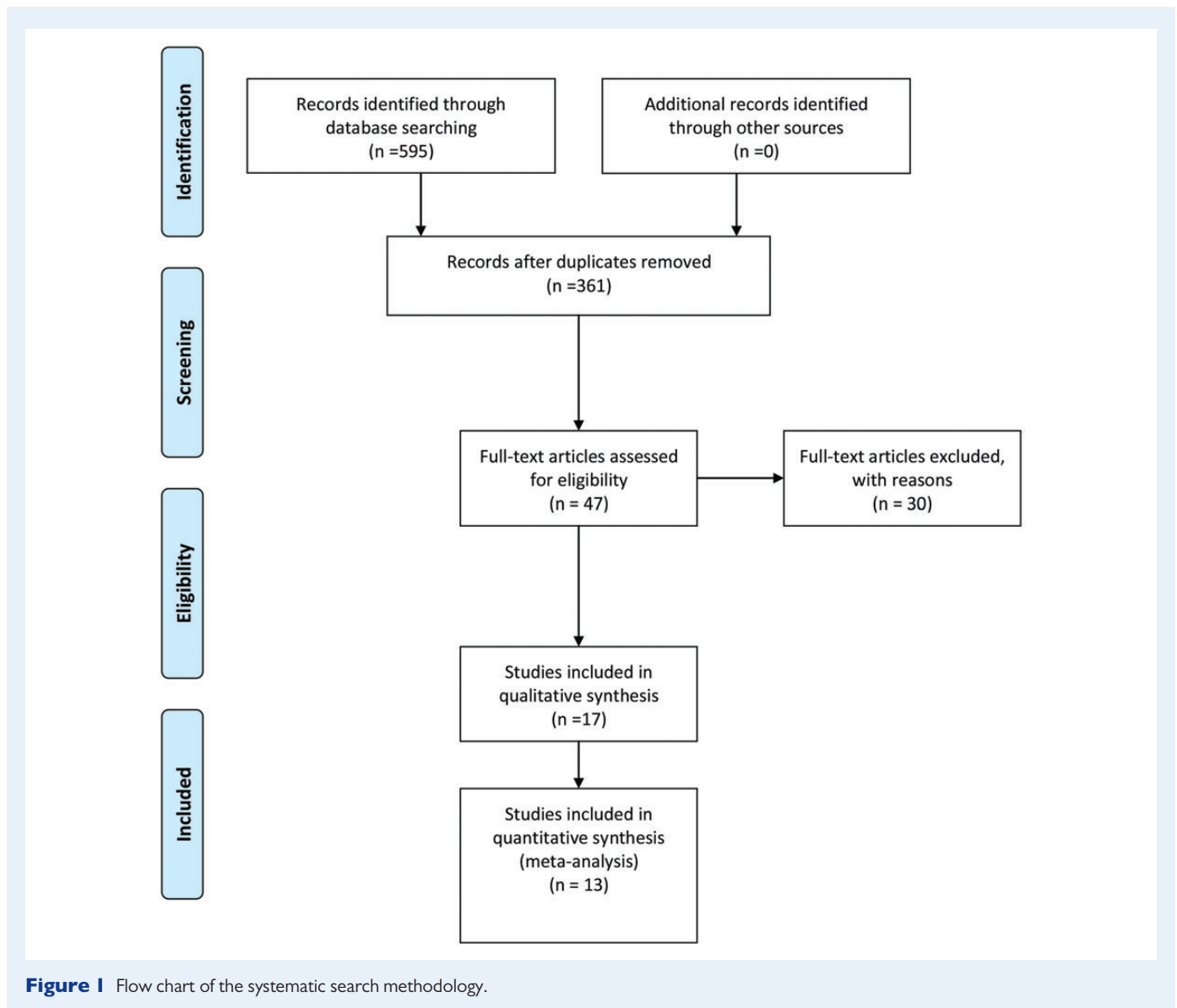
The systematic search of the biomedical databases produced 595 hits; after excluding duplicates 361 citations were identified (Fig. 1). Unpublished literature (open grey website or hand searching of references) meeting the search indices was not identified. After excluding articles based on the title or abstract, 47 articles were assessed fully for eligibility. Thirty studies were excluded with a reason recorded (Supplementary

data, Table S1); thus 17 studies were selected for the systematic review (Nelson et al., 2007; Lee et al., 2009; Gleicher et al., 2010; Majumder et al., 2010; Wang et al., 2010; Friden et al., 2011; La Marca et al., 2011; Weghofer et al., 2011; Grzegorzczuk-Martin et al., 2012; Arce et al., 2013; Brodin et al., 2013; Khader et al., 2013; Li et al., 2013; Lin et al., 2013; Lukaszuk et al., 2013; Merhi et al., 2013; Mutlu et al., 2013). Four of these were excluded from the meta-analysis as extraction of relevant data was not possible even after contacting the authors (Wang et al., 2010; Weghofer et al., 2011; Lin et al., 2013; Mutlu et al., 2013) and a subgroup of another study for the same reason (Lee et al., 2009). Most of the data from one of the excluded studies (Weghofer et al., 2011) were included in a previous study from the same research group (Gleicher et al., 2010) which contributed to the meta-analysis. One of the studies was in French (Grzegorzczuk-Martin et al., 2012). The characteristics of the studies included in the meta-analysis are listed in Table 1.

Pooled estimates

We present data of 6856 cycles (6306 women) undergoing IVF or ICSI. The studies were categorized into those including only women with expected low ovarian reserve ($n = 542$ women) and those with women with unknown ovarian reserve ($n = 5764$) to minimize the between study heterogeneity. The pooled DOR for AMH predicting a live birth among women with unknown ovarian reserve who present to a fertility clinic was 2.39 (95% CI: 1.85–3.08) (Fig. 2a). The estimated I-squared was 58.9%, suggesting moderate heterogeneity between the studies. After adjustment for age, the DOR obtained by a hierarchical logistic regression model analysis was very similar at 2.48 (95% CI: 1.81–3.22). After adjusting for the assay used to measure AMH, the DOR was almost identical at 2.42 (95% CI: 1.86–3.14). The DOR for women with expected low ovarian reserve was 4.63 (95% CI: 2.75–7.81) (I-squared 0%), with wider CIs due to the small number of pooled studies (Fig. 2b). Hierarchical logistic regression analysis was not conducted for this subgroup of studies because of the small sample size ($n = 3$ studies). Since the 95% CIs of the DOR of the sensitivity analysis overlapped, the pooled DOR for all studies was estimated at 2.63 (95% CI: 2.02–3.40) with estimated I-squared of 62.1%. After adjustment for age the DOR for all studies was 2.67 (95% CI: 2.06–3.48). After including AMH assay as a covariate, the adjusted DOR obtained by the hierarchical model was 2.66 (95% CI: 2.06–3.43).

For the prediction of live birth in women with unknown ovarian reserve ($n = 10$ studies), the hierarchical summary receiver operating characteristics (HSROCs) along with study-specific estimates are plotted in Fig. 3. The parameters of the plot did not change substantially after including age or AMH assay as covariates (data not shown). The summary ROC and 95% CIs do not cross the line of no-discrimination. The AUC was 0.61 (CI 0.56–0.65). The overall summary estimates of the above 10 studies for serum AMH and live birth were sensitivity of 83.7% (95% CI: 72.5–90.9%) and specificity of 32.0% (95% CI: 21.6–44.6%) with adjusted for age sensitivity of 83.7% (95% CI: 72.5–90.1%) and specificity of 32.6% (95% CI: 21.8–45.5%). However, caution is required interpreting the summary sensitivity and specificity as the pooled studies have similar but not identical AMH thresholds (Table 1), and summary sensitivity and specificity can vary according to the threshold used. The adjusted for age positive likelihood ratio was 1.24 (95% CI: 1.14–1.36) and negative likelihood ratio 0.5 (95% CI: 0.39–0.65).



Study quality assessment and publication bias

The quality assessment of the selected studies is represented as percentage of high, low or unclear bias in each domain assessed by the QUADAS 2 tool (Supplementary data, Fig. S1). Most studies reported live birth per cycle (or per patient if they included solely the first stimulated cycle (Nelson *et al.*, 2007; Lee *et al.*, 2009; Majumder *et al.*, 2010; La Marca *et al.*, 2011; Grzegorzczuk-Martin *et al.*, 2012; Khader *et al.*, 2013), one study reported live birth per ovum retrieval (Friden *et al.* 2011) and two reported the cumulative live birth rate (Arce *et al.*, 2013; Li *et al.*, 2013). The majority of the studies measured AMH using the Diagnostic Systems Laboratories Inc. (DSL, Webster, TX, USA) assay (Nelson *et al.*, 2007; Lee *et al.*, 2009; Gleicher *et al.*, 2010; Majumder *et al.*, 2010; Friden *et al.*, 2011; Brodin *et al.*, 2013; Khader *et al.*, 2013; Lukaszuk *et al.*, 2013; Merhi *et al.*, 2013) with the remainder using the Immunotech-Beckman Coulter (IBC, Marseille, France) assay (La Marca *et al.*, 2011; Grzegorzczuk-Martin *et al.*, 2012) or the Beckman Coulter Generation II assay (Arce *et al.*, 2013; Li *et al.*, 2013). With respect to potential

selection bias, two studies excluded women with polycystic ovary syndrome (La Marca *et al.*, 2011; Grzegorzczuk-Martin *et al.*, 2012), one excluded couples with severe male factor infertility (La Marca *et al.*, 2011), some studies included only women with low expected ovarian reserve defined either by advanced age and/or high FSH and/or low AMH (Gleicher *et al.*, 2010; Friden *et al.*, 2011; Weghofer *et al.*, 2011; Merhi *et al.*, 2013).

The funnel plot (Supplementary data, Fig. S2) visually suggests asymmetry raising the possibility that studies with small sample size and results lacking statistical significance may be missing; however, the statistical test for funnel plot asymmetry did not reach statistical significance ($P = 0.25$).

Discussion

This meta-analysis of 6306 women suggests that AMH has some association with predicting live birth in women undergoing IVF; however, its

Table 1 Characteristics of the studies included in the meta-analysis.

Study	Population	AMH test	Outcome
a. Characteristics of studies of women undergoing IVF/ICSI			
Li <i>et al.</i> (2013)	<i>n</i> = 1026 consecutive women undergoing their first cycle of IVF/ICSI from 2007 to 2009 and were stimulated with either the long GnRH agonist protocol or the GnRH antagonist protocol. Egg donation and preimplantation cycles were excluded. Study subjects had a median age of 35 years (IQ, 33–38) and had various aetiologies of infertility (male infertility was the majority, followed by tubal, endometriosis, anovulation and unexplained). Retrospective cohort study design.	Gen II Assay was used No threshold was used and sensitivity and specificity were extracted (Plot digitizer software) from graph	Live birth after the first fresh cycle, no frozen cycles were included in the analysis. <i>n</i> = 383 achieved a live birth. <i>n</i> = 643 did not have a live birth after a fresh IVF cycle
Arce <i>et al.</i> (2013)	<i>n</i> = 749 women aged 21–34 years with either male or unexplained infertility stimulated with the GnRH antagonist protocol. Excluded women with polycystic ovary syndrome (PCOS), endometriosis and previous poor response. Women had FSH I – 12 IU/l and antral follicle count \geq 10. It was a secondary analysis of data prospectively collected in a randomized assessor blinded trial.	Gen II Assay was used. Threshold \geq 13 pmol/l (1.82 ng/ml) was used	<i>n</i> = 291 live birth after the first fresh cycle
Lukaszuk <i>et al.</i> (2013)	<i>n</i> = 619 women (of 2495) had anti-Müllerian hormone (AMH) measured and underwent their first ICSI from 2005 to 2009. Median age was 32 (IQ, 29–34) with male infertility, anovulation, tubal factor, endometriosis and unexplained infertility. They underwent pituitary suppression with GnRH agonist protocol. Retrospective observational study.	DSL ELISA kit Threshold \geq 1.9 ng/ml (13.6 pmol/l) was used	<i>n</i> = 39 women with AMH below the threshold had a live birth. <i>n</i> = 252 women with AMH above the threshold had a live birth
Brodin <i>et al.</i> (2013)	<i>n</i> = 892 consecutive women with a maximum age of 42 years (median 36 years) underwent 1230 cycles of IVF/ICSI stimulated by the long GnRH agonist protocol. Aetiology of infertility included male factor, unexplained, tubal factor and endometriosis. Prospective data collection was conducted.	DSL ELISA kit Threshold \geq 0.84 ng/ml (6 pmol/l) was used	<i>n</i> = 33 women with AMH below the threshold had a live birth (per stimulated cycle). <i>n</i> = 222 women with AMH above the threshold had a live birth (per stimulated cycle)
Khader <i>et al.</i> (2013)	<i>n</i> = 822 women aged 25–42 years and had had their first IVF/ICSI cycle from 2006 to 2010. They underwent a GnRH agonist or antagonist protocol and the causes of infertility included male factor, anovulation, tubal disease, endometriosis and unexplained infertility. Data were collected prospectively in a registry and it was a retrospective cohort study.	DSL ELISA kit Threshold \geq 0.73 ng/ml (5.2 pmol/l) was used	<i>n</i> = 5 women with AMH below the threshold had a live birth. <i>n</i> = 237 women with AMH above the threshold had a live birth
Grzegorzczak-Martin <i>et al.</i> (2012)	<i>n</i> = 704 women attending for their first IVF/ICSI study from 2006 to 2009 and underwent a long GnRH agonist protocol. Women with PCOS, one ovary, older than 43 years of age were excluded. Also, women with FSH \geq 10 IU/l and AMH $>$ 2 ng/ml were excluded from the authors. We further excluded from the analysis women with FSH \geq 10 IU/l and AMH \leq 2 ng/ml (<i>n</i> = 54, so <i>n</i> = 650 included in the meta-analysis) because they were more likely to be poor responders. Retrospective cohort study.	IBC ELISA kit Threshold \geq 2 ng/ml was used	<i>n</i> = 19 women with AMH below the threshold had a live birth. <i>n</i> = 100 women with AMH above the threshold had a live birth

La Marca <i>et al.</i> (2011)	<i>n</i> = 381 women attended for their first IVF/ICSI cycle from 2005 to 2008, aged up to 42 years (mean \pm SD, 34.8 \pm 4.8 years) and were stimulated by a long GnRH protocol. Couples with severe male infertility (sperm count $< 1 \times 10^6$ /ml or normal forms $< 5\%$) or women with systematic diseases were excluded from the study. Data were collected prospectively in a registry and it was a retrospective observational study.	IBC ELISA kit Threshold ≥ 0.4 ng/ml was used	<i>n</i> = 3 women with AMH below the threshold had a live birth. <i>n</i> = 98 women with AMH above the threshold had a live birth
Majumder <i>et al.</i> (2010)	<i>n</i> = 162 women undergoing their first IVF/ICSI cycle from 2005 to 2006 and aged 23 to 39 years (Mean \pm SD, 31.8 \pm 3.79). They underwent a long GnRH agonist protocol for the common causes of infertility. It was a prospective observational study.	DSL ELISA kit Threshold ≥ 19.3 pmol/l (2.7 ng/ml) had 65.8% sensitivity and 54.8% specificity of predicting live birth (per embryo transfer)	<i>n</i> = 38 had a live birth (out of 137 women having had an embryo transfer)
Lee <i>et al.</i> (2009)	<i>n</i> = 336 underwent their first IVF/ICSI cycle from March 2007 to December 2007. They underwent a long GnRH agonist protocol for male factor, tubal factor, other female factor or unexplained infertility. Women with PCOS were excluded. <i>n</i> = 213 were under 35 years of age (mean \pm SD, 30.8 \pm 0.2) and <i>n</i> = 123 were ≥ 35 years of age (mean \pm SD, 38.6 \pm 0.2). It was a prospective cohort study.	DSL ELISA kit Threshold ≥ 1.68 ng/ml (12 pmol/l) had 60% sensitivity and 66.2% specificity of predicting live birth (per embryo transfer) in women ≥ 35 years	<i>n</i> = 40 women ≥ 35 years had a live birth (out of 114 women having had an embryo transfer). There were no extractable data for women < 35 years
Nelson <i>et al.</i> (2007)	<i>n</i> = 340 consecutive patients undergoing their first stimulated cycle with a GnRH agonist protocol. The patients had a median age of 34 years (IQ: 31–37 year). Prospective cohort study.	DSL ELISA kit Threshold ≥ 7.8 pmol/l (1.1 ng/ml)	<i>n</i> = 93 women achieved a live birth
b. Characteristics of studies of women with expected low ovarian reserve (LOR) undergoing IVF/ICSI			
Merhi <i>et al.</i> (2013)	<i>n</i> = 120 historic cohort with LOR underwent IVF/ICSI with a GnRH agonist protocol from January 2008 to June 2013. The participants were over 35 years of age and LOR defined as FSH ≥ 10 IU/l. Participants were subdivided in three groups based on AMH levels with mean age \pm SD group of 41.2 \pm 3.3, 39.3 \pm 3 and 40.2 \pm 2.8 years.	DSL ELISA kit Threshold ≥ 0.8 ng/ml (5.7 pmol/l)	<i>n</i> = 9 women achieved a live birth
Friden <i>et al.</i> (2011)	<i>n</i> = 127 women aged 39–46 years undergoing their first stimulated cycle from November 2006 to December 2008. They underwent standard antagonist or agonist protocols. There is no information regarding the study design.	DSL ELISA kit Threshold ≥ 8.6 pmol/l (1.2 ng/ml)	<i>n</i> = 6 women with AMH below the threshold had a live birth. <i>n</i> = 8 women with AMH above the threshold had a live birth
Gleicher <i>et al.</i> (2010)	<i>n</i> = 295 with LOR underwent 507 cycles and were stimulated with dehydroepiandrosterone (DHEA) and microdose agonist. LOR was defined arbitrarily as FSH ≥ 10 IU/l or/and abnormally low age-specific AMH. <i>n</i> = 174 women had AMH ≤ 1.05 ng/ml and had a mean age 39.6 years (SD: 4.6). <i>n</i> = 121 women had AMH > 1.05 ng/ml and mean age 35.2 years (SD: 5.4).	DSL ELISA kit Threshold > 1.05 ng/ml (7.5 pmol/l) had a sensitivity of 73.6% and specificity of 67.4% in predicting live birth per stimulating cycle.	<i>n</i> = 43 women had a live birth (507 cycles) (data extracted with the Plot digitizer software)

PCOS, polycystic ovary syndrome; DSL, Diagnostic Systems Laboratories; IBC, Immunotech-Beckman Coulter.

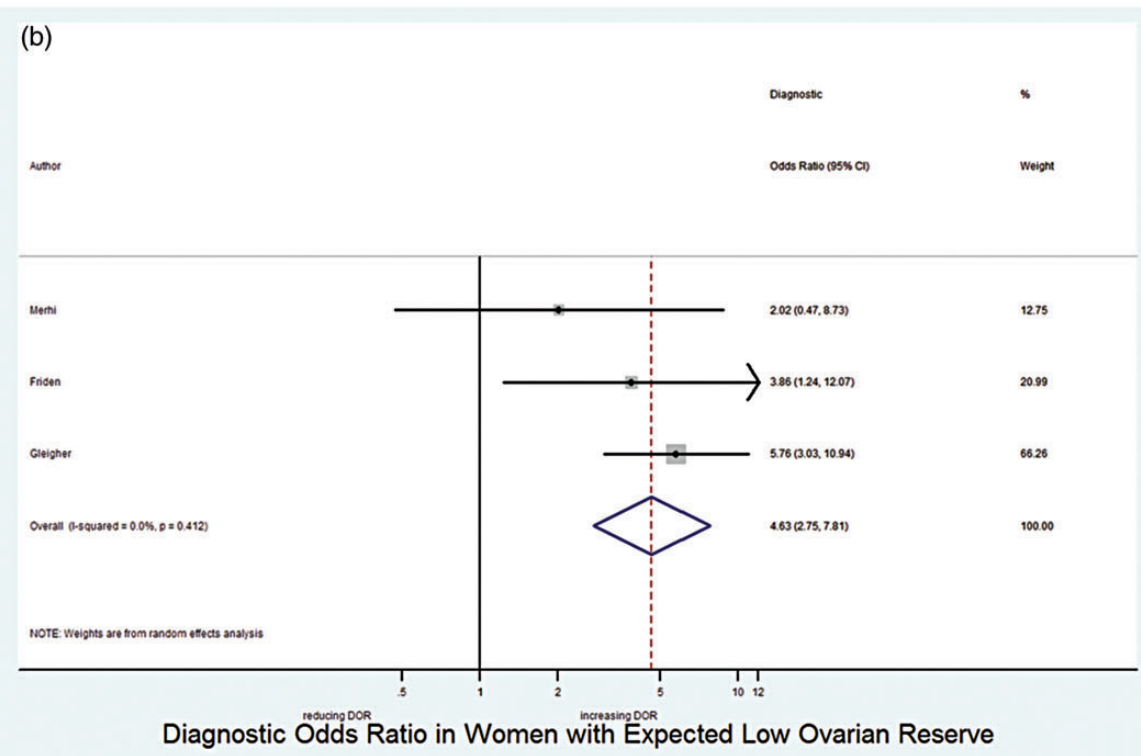
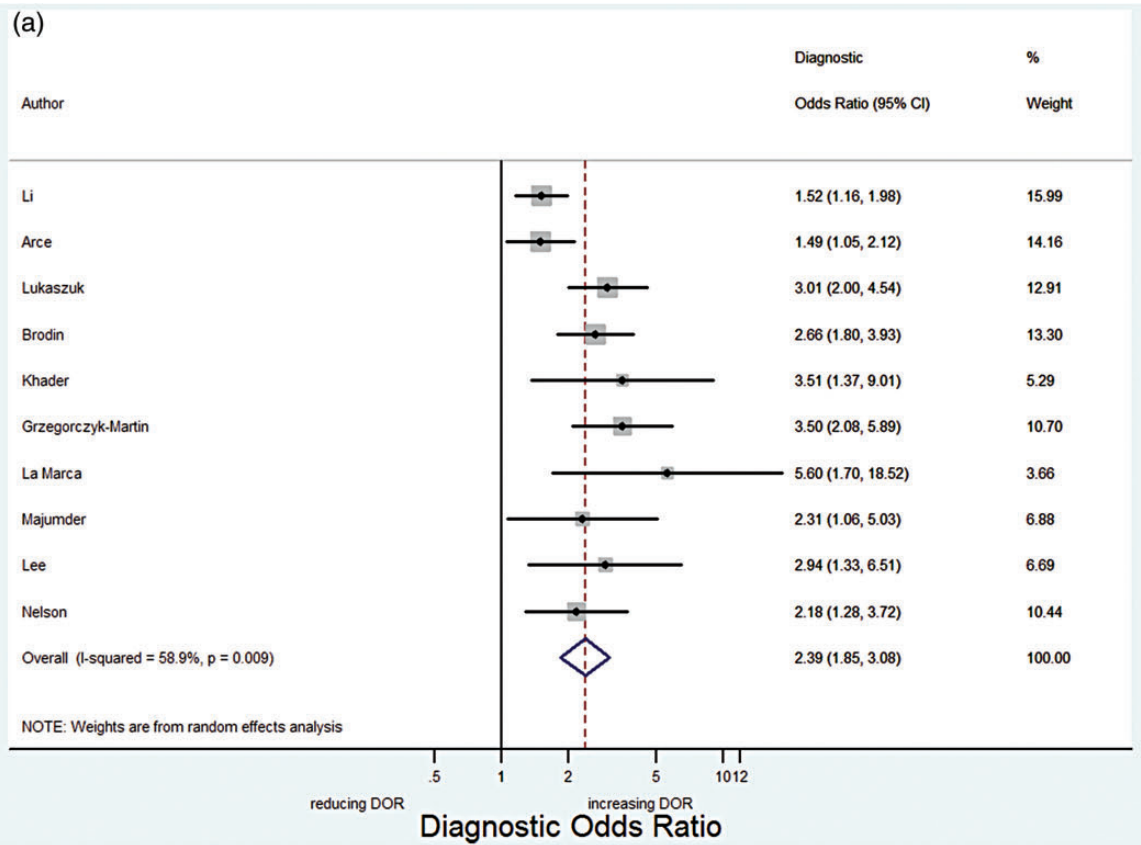


Figure 2 (a) Forest plot of diagnostic odds ratio (DOR) of 10 studies including women with unknown ovarian reserve before being included in the pooled studies. (b) Forest plot of DOR of three studies including women with expected low ovarian reserve before being included in the pooled studies.

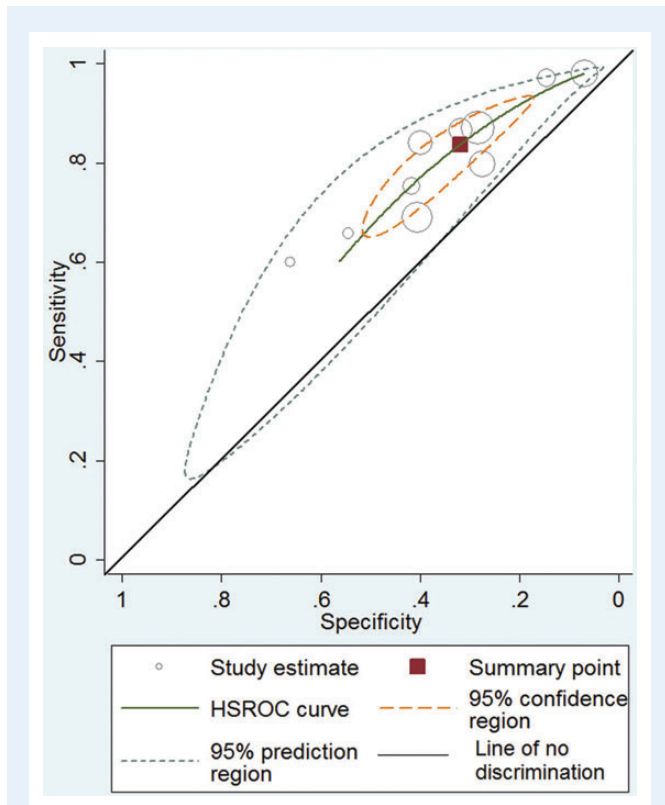


Figure 3 Hierarchical summary receiver operating characteristic curve (HSROC) of AMH in the prediction of live birth after IVF/ICSI with 95% confidence region, 95% prediction region and diagonal line of no discrimination. The area under the curve (AUC) is 0.61 (CI 0.56–0.65).

predictive accuracy is poor. The pooled DOR among 5764 women with unknown ovarian reserve was 2.48 after adjusting for age. In the remaining women with expected low ovarian reserve ($n = 542$), AMH had a better, albeit still small, predictive effect (DOR = 4.63); however, this needs to be substantiated in larger studies. Although the DOR was greater than unity in all the pooled studies, it was consistently low; it is known that useful tests with good predictive accuracy tend to have DOR above 20 (Fischer *et al.*, 2003). The HSROC model and 95% CIs of the pooled data did not cross the no-discrimination line, indicating that AMH has some value in predicting live birth. In addition, the prediction 95% confidence region, which suggests the confidence region for a forecast of the true specificity and sensitivity in a future study, includes the line of no-discrimination, if only marginally; this raises the possibility of a positive predictive value of AMH being found in future studies. However, it is established that tests with likelihood ratios ranging from 0.33 to 3 rarely change clinical decisions (Jaeschke *et al.*, 1994); therefore, the small positive (1.24) and the large negative likelihood ratio (0.5) found in our study indicate that AMH alone is unlikely to alter a clinical decision based on the chance of live birth after IVF/ICSI.

The potential value of ovarian reserve tests, including AMH, in predicting the likelihood of pregnancy after assisted conception has been contentious. Initial meta-analysis and more recent individual patient data meta-analysis did not demonstrate an association (Broekmans *et al.*, 2006; Broer *et al.*, 2013). This was despite ovarian reserve tests, and in

particular AMH, being strongly associated with ovarian response and oocyte yield (Broer *et al.*, 2009, 2011, 2013), which is a known major determinant of live birth (Sunkara *et al.*, 2011). The current meta-analysis has used a solid methodology by pooling a large number of studies with the outcome of live birth rather than pregnancy, so the conclusions should not be viewed as contradictory but supplementary.

AMH may not be associated with a positive pregnancy test or ongoing pregnancy, with the small association with live birth only becoming apparent after non-continuing pregnancies are lost. Although we did not test the value of AMH in predicting non-continuing pregnancies or ongoing pregnancies in the current review, we could consider this as a possible biological explanation. This would suggest that AMH is not only a marker of ovarian response and oocyte quantity (La Marca *et al.*, 2010; Broer *et al.*, 2011, 2013; Arce *et al.*, 2013), but also that it has a (limited) association with oocyte quality. The decline in oocyte quality with increasing age is well established (Nybo Andersen *et al.*, 2000), but the literature on ovarian reserve tests and oocyte quality is inconsistent. The positive association of AMH with increasing cumulative live birth rate is attributed to the greater availability of oocytes and embryos and not better oocyte quality by some (Arce *et al.*, 2013) but was found to be independent of age and oocyte yield by others (Brodin *et al.*, 2013). Several studies have not observed an association between AMH and oocyte or embryo quality (Smeenk *et al.*, 2007; Lie Fong *et al.*, 2008; Guerif *et al.*, 2009; Mashiach *et al.*, 2010; Riggs *et al.*, 2011; Anckaert *et al.*, 2012; Kedem-Dickman *et al.*, 2012; Arce *et al.*, 2013), whereas others report a positive association (Ebner *et al.*, 2006; Majumder *et al.*, 2010; Irez *et al.*, 2011; Brodin *et al.*, 2013; Lin *et al.*, 2013). As assessment of oocyte and embryo quality has largely focused on morphology rather than objective measures of euploid status and developmental potential (Plante *et al.*, 2010; Kline *et al.*, 2011), the alternative outcome of a live birth has been used by some researchers to assess oocyte quality. Many of these studies were small (Ebner *et al.*, 2006; Smeenk *et al.*, 2007; Lie Fong *et al.*, 2008; Majumder *et al.*, 2010; Mashiach *et al.*, 2010; Irez *et al.*, 2011; Riggs *et al.*, 2011; Kedem-Dickman *et al.*, 2012; Lin *et al.*, 2013), raising the possibility of a beta (false negative) error (Riggs *et al.*, 2011). The present analysis indicates that AMH has some value, albeit with poor accuracy, in predicting live birth, and that this relationship is independent of age or AMH assay. While most circulating AMH derives from small antral follicles (Weenen *et al.*, 2004; Jeppesen *et al.*, 2013), AMH expression persists in the cumulus cells surrounding the oocyte at the time of ovulation (Salmon *et al.*, 2004) providing a potential basis for a relationship with oocyte quality distinct from its relationship with follicle number.

The included studies in our analysis reported AMH according to the IBC, the DSL or the Beckman Coulter Generation II assay. Although the IBC and DSL assays differ both in their pairs of monoclonal antibodies and standardization so do not give comparable values, the conversion formula of the DSL assay data into IBC values of $2.02 * DSL = IBC$ has been used consistently for data aggregation studies (Hehenkamp *et al.*, 2006). In contrast, the Generation II assay was recently released after harmonization of the other two assays by incorporating the antibodies used in the DSL assay but being calibrated to the Immunotech assay and thereby anticipated to give equivalent values (Nelson and La Marca 2011). However, since the commercial release of the AMH Generation II assay it has been demonstrated that there is a systematic shift in assay calibration and AMH values generated with the AMH Generation II assay were significantly lower compared with the DSL assay for women

of similar age (Nelson et al., 2013a). Therefore, the different AMH assays with the current calibration concerns could be a source of significant heterogeneity in our pooled analysis; however, after adjustment for AMH assay the DOR for both women of unknown ovarian reserve and all women was only changed modestly confirming further that, irrespective of the assay, measured AMH has some value in the prediction of live birth.

Clinical application

The immediate clinical implication of the present finding is that AMH independently of age provides additional information for couples considering assisted reproduction. However, its diagnostic accuracy in live birth is poor and should not be used to alter clinical decisions and exclude couples from IVF/ICSI based on a low AMH. In addition, these data do not justify adoption of an AMH threshold for access to such treatments and further studies are needed to investigate whether a universal AMH threshold is possible, or appropriate.

To date a wide range of prediction models has been developed to facilitate prognostication of the likelihood of success after assisted conception, with none of these going through the three classical phases of model development (Nelson and Lawlor 2011; van Loendersloot et al., 2013). A recent systematic review and meta-analysis which analysed nine predictive factors in IVF identified that female age and baseline FSH were inversely associated with the likelihood of success after IVF (van Loendersloot et al., 2010). As AMH is a stronger associate of ovarian reserve than FSH (Broekmans et al., 2006; Hansen et al., 2011), this suggests that future prediction models should consider AMH as an alternative covariate. Critical assessment of whether inclusion of AMH improves the prediction characteristics of existing models, when compared with updating existing prediction models with adjustment or recalibration to account for local circumstances, will be a critical step in confirming its clinical utility in prediction of live birth.

Strengths and limitations

This is the first study presenting pooled data of a large number of cycles to assess the predictive value of serum AMH in live birth after IVF/ICSI. The strengths of the review lie in the extensive search strategy, adherence to recent guidelines (Cochrane 2011), inclusion of non-English studies and robust statistical analysis in accordance with established guidance for diagnostic tests (Irwig et al., 1994; Khan et al., 2001). Although the process of systematic literature review and meta-analysis is a robust way of generating a more powerful estimate of true-effect size with less random error than individual studies, it does have limitations and the inferences assumed by the data are subject to the limitations and bias of the primary studies. Heterogeneity of the studies must be addressed as it may affect the justification for pooling the data into one analysis. In the case of the present meta-analysis, heterogeneity may have been caused by different baseline characteristics in study participants, different stimulating protocols, variation in the AMH threshold and assay across different studies and study quality characteristics. However, the statistical estimation of heterogeneity was within acceptable levels for pooling studies. In addition, one of the advantages of the HSROC analysis is that it takes into account the full range of variation in the data, differentiating within study from between study variability and systematic from random variability (Gatsonis and Paliwal 2006). Secondly, our pooled analysis did not include data from some studies (Wang et al., 2010; Lin et al., 2013; Mutlu et al., 2013) and from a subgroup from

another study (Lee et al., 2009) (overall $n = 1956$ women) as the data to conduct a 2×2 table were not extractable or available even after contacting the authors, which we acknowledge may have introduced bias. The largest of these studies showed a positive relationship between AMH and live birth although it attenuated after age stratification (Wang et al., 2010), while the others did not show an association possibly due to small sample number ($n = 83$ (Lin et al., 2013); $n = 213$ (Lee et al., 2009); $n = 192$ (Mutlu et al., 2013)). Also, while the funnel plot analysis raises the possibility that small studies showing non-significant or negative association between AMH and live birth may be missing, an asymmetrical funnel plot does not prove a specific type of bias and may indicate small study effect, i.e. smaller studies have the tendency to inflate the summary effect (Sterne et al., 2011) and moreover, the apparent funnel asymmetry was not statistically significant. Our methodology tried to minimize the possibility of publication bias by implementing a robust search technique without language limitations and by contacting the authors when the relevant information was not extractable. In addition, the qualitative assessment of the included studies indicates that the majority of the studies had low risk of methodological bias (Supplementary data, Fig. S1). Our pooled analysis does not derive a summary diagnostic threshold for AMH, as this would be inappropriate due to the individual studies utilizing diverse populations with different fertility potentials and ethnic backgrounds, different AMH cut-off points and different AMH assays.

Conclusion

Based on the current evidence, we found that AMH adds some value in predicting live birth, and this is independent of age or AMH assay. Although the CIs of the DOR for AMH do not cross unity, its predictive accuracy is poor and should not be over-interpreted. These findings were consistent across all of the studies examined, but the existing evidence would be inappropriate to determine a widely applicable threshold value due to the heterogeneity between studies, and likewise to derive the relative risks of live birth across the clinically relevant range of AMH values or exclude women from IVF/ICSI. This evidence can only be obtained by prospectively designed studies of test accuracy with adequate clinical size and attention to limiting bias and appropriate outcome measures (possibly including cost effectiveness analysis) using a decision tree model. This will allow the trade-off between positive and negative benefits to be truly evaluated.

Supplementary data

Supplementary data are available at <http://humup.oxfordjournals.org/>.

Authors' roles

S.M.N. and R.A.A. conceived the idea of the study. S.I. and S.M.N. did the systematic search, assessed the eligible studies and extracted the data. S.I. pooled the data and conducted the statistical analysis. T.W.K. contributed to the statistical analysis. O.W. reviewed the methodology. S.I., S.M.N. and R.A.A. drafted the paper. All the authors contributed to the final version of the paper.

Funding

No funding was received for this study.

Conflict of interest

S.M.N. and R.A.A. have undertaken consultancy work for Beckman Coulter. The other authors have no conflicts to declare.

References

- Anckaert E, Smits J, Schiettecatte J, Klein BM, Arce JC. The value of anti-Müllerian hormone measurement in the long GnRH agonist protocol: association with ovarian response and gonadotrophin-dose adjustments. *Hum Reprod* 2012; **27**:1829–1839.
- Arce JC, La Marca A, Mirner Klein B, Nyboe Andersen A, Fleming R. Antimüllerian hormone in gonadotropin releasing-hormone antagonist cycles: prediction of ovarian response and cumulative treatment outcome in good-prognosis patients. *Fertil Steril* 2013; **99**:1644–1653.
- Brodin T, Hadziostmanovic N, Berglund L, Olovsson M, Holte J. Antimüllerian hormone levels are strongly associated with live-birth rates after assisted reproduction. *J Clin Endocrinol Metab* 2013; **98**:1107–1114.
- 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; **12**:685–718.
- Broer SL, Mol BWJ, Hendriks D, Broekmans FJM. The role of antimüllerian hormone in prediction of outcome after IVF: comparison with the antral follicle count. *Fertil Steril* 2009; **91**:705–714.
- Broer SL, Dolleman M, Opmeer BC, Fauser BC, Mol BW, Broekmans FJM. AMH and AFC as predictors of excessive response in controlled ovarian hyperstimulation: a meta-analysis. *Hum Reprod Update* 2011; **17**:46–54.
- Broer SL, van Disseldorp J, Broeze KA, Dolleman M, Opmeer BC, Bossuyt P, Eijkemans MJC, Mol BWJ, Broekmans FJM. Import Study Group. 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; **19**:26–36.
- Cochrane Handbook for DTA reviews. 2011. <http://srdta.cochrane.org/handbook-dta-reviews>.
- Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005; **58**:882–893.
- Ebner T, Sommergruber M, Moser M, Shebl O, Schreier-Lechner E, Tews G. Basal level of anti-Müllerian hormone is associated with oocyte quality in stimulated cycles. *Hum Reprod* 2006; **21**:2022–2026.
- Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med* 2003; **29**:1043–1051.
- Friden B, Sjoblom P, Menezes J. Using anti-Müllerian hormone to identify a good prognosis group in women of advanced reproductive age. *Aust N Z J Obstet Gynaecol* 2011; **51**:411–415.
- Gatsonis C, Paliwal P. Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer. *AJR Am J Roentgenol* 2006; **187**:271–281.
- Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; **56**:1129–1135.
- Gleicher N, Weghofer A, Barad DH. Anti-Müllerian hormone (AMH) defines, independent of age, low versus good live-birth chances in women with severely diminished ovarian reserve. *Fertil Steril* 2010; **94**:2824–2827.
- Grzegorzczak-Martin V, Khrouf M, Bringer-Deutsch S, Mayenga JM, Kulski O, Cohen-Bacrie P, Benaim JL, Belaïsch-Allart J. Pronostic en fécondation in vitro des patientes ayant une AMH basse et une FSH normale. *Gynécologie Obstétrique Fertil* 2012; **40**:411–418.
- Gueriff F, Lemseffer M, Couet ML, Gervereau O, Ract V, Royere D. Serum antimüllerian hormone is not predictive of oocyte quality in vitro fertilization. *Ann Endocrinol* 2009; **70**:230–234.
- Hansen KR, Hodnett GM, Knowlton N, Craig LB. Correlation of ovarian reserve tests with histologically determined primordial follicle number. *Fertil Steril* 2011; **95**:170–175.
- Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics (Oxford, England)* 2007; **8**:239–251.
- Hehenkamp WJ, Looman CW, Themmen AP, de Jong FH, Te Velde ER, Broekmans FJ. Anti-Müllerian hormone levels in the spontaneous menstrual cycle do not show substantial fluctuation. *J Clin Endocrinol Metab* 2006; **91**:4057–4063.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**:557–560.
- Honnma H, Baba T, Sasaki M, Hashiba Y, Oguri H, Fukunaga T, Endo T, Asada Y. Serum anti-Müllerian hormone levels affect the rate of ongoing pregnancy after in vitro fertilization. *Reprod Sci* 2013; **20**:51–59.
- Irez T, Ocal P, Guralp O, Cetin M, Aydogan B, Sahmay S. Different serum anti-Müllerian hormone concentrations are associated with oocyte quality, embryo development parameters and IVF-ICSI outcomes. *Arch Gynecol Obstet* 2011; **284**:1295–1301.
- Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994; **120**:667–676.
- Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; **271**:703–707.
- Jeppesen JV, Anderson RA, Kelsey TW, Christiansen SL, Kristensen SG, Jayaprakasan K, Raine-Fenning N, Campbell BK, Yding Andersen C. Which follicles make the most anti-Müllerian hormone in humans? Evidence for an abrupt decline in AMH production at the time of follicle selection. *Mol Hum Reprod* 2013; **19**:519–527.
- Kedem-Dickman A, Maman E, Yung Y, Yerushalmi GM, Hemi R, Hanochi M, Dor J, Hourvitz A. Anti-Müllerian hormone is highly expressed and secreted from cumulus granulosa cells of stimulated preovulatory immature and atretic oocytes. *Reprod Biomed Online* 2012; **24**:540–546.
- Khader A, Lloyd SM, McConnachie A, Fleming R, Grisendi V, La Marca A, Nelson SM. External validation of anti-Müllerian hormone based prediction of live birth in assisted conception. *J Ovarian Res* 2013; **6**:3.
- Khan KS, Dinnes J, Kleijnen J. Systematic reviews to evaluate diagnostic tests. *Eur J Obstet Gynecol Reprod Biol* 2001; **95**:6–11.
- 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; **26**:1537–1550.
- La Marca A, Sighinolfi G, Radi D, Argento C, Baraldi E, Arsenio AC, Stabile G, Volpe A. Anti-Müllerian hormone (AMH) as a predictive marker in assisted reproductive technology (ART). *Hum Reprod Update* 2010; **16**:113–130.
- La Marca A, Nelson SM, Sighinolfi G, Manno M, Baraldi E, Roli L, Xella S, Marsella T, Tagliasacchi D, D'Amico R et al. Anti-Müllerian hormone-based prediction model for a live birth in assisted reproduction. *Reprod Biomed Online* 2011; **22**:341–349.
- Lee TH, Liu CH, Huang CC, Hsieh KC, Lin PM, Lee MS. Impact of female age and male infertility on ovarian reserve markers to predict outcome of assisted reproduction technology cycles. *Reprod Biol Endocrinol* 2009; **7**:100.
- Li HW, Lee VC, Lau EY, Yeung WS, Ho PC, Ng EH. Role of baseline antral follicle count and anti-müllerian hormone in prediction of cumulative live birth in the first in vitro fertilisation cycle: a retrospective cohort analysis. *PLoS One* 2013; **8**:e61095.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; **339**:b2700.
- Lie Fong S, Baart EB, Martini E, Schipper I, Visser JA, Themmen AP, de Jong FH, Fauser BJ, Laven JS. Anti-Müllerian hormone: a marker for oocyte quantity, oocyte quality and embryo quality? *Reprod Biomed Online* 2008; **16**:664–670.
- Lin WQ, Yao LN, Zhang DX, Zhang W, Yang XJ, Yu R. The predictive value of anti-müllerian hormone on embryo quality, blastocyst development, and pregnancy rate following in vitro fertilization-embryo transfer (IVF-ET). *J Assist Reprod Genet* 2013; **30**:649–655.
- Lukaszuk K, Kunicki M, Liss J, Lukaszuk M, Jakiel G. Use of ovarian reserve parameters for predicting live births in women undergoing in vitro fertilization. *Eur J Obstet Gynecol Reprod Biol* 2013; **168**:173–177.
- Majumder K, Gelbaya TA, Laing I, Nardo LG. The use of anti-Müllerian hormone and antral follicle count to predict the potential of oocytes and embryos. *Eur J Obstet Gynecol Reprod Biol* 2010; **150**:166–170.
- Mashiach R, Amit A, Hasson J, Amzalag S, Almog B, Ben-Yosef D, Lessing JB, Limor R, Azem F. Follicular fluid levels of anti-Müllerian hormone as a predictor of oocyte maturation, fertilization rate, and embryonic development in patients with polycystic ovary syndrome. *Fertil Steril* 2010; **93**:2299–2302.
- Merhi Z, Zapantis A, Berger DS, Jindal SK. Determining an anti-müllerian hormone cutoff level to predict clinical pregnancy following in vitro fertilization in women with severely diminished ovarian reserve. *J Assist Reprod Genet* 2013; **30**:1361–1365.
- Mutlu MF, Erdem M, Erdem A, Yildiz S, Mutlu I, Arisoy O, Oktem M. Antral follicle count determines poor ovarian response better than anti-müllerian hormone but age is the

- only predictor for live birth in *in vitro* fertilization cycles. *J Assist Reprod Genet* 2013; **30**:657–665.
- Nelson SM, La Marca A. The journey from the old to the new AMH assay: how to avoid getting lost in the values. *Reprod Biomed Online* 2011; **23**:411–420.
- Nelson SM, Lawlor DA. Predicting live birth, preterm delivery, and low birth weight in infants born from *in vitro* fertilisation: a prospective study of 144018 treatment cycles. *PLoS Med* 2011; **8**:e1000386.
- 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; **22**:2414–2421.
- Nelson SM, Yates RW, Lyall H, Jamieson M, Traynor I, Gaudoin M, Mitchell P, Ambrose P, Fleming R. Anti-Mullerian hormone-based approach to controlled ovarian stimulation for assisted conception. *Hum Reprod* 2009; **24**:867–875.
- Nelson SM, Iliodromiti S, Fleming R, Anderson R, McConnachie A, Messow CM. Reference range for the antimullerian hormone Generation II assay: a population study of 10,984 women, with comparison to the established Diagnostics Systems Laboratory nomogram. *Fertil Steril* 2013a. doi:pii: S0015-0282(13)03167-1.
- Nelson SM, Telfer EE, Anderson RA. The ageing ovary and uterus: new biological insights. *Hum Reprod Update* 2013b; **19**:67–83.
- Nybo Andersen AM, Wohlfahrt J, Christens P, Olsen J, Melbye M. Maternal age and fetal loss: population based register linkage study. *BMJ* 2000; **320**:1708–1712.
- 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; **18**:1–11.
- Plante BJ, Beamon C, Schmitt CL, Moldenhauer JS, Steiner AZ. Maternal antimullerian hormone levels do not predict fetal aneuploidy. *J Assist Reprod Genet* 2010; **27**:409–414.
- Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; **58**:982–990.
- Riggs R, Kimble T, Oehninger S, Bocca S, Zhao Y, Leader B, Stadtmayer L. Anti-Mullerian hormone serum levels predict response to controlled ovarian hyperstimulation but not embryo quality or pregnancy outcome in oocyte donation. *Fertil Steril* 2011; **95**:410–412.
- Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001; **20**:2865–2884.
- Salmon NA, Handyside AH, Joyce IM. Oocyte regulation of anti-Mullerian hormone expression in granulosa cells during ovarian follicle development in mice. *Dev Biol* 2004; **266**:201–208.
- 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 intracytoplasmic sperm injection. *Fertil Steril* 2007; **87**:223–226.
- Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, Carpenter J, Rucker G, Harbord RM, Schmid CH et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011; **343**:d4002.
- 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; **26**:1768–1774.
- Takwoingi Y, Deeks JJ. MetaDAS. A SAS macro for meta-analysis of diagnostic accuracy studies. User Guide Version 1.3, 2010. <http://srdta.cochrane.org/>
- 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; **16**:577–589.
- van Loendersloot L, Repping S, Bossuyt PM, van der Veen F, van Wely M. Prediction models in *in vitro* fertilization; where are we? *J Advanc Res* 2013. doi:10.1016/j.jare.2013.05.002.
- van Rooij IA, Broekmans FJM, Hunault CC, Scheffer GJ, Eijkemans MJC, de Jong FH, Themmen APN, te Velde ER. Use of ovarian reserve tests for the prediction of ongoing pregnancy in couples with unexplained or mild male infertility. *Reprod Biomed Online* 2006; **12**:182–190.
- Wallace WH, Kelsey TW. Human ovarian reserve from conception to the menopause. *PLoS One* 2010; **5**:e8772.
- Wang JG, Douglas NC, Nakhuda GS, Choi JM, Park SJ, Thornton MH, Guarnaccia MM, Sauer MV. The association between anti-Mullerian hormone and IVF pregnancy outcomes is influenced by age. *Reprod Biomed Online* 2010; **21**:757–761.
- Weenen C, Laven JS, Von Bergh AR, Cranfield M, Groome NP, Visser JA, Kramer P, Fauser BC, Themmen AP. Anti-Mullerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. *Mol Hum Reprod* 2004; **10**:77–83.
- Weghofer A, Dietrich W, Barad DH, Gleicher N. Live birth chances in women with extremely low-serum anti-Mullerian hormone levels. *Hum Reprod* 2011; **26**:1905–1909.
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**:529–536.