Endometrial thickness and pregnancy rates after IVF: a systematic review and meta-analysis

Annemieke Kasius†, Janine G. Smit†*, Helen L. Torrance1, Marinus J.C. Eijkemans2, Ben Willem Mol3, Brent C. Opmeer4, and Frank J.M. Broekmans1

1Department of Reproductive Medicine and Gynecology, University Medical Center of Utrecht, Room Number: F05.126, Postbox 85500, 3508 GA Utrecht, The Netherlands 2Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, 3584 CG Utrecht, The Netherlands 3The Robinson Institute, School of Paediatrics and Reproductive Health, University of Adelaide, 5000 SA, Australia 4Clinical Research Unit, Academic Medical Center, Amsterdam, The Netherlands

*Correspondence address. E-mail: jsmit6@umcutrecht.nl
Submitted on November 15, 2013; resubmitted on February 12, 2014; accepted on February 24, 2014

TABLE OF CONTENTS

• Introduction
• Methods
  Search and selection strategy
  Eligibility criteria
  Quality assessment
  Statistical analysis
• Results
  Systematic search, selection and data extraction
  Study characteristics
  ROC analysis
  Odds ratios
  Meta-regression
• Discussion
  Limitations
  Strengths
  Implications for clinical practice
• Summary

BACKGROUND: Thin endometrium on ultrasound in the course of ovarian hyperstimulation has been thought to be associated with poor success rates after IVF, even in the absence of prior intrauterine surgery or infection. To assess the clinical significance of endometrial thickness (EMT) for IVF outcome, we performed a systematic review and meta-analysis.

METHODS: The electronic databases Pubmed, Cochrane and Embase were searched up to October 2013 for articles that studied the association between EMT and IVF outcome. The articles had to be written in the English or Dutch language. Studies were included if two-by-two tables for EMT and pregnancy rates could be constructed. Study quality was scored using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist. Summary receiver operating characteristic (sROC) curves were estimated to assess the accuracy of EMT in the
The relevance of endometrial thickness in IVF

**Introduction**

Three decades after the introduction of IVF, ongoing pregnancy rates per cycle vary between 8.6 and 46.2% (Centre for Medically Assisted Procreation, 2009; American Society for Reproductive Medicine, 2011; Human Fertilisation & Embryology Authority, 2011; Nederlandse Vereniging voor Obstetricie & Gynaecologie, 2011). Considering these relatively limited success rates versus the associated risks (Källén, 2011), the estimated sROC curve indicates a virtually absent discriminatory capacity of EMT in the prediction of pregnancy. At present, results from studies that investigated the relationship between EMT and the occurrence of pregnancy after IVF treatment are conflicting. Some studies found that thin EMT negatively affects pregnancy rates after fertility treatment (Kovacs et al., 2003; El-Toukhfy et al., 2008), while other studies could not confirm this (Oliveira et al., 1997; Lesny et al., 1999; Yuval et al., 1999; Merce et al., 2008). This lack of consensus can possibly be explained by the fact that no exact definition of thin endometrium as assessed by ultrasound exists (Senturk and Erel, 2008). Therefore, there is no solid conclusion on the clinical significance of EMT for the chance of conceiving after IVF.

Despite all this, EMT assessment has become part of standard monitoring during fertility treatment. Clinicians and infertile couples may thus be faced with the dilemma of whether they should proceed with the fresh IVF cycle or should choose to cryopreserve the embryos for potentially better chances to conceive with an endometrium developed under natural conditions in subsequent cryocycles.

**Methods**

**Search and selection strategy**

The electronic databases PubMed, Cochrane and Embase were searched until October 2013 for articles which described EMT and outcome after IVF treatment. The following keywords and their synonyms were used: (‘endometrial thickness’ or ‘thin endometrium’ or ‘endometrial characteristics’ or ‘endometrial stripe’ or ‘endometrial receptivity’ or ‘embryo implantation’) and (‘IVF’ or ‘IVF treatment’ or ‘ICSI’ or ‘infertility treatment’ or ‘assisted reproductive technology’) and (‘pregnancy’ or ‘IVF outcomes’ or ‘live-birth rate’ or ‘clinical outcomes’). No restrictions on date or type of publication were applied to the search (Supplementary data).

Titles and abstracts of all identified studies were screened and the full paper of the preselected articles was read by two researchers (A.K. and J.G.S). Both researchers extracted the data from the article independently. Any disagreement between the two researchers was resolved through discussion, or in case of persistent disagreement, by consultation with a third expert (F.B.).

**Eligibility criteria**

All cohort studies and RCTs investigating the effect of EMT on IVF outcome were considered eligible for inclusion. The articles had to be written in the English or Dutch language.

The patient population consisted of subfertile women undergoing IVF or ICSI treatment. Women of all ages were included. Studies that included cycles with donor oocytes or women with intrauterine pathology [e.g. uterine polyps, submucosal or intramural myoma’s and adhesions (Asherman syndrome)] were excluded from analysis.
EMT had to be measured by TVU on the day of ovulation triggering as the maximal echogenic distance between the junction of the endometrium and myometrium in the mid-sagittal plane.

The main study outcomes were live birth (defined as a live born baby $\geq 24$ weeks of gestation), ongoing pregnancy (defined as the presence of a living intrauterine fetus on TVU at the 20th week of gestation) and clinical pregnancy (defined as the presence of a gestational sac on TVU or other definitive clinical signs). Other outcomes were biochemical pregnancy (defined as the presence of positive hCG test combined with the absence of pregnancy findings on TVU), miscarriage (defined as a fetal loss prior to the 20th week of gestation) and ectopic pregnancy (defined as a pregnancy that implants outside of the uterus). All outcomes were reported per cycle.

Quality assessment
Each selected study was scored for their relevance and methodological quality by using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklists for Observational Studies, version October/November 2007. Furthermore, the following characteristics of the studies were taken into consideration: patient sampling (consecutive or other), data collection method (prospective or retrospective), study design (cohort or RCT), blinding (i.e. blinding of the sonographers for the medical history and the stimulation protocol used), selection bias (i.e. exclusion of cases because of reasons which affect the possibility of generalizing the results, for example women with previous pregnancies), verification bias (i.e. adapting the treatment protocol for optimizing the endometrium to improve the chance to conceive), information bias (i.e. measuring EMT by multiple sonographers), attrition bias (i.e. completeness of the outcome measures) and the stimulation protocol used (GnRH antagonist of GnRH agonist, urinary FSH or recombinant FSH).

Statistical analysis

Bivariate model and summary receiver operating characteristic
To evaluate the overall accuracy, including the whole range of possible thresholds, we estimated the summary receiver operating characteristic (sROC) curve using the bivariate model (Reitsma et al., 2005). The rationale behind the bivariate model is that reported estimates for sensitivity and specificity from different studies may be based on different positivity thresholds (explicitly due to cut-off values used, or implicitly related to the assessment method or device used). Consequently, part of the observed variability in the reported values will reflect a shift along an underlying ROC curve, reflected in a negative correlation between sensitivity and specificity. The bivariate model accommodates this correlation by including a covariance term in the model. Subsequently, sROC curves can be estimated from the model parameters (Harbord et al., 2007) and plotted with the original data points in the ROC space. Estimates were based on averaged model results from repeated stratified bootstrap samples to account for multiple accuracy points reported by the same study.

We also estimated the predictive accuracy for specific cut-off values of EMT (7, 8 and 9 mm), resulting in pooled estimates for sensitivity and specificity for each cut-off, as well as sROC curves reflecting implicit differences based on different positivity thresholds.

Odds ratios
ORs with 95% confidence intervals (CI) were calculated and displayed in a forest plot to determine whether the frequently reported cut-off of 7 mm is associated with reduced pregnancy rates. Two-by-two tables were constructed for the different outcome measures at a cut-off of 7 mm (live birth, ongoing pregnancy, clinical pregnancy). ORs of original studies were pooled with RevMan software (Version 5.1, The Cochrane Collaboration, 2008), using a random effect model. This model was chosen for explicating the heterogeneity between the included studies and to estimate the between-study variance. The amount of heterogeneity was quantified by using the $I^2$ statistic, which represents the percentage of total variability across the studies that is due to heterogeneity instead of chance. Low heterogeneity is defined as a value $<50$% (Higgins et al., 2003). Finally, whenever possible, we analysed whether the results and significance of the meta-analysis would change when the analysis was limited to prospective or retrospective studies only or when different cut-offs were used.

Meta-regression
Meta-regression analysis was performed with EMT $\leq 7$ mm as a dependent variable and female age and number of oocytes as independent variables to determine if differences in female age and number of oocytes could be of influence on the estimated effect of thin EMT. The meta-regression analysis was performed using the ‘metafor’ package, R Version 3.0.1. If the $P$-value was $<0.05$, results were considered to indicate statistical significance.

Results

Systematic search, selection and data extraction
The electronic search resulted in 1170 hits. Following the screening of titles, removing of duplicates, and exclusion of articles without full text, 89 studies were identified to be potentially eligible for inclusion (Fig. 1).

After reading the manuscripts and assessing the inclusion criteria and methodological quality, 23 studies were eligible for final inclusion. From 17 of these studies the appropriate data for calculating the absolute pregnancy rates could be extracted. Nine authors were contacted and asked for further information, of which five authors provided the appropriate data necessary to construct 2 $\times$ 2 tables (De Geyter et al., 2000; Grant et al., 2007; Singh et al., 2011) but could still be included because it was possible to construct 2 $\times$ 2 tables for other cut-off values. Thus, 22 studies were finally included in this systematic review and meta-analysis.

From 11 included studies, 2 $\times$ 2 tables for the endometrial cut-off value of $\leq 7$ versus $>7$ mm could be constructed (Basir et al., 2002; Amir et al., 2007; Richter et al., 2007; Al-Ghamdi et al., 2008; Dechaud et al., 2008; Bozdag et al., 2009; Okohue et al., 2009; Chen et al., 2010; Kinay et al., 2010; Kuc et al., 2011; Zhao et al., 2012). Eleven studies reported several other cut-off values (Rinaldi et al., 1996; Weissman et al., 1999; De Geyter et al., 2000; Dietterich et al., 2002; Aboulghar et al., 2005; Rashidi et al., 2005; Zhang et al., 2005; Grant et al., 2007; Traub et al., 2009; Singh et al., 2011; Yoeli et al., 2004). All authors of the articles included in the analysis for EMT $\leq 7$ versus $>7$ mm were requested to provide additional data on female age and number of oocytes retrieved of which seven replied with the appropriate data.

Study characteristics
Characteristics of included studies are listed in Table I. Most studies were of a retrospective cohort design and described only the clinical pregnancy rate as outcome measure.

Figure 2 shows the overall risk of bias in this meta-analysis. Only three studies described that the researchers were blinded for EMT results or hormonal response (Rinaldi et al., 1996; Yoeli et al., 2004; Dechaud et al., 2007; Richter et al., 2007; Al-Ghamdi et al., 2008; Dechaud et al., 2008; Bozdag et al., 2009; Okohue et al., 2009; Chen et al., 2010; Kinay et al., 2010; Kuc et al., 2011; Zhao et al., 2012). Eleven studies reported several other cut-off values (Rinaldi et al., 1996; Weissman et al., 1999; De Geyter et al., 2000; Dietterich et al., 2002; Aboulghar et al., 2005; Rashidi et al., 2005; Zhang et al., 2005; Grant et al., 2007; Traub et al., 2009; Singh et al., 2011; Yoeli et al., 2004). All authors of the articles included in the analysis for EMT $\leq 7$ versus $>7$ mm were requested to provide additional data on female age and number of oocytes retrieved of which seven replied with the appropriate data.
Selection bias appeared to be present in most of the studies due to the exclusion of specific patient groups such as elderly women, women with previous pregnancies or women undergoing frozen embryo treatment cycles. Also the overall risk of verification bias could not be determined because only five studies described whether the treatment protocol was or was not adapted in the case of a thin EMT (Basir et al., 2002; Zhang et al., 2005; Bozdag et al., 2009; Kinay et al., 2010; Zhao et al., 2012). The risk of information bias remained undetermined because most studies did not describe who measured the EMT (Basir et al., 2002; Zhang et al., 2005; Bozdag et al., 2009; Kinay et al., 2010; Zhao et al., 2012).

Characteristics of included patients are shown in Table II. All participants were scheduled for IVF or ICSI treatment. Cut-off values of EMT ranged from $\leq 6$ to $\leq 26.7$ mm. The mean age in the included studies ranged from 30.5 to 35.6 years.

**ROC analysis**

Observed accuracy estimates (sensitivity and $1 - $ specificity) for all studies and for all cut-off values reported, are plotted in the ROC space. Block size reflects the sample size of the studies for pregnant (horizontal) and non-pregnant (vertical) outcomes, respectively (Fig. 3). Overall, the predictive accuracy of EMT for non-pregnancy was low, as the sROC curve shows that EMT does not discriminate between cases that achieved a clinical pregnancy and cases that did not (area under the ROC: 0.56 and curve close to the $X = Y$ line).

Analyses for the different cut-off values show that sensitivity increases from near zero at a cut-off of $\leq 7$ mm [0.05 (95% CI 0.03–0.09)] to a sensitivity of 0.21 (95% CI 0.18–0.26) at a cut-off of $\leq 9$ mm. The specificity decreases at the same rate from close to 1 at $\leq 7$ mm [0.98 (95% CI 0.97–0.99)] to a minimum level of 0.85 at $\leq 9$ mm (95% CI 0.81–0.87) (Fig. 4A–C).
<table>
<thead>
<tr>
<th>Author</th>
<th>Consecutive</th>
<th>Design</th>
<th>Pro-/retrospective</th>
<th>Blinding</th>
<th>Outcome</th>
<th>Data per cycle</th>
<th>Selection bias</th>
<th>Information bias</th>
<th>Attrition bias</th>
<th>Verification bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rinaldi et al. (1996)</td>
<td>Yes</td>
<td>Cohort</td>
<td>NIA</td>
<td>Yes</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Weissman et al. (1999)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>CP, MR</td>
<td>Yes</td>
<td>Yes</td>
<td>NIA</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>De Geyter et al. (2000)</td>
<td>NIA</td>
<td>Cohort</td>
<td>Prospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>NIA</td>
<td>No</td>
<td>Yes</td>
<td>NIA</td>
</tr>
<tr>
<td>Basir et al. (2002)</td>
<td>NIA</td>
<td>Cohort</td>
<td>Prospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Dietterich et al. (2002)</td>
<td>No</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>CP, MR</td>
<td>Yes</td>
<td>Yes</td>
<td>NIA</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Yoeli et al. (2004)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Prospective</td>
<td>Yes</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Aboulishar et al. (2005)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Prospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Rashidi et al. (2005)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Prospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>NIA</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Zhang et al. (2005)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>NIA</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Amir et al. (2007)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>BP</td>
<td>Yes</td>
<td>Yes</td>
<td>NIA</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Grant et al. (2007)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Richter et al. (2007)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>LB, OP, CP, MR</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Al-Ghamdi et al. (2008)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Dechaud et al. (2008)</td>
<td>NIA</td>
<td>Cohort</td>
<td>Prospective</td>
<td>Yes</td>
<td>OP</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Bozdag et al. (2009)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>No</td>
<td>NIA</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Okohue et al. (2009)</td>
<td>NIA</td>
<td>Cohort</td>
<td>Prospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Traub et al. (2009)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>NIA</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Chen et al. (2010)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Kinay et al. (2010)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Prospective</td>
<td>No</td>
<td>CP, BP</td>
<td>Yes</td>
<td>Yes</td>
<td>NIA</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Kuc et al. (2011)</td>
<td>NIA</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>LB, CP, BP</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Singh et al. (2011)</td>
<td>NIA</td>
<td>Cohort</td>
<td>Prospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>NIA</td>
</tr>
<tr>
<td>Zhao et al. (2012)</td>
<td>Yes</td>
<td>Cohort</td>
<td>Retrospective</td>
<td>No</td>
<td>CP</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Studies were scored on patient sampling (consecutive or other), study design, data collection (prospective or retrospective) blinding (i.e. blinding of the sonographers for the medical history and the stimulation protocol used), selection bias (i.e. exclusion of cases because of reasons which affect the possibility of generalizing the results), information bias (i.e. measuring EMT by multiple sonographers), attrition bias (i.e. completeness of the outcome measures) and the stimulation protocol used. BP, biochemical pregnancy; CR, clinical pregnancy; OR, ongoing pregnancy; LR, live birth; MR, miscarriage rate; EP, ectopic pregnancy; NIA, no information available.
Insufficient data were present to perform ROC analysis for ongoing pregnancy rate or live birth.

**Odds ratios**

Live birth and ongoing pregnancy rate were included in the same meta-analysis. The meta-analysis for live birth and ongoing pregnancy rates together demonstrated a trend towards lower pregnancy rates for women with thin endometrium, although these differences did not reach statistical significance [OR 0.38 (95% CI 0.09–1.54), P = 0.18].

The I² statistic was 11% (Fig. 5). Insufficient data were available to perform this analysis for other cut-offs.

The forest plot with the calculated ORs for clinical pregnancy rate is shown in Fig. 6. The probability of conceiving with EMT ≤ 7 mm was significantly lower [OR 0.42 (95% CI 0.27–0.67), P = 0.0003] when compared with EMT > 7 mm. The I² statistic for heterogeneity was 41%, indicating that study heterogeneity was low.

As shown in Fig. 6, additional analysis for clinical pregnancy was performed to investigate whether the results would change when meta-analysis was performed with retrospective or prospective...
Figure 3  Summary ROC curve. EMT in the prediction of clinical pregnancy for all studies and all cut-off values reported. Block size reflects the sample size of the studies. EMT has no discriminatory capacity for clinical pregnancy (AUC-ROC 0.56 and curve close to the X = Y line). ROC, receiver operating characteristic; AUC, area under the curve.

studies only. The results for the retrospective studies remained significant [OR 0.46 (95% CI 0.29–0.71), P = 0.0006], while the results for prospective studies did not reach significance [OR 0.22 (95% CI 0.03–1.53), P = 0.13].

Although we focused on a cut-off of 7 mm in the present review, analyses for other cut-offs were also performed. As shown in Table III, there seems to be a positive association between EMT and prospects of clinical pregnancy. The probability of clinical pregnancy improves from low cut-offs [EMT ≤ 6 mm: OR 0.46 (95% CI 0.20–1.05)] to high cut-offs [EMT ≤ 12 mm: OR 0.89 (95% CI 0.55–1.44)], with significantly decreased pregnancy rates up to an EMT value of ≤ 10 mm.

There were insufficient data for miscarriage rate and biochemical and ectopic pregnancy to perform a meta-analysis.

Meta-regression

Only seven studies could be included in the meta-regression (Richter et al., 2007; Al-Ghamdi et al., 2008; Bozdag et al., 2009; Okohue et al., 2009; Chen et al., 2010; Kinay et al., 2010; Kuc et al., 2011). As shown in Table IV, the mean age in women with EMT ≤ 7 mm was higher (P-value < 0.001) and the mean number of oocytes retrieved was lower (P-value < 0.001) compared with the group with EMT > 7 mm. These factors are known to influence the chance of conceiving after IVF (Sharma et al., 2002), and we therefore considered them to be confounding factors for the effect of EMT on pregnancy prospects.

In the meta-regression, the characteristics female age (I² statistic 0%, P = 0.055) and number of oocytes retrieved (I² statistic 35%, P = 0.49) were not significantly associated with the outcome pregnancy. Consequently, analyses were not adjusted for female age and number of oocytes.

Meta-regression could not be performed for other cut-offs because insufficient data were available.

Discussion

The current review and meta-analysis summarizes the available knowledge concerning the accuracy of EMT in the prediction of pregnancy after IVF treatment. From the sROC curves (Fig. 3) it becomes clear that EMT has no predictive capacity for the occurrence of pregnancy. The discriminatory capacity slightly improved for lower cut-off values of EMT. However, the accuracy in predicting the occurrence of pregnancy remained poor for all cut-off values (Fig. 4A–C). Although EMT cannot be used to predict IVF outcome in terms of the occurrence of pregnancy (pregnant versus not pregnant), it does seem to be a factor for the assessment of the probability of conceiving after IVF. For clinical pregnancy rates, the probability of pregnancy was significantly lower in the group with thin EMT [EMT ≤ 7 mm: OR 0.42 (95% CI 0.27–0.67) P = 0.0003] (Fig. 6). Positive and negative predictive values for the outcome clinical pregnancy were 77 and 48%, respectively.

Because differences in female age and number of oocytes seemed to be present, the meta-regression had to rule out possible confounding. Surprisingly, female age and number of oocytes retrieved were not associated with pregnancy prospects. This finding can be explained by the fact that conventional relationships between pregnancy chances and oocyte number and female age were not consistent across studies, with some studies even showing opposite effects. This phenomenon is also known as ‘ecological fallacy’. Therefore, the confounding role for oocyte number and female age cannot be ruled out.

To the best of our knowledge, this is the first systematic review and meta-analysis that investigated both the independent predictive capacity and the prognostic value of EMT on pregnancy outcomes after IVF treatment. Studies investigating the significance of EMT in IVF or ICSI treatment cycles are conflicting. Several studies found an association between EMT and pregnancy (Kovacs et al., 2003; Traub et al., 2009), while others could not establish a significant correlation between EMT and the chance to conceive (Coulam et al., 1994; Csemiczky et al., 1999; IJland et al., 1999; Barker et al., 2009). In the meta-analysis of Momeni et al. (2011) the mean EMT appeared to be significantly higher in pregnant women compared with non-pregnant women [mean difference 0.4 mm (95% CI 0.22–0.58)]. Furthermore, the accuracy of EMT in predicting pregnancy appeared to be poor in many studies, which is in line with our findings (Basir et al., 2002; Rashidi et al., 2005; Zhang et al., 2005; Grant et al., 2007; Richter et al., 2007; Al-Ghamdi et al., 2008; Traub et al., 2009).

It is unclear why thin EMT is associated with poorer pregnancy prospects. Unfortunately, no studies have investigated the endometrial histology of patients with thin EMT. It has been speculated that basal layer endometrium oxygen concentrations are increased in patients with thin EMT. High oxygen levels might be detrimental for embryo implantation (Casper, 2011). In the review by Haouzi et al. (2012) different gene and protein-expression profiles are described which may play a critical role in endometrial receptivity in natural and stimulated cycles. In future research, molecular tools and morphological characteristics of the endometrium could possibly be studied to clarify if a relationship exists between thin EMT and endometrial receptivity. In addition, it...
Figure 4  ROC analyses for clinical pregnancy rate for different cut-offs of EMT. Pooled estimates for sensitivity and specificity are presented for (A) 7 mm, (B) 8 mm and (C) 9 mm. The pooled sensitivity and specificity values show that the discriminatory capacity slightly improves for lower cut-off values but the accuracy in predicting the occurrence of pregnancy remains poor. Sens, sensitivity; spec, specificity.
would be interesting to investigate the value of EMT in patients undergoing frozen embryo transfer cycles because in this population the possible negative effect of ovarian hyperstimulation on endometrial development is absent. Only one study reported on EMT in natural frozen embryo transfer cycles (Basir et al., 2002), and showed that pregnancy rates were significantly lower in patients with suboptimal endometrial development, suggesting an independent predictive role of EMT.

Endometrial receptivity may be reflected by endometrial stripe pattern. Eight studies also investigated the effect of endometrial pattern on pregnancy outcome (Dietterich et al., 2002; Rashidi et al., 2005; Dechaud et al., 2008; Bozdag et al., 2009; Chen et al., 2010; Kuc et al., 2011; Singh et al., 2011; Zhao et al., 2012) and the results are conflicting. Some studies found an association between endometrial pattern and IVF outcome (Dechaud et al., 2008; Chen et al., 2010; Kuc et al., 2011; Zhao et al., 2012), whilst other studies showed no significant correlation (Dietterich et al., 2002; Rashidi et al., 2005; Bozdag et al., 2009; Singh et al., 2011). Unfortunately, meta-analysis could not be performed because different classification systems for endometrial pattern were used in the separate studies.

**Figure 5** Forest plot of ongoing pregnancy and live birth for women with EMT ≤ 7 mm and EMT > 7 mm. Chances of live birth and ongoing pregnancy are lower for women with EMT ≤ 7 mm, but the results are not significant. The I² statistic of 11% indicates that study heterogeneity was low. CI, confidence interval.

**Figure 6** Forest plot of clinical pregnancy for women with EMT ≤ 7 mm and women with EMT > 7 mm. The probability of clinical pregnancy is significantly lower for women with EMT ≤ 7 mm. The I² statistic was 41%, indicating that study heterogeneity was low.
The relevance of endometrial thickness in IVF

Limitations

Some limitations of this systematic review need to be addressed. Although study heterogeneity was low, differences in study quality characteristics and study populations were present. Furthermore, the definition of thin endometrium was not uniform in all studies and different cut-off values were used. By constructing an ROC curve it was possible to assess the effect of different cut-off values on the accuracy of EMT. Additionally, an EMT ≤ 7 mm was defined, by the authors, as thin EMT in the prognostic analyses. This cut-off was chosen because most of the studies reported thin EMT as ≤ 7 mm. It must be noted that pregnancy rates improved with increasing EMT, but that pregnancy rates still appeared to be significantly lower up to an EMT value of 10 mm (Table III). It may therefore be incorrect to use one specific cut-off value to discriminate thin from normal EMT.

Another limitation is the reproducibility of EMT assessment. Karlsson et al. (1994) described a mean discrepancy of 1.5 mm in measurements of the endometrium between experienced and inexperienced sonographers. Delisle et al. (1998), however, described a good inter- and intra-observer reproducibility of 95 and 94%, respectively. Also differences in equipment (including resolution of ultrasound machines) could cause differences in EMT measurements. For these reasons, a bivariate model was used for the ROC analysis in the present review, which allows for variations in the assessment of EMT and the choice of cut-off values.

It must also be mentioned that only seven studies could be included in the meta-regression because the other studies reported insufficient data. As a consequence it may not be correct to make conclusions about the influence of potential confounders for the review as a whole. Moreover, meta-regression was only performed for female age and number of oocytes retrieved. Insufficient data were available for other potential confounders such as embryo quality and hormonal parameters. Further research must reveal the possible interrelation of other factors with EMT.

Lastly, many retrospective studies were included which are associated with an inevitable risk of bias. The overall risk of bias in this meta-analysis appeared to be high (Fig. 2). Only three studies described binding of the sonographers and selection bias appeared to be present in most of the studies. The risk of verification and information bias remained undetermined due to insufficient information. Results were no longer significant and the effect size decreased by half if only prospective studies were

<table>
<thead>
<tr>
<th>Table III</th>
<th>ORs and 95% CI for clinical pregnancy at different EMT cut-off values.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cut-off (mm)</td>
<td>No. of patients</td>
</tr>
<tr>
<td>6</td>
<td>6886</td>
</tr>
<tr>
<td>7</td>
<td>10,612</td>
</tr>
<tr>
<td>8</td>
<td>7247</td>
</tr>
<tr>
<td>9</td>
<td>8,221</td>
</tr>
<tr>
<td>10</td>
<td>7,446</td>
</tr>
<tr>
<td>11</td>
<td>7,451</td>
</tr>
<tr>
<td>12</td>
<td>7,873</td>
</tr>
</tbody>
</table>

ORs for clinical pregnancy improve with increasing thickness. However, ORs remain significantly lower up to a thickness of 10 mm.

<table>
<thead>
<tr>
<th>Table IV</th>
<th>Female age and number of oocytes retrieved in cases with EMT ≤ 7 mm versus EMT &gt; 7 mm.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>Age (years) ≤7 mm</td>
</tr>
<tr>
<td>Richter et al. (2007)</td>
<td>34.9 (±3.6)</td>
</tr>
<tr>
<td>Al-Ghamdi et al. (2008)</td>
<td>33.3 (±4.5)</td>
</tr>
<tr>
<td>Bozdag et al. (2009)</td>
<td>33.1 (±4.7)</td>
</tr>
<tr>
<td>Okohue et al. (2009)</td>
<td>30.8 (±3.3)</td>
</tr>
<tr>
<td>Chen et al. (2010)</td>
<td>31.7 (±3.8)</td>
</tr>
<tr>
<td>Kinay et al. (2010)</td>
<td>36.7 (±3.2)</td>
</tr>
<tr>
<td>Kuc et al. (2011)</td>
<td>31.3 (±3.1)</td>
</tr>
</tbody>
</table>

Data are shown as mean (SD). Mean age in women with EMT ≤ 7 mm was significantly higher (P < 0.001) and the mean number of oocytes retrieved was significantly lower (P < 0.001) compared with the group with EMT > 7 mm.
 included. This may have been due to the small number of cases in these studies (n = 357). Therefore, the significance of the results is probably mainly due to the effect of the retrospective studies.

**Strengths**

To the best of our knowledge this is the first review that not only summarizes the evidence on the significance of EMT in IVF but also investigates the role of possible confounders. An extended search was performed in the electronic databases Pubmed, Embase and Cochrane, resulting in 22 studies that could be included. The quality of the included studies was assessed separately by two authors. Finally, different cut-off values were included in the analyses.

**Implications for clinical practice**

Physicians are regularly confronted with the difficult decision of whether treatment should be continued in cycles with thin EMT or whether embryos should be cryopreserved for potentially better pregnancy chances in subsequent frozen embryo replacement cycles. Based on the findings of the current review it can be concluded that cancelling IVF treatment cycles seems not to be justified based on solely a poor EMT. It is an important finding that pregnancy prospects improve with a thicker EMT. However, the results must be interpreted with caution because many retrospective studies were included. The significance of the results disappeared when retrospective studies were excluded from the meta-analysis and the effect of potential confounders in these studies therefore cannot be ruled out. Finally, differences in pregnancy prospects in relation to EMT as reported in the literature may not be sufficient reason to have couples refrain from continuing treatment.

Further research must reveal the possible interrelation of the factors female age and oocyte number with EMT and thereby with the occurrence of pregnancy. Such research should preferably be based on individual patient data meta-analysis from multiple studies, which would solve many of the methodological problems mentioned above. With such a large data set, the accuracy of combining prognostic variables could be assessed properly and the real independent predictive capacity of EMT could be determined. Furthermore, a histology study might be of value to unravel the pathophysiology at the endometrium level.

**Summary**

This systematic review and meta-analysis demonstrates that EMT has limited capacity to identify women who have a low chance to conceive after IVF. The frequently reported cut-off of 7 mm is only related to a lower chance of pregnancy, but occurs infrequently. Therefore, the use of EMT as a tool to decide on cycle cancellation, freezing of all embryos or refraining from further IVF treatment seems not to be justified based on the current meta-analysis. However, further research on the real independent significance of EMT in IVF is needed because of methodological weaknesses of the included studies.

**Acknowledgements**

The authors thank the corresponding authors for their response to the requests for additional information.

**Supplementary data**

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

**Authors’ roles**

A.K. and J.S. searched the electronic databases, selected studies, performed data extraction and statistical analysis and took the lead in writing the manuscript. M.J.C.E. provided statistical support and performed the meta-regression analyses. B.O. provided the ROC analysis. J.F.M.B., H.L.T. and B.W.M. revised several drafts of the manuscript, all authors confirmed with the final version of the manuscript.

**Funding**

No monetary compensation was received for conducting this review.

**Conflict of interest**

None of the authors has any conflict of interest related to this work.

**References**


