Human Reproduction Open, pp. 1–18, 2019
doi:10.1093/hropen/hoy025

human reproduction open

REVIEW

Endometrial scratching prior to IVF; does it help and for whom? A systematic review and meta-analysis

N.E. van Hoogenhuijze [1,*], J.C. Kasius[2], F.J.M. Broekmans[1], J. Bosteels[3], and H.L. Torrance[1]

[1]Department of Reproductive Medicine and Gynecology, University Medical Center Utrecht, Utrecht University, Room Number: F05.126, Postbox 85500, 3508 GA Utrecht, The Netherlands [2]Department of Gynaecology, Amsterdam University Medical Centre, location AMC, Postbox 22660, 1100 DD Amsterdam, The Netherlands [3]Department of Obstetrics and Gynaecology, Imelda Hospital, Bonheiden, Belgium Centre for Evidence-Based Medicine, Cochrane Belgium, 3000 Leuven, Belgium

*Correspondence address. E-mail: n.e.vanhoogenhuijze@umcutrecht.nl orcid.org/0000-0001-9057-1306

Submitted on June 16, 2018; resubmitted on December 6, 2018; editorial decision on December 14, 2018; accepted on December 19, 2018

STUDY QUESTION: What is the effect of endometrial scratching in patients with or without prior failed ART cycles on live birth (LBR) and clinical pregnancy rates (CPR)?

SUMMARY ANSWER: It remains unclear if endometrial scratching improves the chance of pregnancy and, if so, for whom.

WHAT IS KNOWN ALREADY: Endometrial scratching is hypothesized to improve embryo implantation in ART. Multiple studies have been published, but it remains unclear if endometrial scratching actually improves pregnancy rates and, if so, for which patients.

STUDY DESIGN, SIZE, DURATION: For this review, a systematic search for published articles on endometrial scratching and ART was performed on 12 February 2018, in Pubmed, Embase and the Cochrane Library.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Randomized controlled trials (RCTs) that evaluated endometrial scratching in the cycle prior to the stimulation cycle and reported CPR or LBR were included. RCTs investigating the effect of scratching during the stimulation cycle, or prior to cryo-thaw cycles were excluded. Studies were assessed using the Cochrane Risk of Bias tool. The effect of scratching was assessed for three different patient groups: patients with no prior IVF/ICSI treatment (Group 0), patients with one failed full IVF/ICSI cycle, including cryo-thaw cycles (Group 1) and patients with two or more failed full IVF/ICSI cycles (Group 2). A meta-analysis was performed when statistical heterogeneity was low; otherwise, a descriptive analysis was performed.

MAIN RESULTS AND THE ROLE OF CHANCE: Fourteen RCTs involving 2537 participants were included. Most RCTs contained a high or unclear risk of bias on one or more items. Substantial clinical and statistical heterogeneity was present; therefore meta-analysis for LBR and CPR could only be performed on Group 1. For this group, no differences between scratch and control were found for both LBR (risk ratio (RR) 1.01 [95%CI 0.68–1.51]) and CPR (RR 1.04 [95%CI 0.74–1.45]). For Groups 0 and 2, pooled analysis could not be performed, and for both groups the results of the individual RCTs were negative, neutral and positive. Miscarriage and multiple pregnancy rates were evaluated for the three groups (0, 1 and 2) together. Both outcomes were not significantly different between scratch and control (miscarriage rate RR 0.82 [95%CI 0.57–1.17] and multiple pregnancy rate RR 1.06 [95%CI 0.84–1.35]). Subgroup analysis, excluding trials with a risk of unintentional endometrial injury in the control group, was performed for Group 0 and 2 for LBR and CPR, and for the overall groups for miscarriage rate and multiple pregnancy rate. This reduced the heterogeneity and allowed for pooled analysis in these subgroups. Results of pooled analysis for the subgroups of Group 0 and 2 showed no significant difference for LBR, but CPR was significantly improved after endometrial scratching (Group 0 RR 1.28 [95%CI 1.02–1.62] and Group 2 RR 2.03 [95%CI 1.20–3.43]). Subgroup analysis of the overall groups showed no significant difference for miscarriage and multiple pregnancy rate.

LIMITATIONS REASONS FOR CAUTION: The main limitations were that many RCTs had a high or unclear risk of bias on one or several items, clinical heterogeneity was still present despite categorizing into three populations, and that not all RCTs could be included in the analyses because separate data for our three groups could not be provided.

WIDER IMPLICATIONS OF THE FINDINGS: It remains unclear if endometrial scratching improves the chance of pregnancy for women undergoing ART and, if so, for whom. This means endometrial scratching should not be offered in daily practice until results from large and well-designed RCTs and an individual patient data analysis become available.

STUDY FUNDING/COMPETING INTERESTS: No specific funding was sought for the study. The Department of Reproductive Medicine and Gynaecology funds of the University Medical Center of Utrecht were used to support the authors throughout the study period and preparation of the manuscript. None of the authors has a conflict of interest to declare.

REGISTRATION NUMBER: Not applicable.

Key words: endometrial scratching / endometrial injury / IVF / ICSI / implantation failure / embryo implantation / ART

WHAT DOES THIS MEAN FOR PATIENTS?

Many couples having problems to conceive receive fertility treatment such as IVF or ICSI. During these treatments, embryos are cultured in the lab and subsequently transferred into the woman's womb. However, approximately only one out of three (35%) embryo transfers results in pregnancy. This leads to a high physical, emotional and financial burden for these couples.

Previous research has suggested that mild injury to the lining of the womb, called 'endometrial scratching', may improve embryo implantation. While multiple studies have been carried out, two important problems are that they were performed on small groups of women and that the women differed greatly in, for example, age, duration of infertility and number of previous treatments.

This review summarizes findings from all the published studies and analyses the data for groups of women who are more similar, i.e. women with no prior IVF/ICSI treatment, women with one failed IVF/ICSI cycle, and women with two or more failed IVF/ICSI cycles. The results show that endometrial scratching does not improve the chance of pregnancy or live birth in women with one failed IVF/ICSI treatment. However, this conclusion is based on studies that are of moderate quality. For the other two groups, the results of the various studies differed so much (negative, neutral or positive effect) that no conclusion could be drawn. Thus, it is still unclear if endometrial scratching really improves the chance of pregnancy, and if so, for whom.

Introduction

Subfertility is a problem that affects many couples globally. In developed countries, 3.5–16.7% of the population suffers from infertility with more than half of these couples seeking medical assistance (Boivin *et al.*, 2007). Some couples never start medical treatment, but others pursue pregnancy with treatments such as IUI, IVF or ICSI (Boivin *et al.*, 2007). This results in over 500 000 IVF/ICSI treatments being carried out annually in Europe (De Geyter *et al.*, 2018). While the efficacy of these treatments has improved, embryo implantation still only occurs in approximately 35% of embryo transfers (De Geyter *et al.*, 2018). Endometrial scratching during ART has been advocated to improve the chance of embryo implantation, but to date it is unclear how it should be performed, whether it is effective, and for whom it might be effective (Panagiotopoulou *et al.*, 2015; Zygula *et al.*, 2016). Despite these concerns, scratching has been implemented widely with a staggering 83% of physicians recommending it to patients undergoing IVF or ICSI in the UK, New Zealand and Australia (Lensen *et al.*, 2016).

Endometrial scratching as a treatment to improve pregnancy rates was first proposed by Barash *et al.* (Granot *et al.*, 2000; Barash *et al.*, 2003). In 2000, they performed endometrial biopsies to study endometrial characteristics of subfertile women who had failed to conceive after IVF. The biopsies were taken in the natural cycle of the month prior to ovarian stimulation for IVF treatment. Coincidentally, they noted a remarkable pregnancy rate (11 out of 12 participants) in women participating in this study (Granot *et al.*, 2000). This led them to conduct a quasi-randomized study from which they concluded that endometrial injury doubled the pregnancy rate in subfertile women (Barash *et al.*, 2003). From then on, multiple studies on the effect of endometrial injury in women undergoing ART have been performed, reporting positive, negative or neutral effects of endometrial injury (Nastri *et al.*, 2015; Panagiotopoulou *et al.*, 2015; Ko and Ng, 2016; Zygula *et al.*, 2016). Along with the rapid publication of these studies, multiple reviews have been published. Comparable to the wide spectrum of results of the randomized controlled trials (RCTs), the reviews also differ in their interpretation of the results. While some reviews concluded that endometrial scratching is associated with increased clinical pregnancy rate (CPR) and/or live birth rate (LBR) (El-Toukhy *et al.*, 2012; Potdar *et al.*, 2012; Nastri *et al.*, 2015; Ko and Ng, 2016), others stated that no conclusion could yet be drawn (Panagiotopoulou *et al.*, 2015; Santamaria *et al.*, 2016; Zygula *et al.*, 2016). Importantly, all reviews stress that (part of) their conclusions are based on moderate-quality studies that were compromised by high clinical heterogeneity, and/or that endometrial scratching should not be implemented in daily practice for all patients since it remains unclear if it is beneficial for all patients or only for certain subgroups. Moreover, while these studies all focused on clinical effect, the biological mechanisms behind the possible effect on implantation have remained hypothetical (Lass *et al.*, 1998; Li and Hao, 2009; Gnainsky *et al.*, 2010,

2015). Taken together, the studies that have been performed so far are of moderate quality at best, apply different methods of endometrial scratching, have been performed on a very diverse population and lack support of a biological mechanism. In line with this, multiple reviews conclude that convincing evidence for a beneficial effect is still lacking (Nastri *et al.*, 2015; Panagiotopoulou *et al.*, 2015; Ko and Ng., 2016; Zygula *et al.*, 2016). Meanwhile, it seems that most clinicians would advise scratching during IVF/ICSI for patients with repeated implantation failure (RIF), leading to implementation of a therapy but with very limited science to support this practice (Lensen *et al.*, 2016).

Since the publication of the abovementioned reviews, multiple RCTs on endometrial injury prior to ART have been carried out. While the first trials mainly focused on patients with RIF, more recent studies have shifted their focus to also include patients in earlier stages of IVF/ICSI treatment (Mahran *et al.*, 2016; Liu *et al.*, 2017; Tk *et al.*, 2017; Maged *et al.*, 2018). The increase in number of published studies not only calls for an updated review of the existing body of evidence, but also has created the possibility to subgroup patients according to the number of failed IVF/ICSI cycles. Moreover, it enables a comparison of RCTs with more similar timing of endometrial injury, namely in the menstrual cycle prior to ovarian stimulation or in the cycle of ovarian stimulation, thereby reducing clinical heterogeneity. Thus, in comparison with previously published reviews, this review could include more trials—a few of which have relatively large populations—while reducing clinical heterogeneity by selecting more similar scratching methods and by defining three different populations.

The aim of this review was therefore to assess whether endometrial scratching in the cycle prior to IVF/ICSI is more effective in patients with an increasing number of previously failed IVF/ICSI cycles, in terms of LBR and CPR.

Materials and Methods

Literature search

A literature search was performed on 12 February 2018 in Pubmed, EMBASE, and Cochrane library. The search contained synonyms for the population under study and intervention (Table I). No language restrictions or other search limitations were applied. Duplicates were removed, after which three reviewers (J.K., J.B. and N.H.) screened studies based on title and abstract (Fig. 1). Studies were selected that were RCTs comparing endometrial scratching performed in the cycle before the stimulation cycle of fresh IVF/ICSI treatment to no scratching or a sham procedure. Thus, case-control studies, comments to the editor, abstracts of unpublished data or data of ongoing trials were excluded. Any disagreement between the reviewers on selected or rejected titles was solved by discussion or by the judgment of a fourth reviewer (H.T.). If the university library could not provide the full text, the corresponding author was contacted.

Subsequently, the full text was screened for additional exclusion criteria. All studies had to report LBR or CPR after the subsequent fresh IVF/ICSI cycle as an outcome parameter. Thus, trials investigating the effect of scratching during the stimulation cycle or prior to cryo-thaw cycles were excluded. Also, studies investigating therapeutic hysteroscopy of intrauterine pathology as a proxy treatment of an endometrial scratching procedure were excluded (trials in which hysteroscopy was performed during fertility workup were not excluded since it was not a proxy treatment).

Risk of bias and quality assessment

Three authors (J.K., J.B. and N.H.) assessed the included studies for risk of bias and methodological quality by using the Cochrane 'Risk of bias assessment tool' (Higgins *et al.*, 2011). By thorough examination of the full text of the selected manuscripts, information was collected on random sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and personnel (detection bias), blinding of outcome assessors (performance bias), incomplete data (attrition bias), selective data reporting (selection bias) and other types of bias. The risk of bias and quality of a study was incorporated into the interpretation of the review findings and fully described in a study characteristic table.

Checking the study results in terms of obvious or insufficient outcomes ensured an awareness of selective reporting within studies.

Outcome measures

The primary outcome measure was LBR from the cycle following randomization. The secondary outcome measures were CPR, miscarriage rate and multiple pregnancy rate. Live birth was defined as the number of patients with a

Table I Exact copy of the query used for each database search: the keywords were 'IVF' and 'endometrial scratching' and synonyms.

Database	Search query
PUBMED	(IVF [Title/Abstract] OR (*in vitro* fertilization [Title/Abstract]) OR implantation [Title/Abstract] OR ICSI [Title/Abstract] OR (intracytoplasmic sperm injection [Title/Abstract]) OR embryo [Title/Abstract] OR (assisted reproduction [Title/Abstract]) OR (assisted reproductive [Title/Abstract])) AND (endometrium [Title/Abstract] OR endometrial [Title/Abstract] OR pipelle [Title/Abstract] OR novak [Title/Abstract]) AND (injury [Title/Abstract] OR scratching [Title/Abstract] OR scratch [Title/Abstract] OR biopsy [Title/Abstract] OR disruption [Title/Abstract])
EMBASE	'ivf':ab,ti OR 'in vitro fertilization':ab,ti OR 'implantation':ab,ti OR 'icsi':ab,ti OR 'intracytoplasmic sperm injection':ab,ti OR 'embryo':ab,ti OR 'assisted reproduction':ab,ti OR 'assisted reproductive':ab,ti AND ('endometrium':ab,ti OR 'endometrial':ab,ti OR 'pipelle':ab,ti OR 'novak':ab,ti) AND ('injury':ab,ti OR 'scratching':ab,ti OR 'scratch':ab,ti OR 'biopsy':ab,ti OR 'disruption':ab,ti)
Cochrane	('IVF':ti,ab,kw OR '*in vitro* fertilization':ti,ab,kw OR 'implantation':ti,ab,kw OR 'ICSI':ti,ab,kw OR 'intracytoplasmic sperm injection':ti,ab,kw OR 'embryo':ti,ab,kw OR 'assisted reproduction':ti,ab,kw OR 'assisted reproductive':ti,ab,kw) AND ('endometrium':ti,ab,kw OR 'endometrial':ti,ab,kw OR 'pipelle':ti,ab,kw OR 'novak':ti,ab,kw) AND ('injury':ti,ab,kw OR 'scratching':ti,ab,kw OR 'scratch':ti,ab,kw OR 'biopsy':ti,ab,kw OR 'disruption':ti,ab,kw)
ClinicalTrials.gov	(IVF OR implantation OR ICSI OR embryo OR 'assisted reproduction' OR 'assisted reproductive') AND (endometrium OR endometrial OR pipelle OR novak) AND (injury OR scratching OR scratch OR biopsy OR disruption)

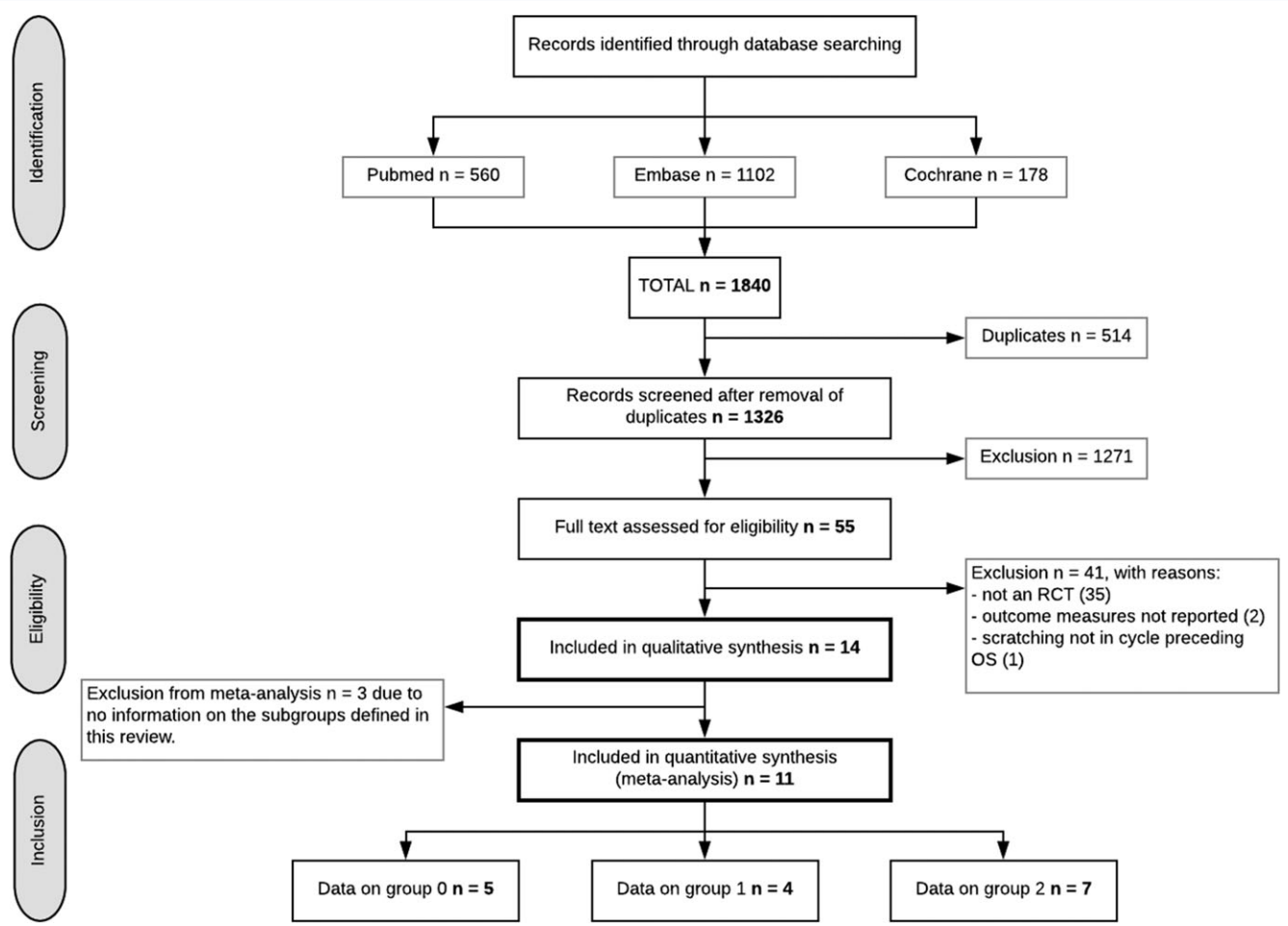


Figure 1 Flowchart of the systematic literature search. Search date: 12 February 2018. *n*, number of records; RCT, randomized controlled trials; OS, ovarian stimulation; Group 0, no prior IVF/ICSI treatment; Group 1, one failed IVF/ICSI treatment; Group 2, two or more failed IVF/ICSI treatments.

live born infant. Clinical pregnancy was defined as visibility of a gestational sac on ultrasound. Miscarriage was defined as any loss of pregnancy before 24 weeks of gestation. Multiple pregnancy was defined as multiple foetuses or multiple gestational sacs on ultrasound, or the birth of a twin or triplet.

Data analysis

Study characteristics were extracted from each manuscript and recorded in a database system (RevMan 5.3, The Cochrane Collaboration, Oxford, UK). To evaluate risk of bias and clinical heterogeneity a summary table describing the study characteristics was constructed. Data were extracted from each manuscript for the three populations of interest: participants with no prior IVF/ICSI treatment (Group 0), participants with one failed complete IVF/ICSI cycle (Group 1), and participants with two or more failed complete IVF/ICSI cycles (Group 2). A complete IVF/ICSI cycle was defined as ovarian stimulation with oocyte retrieval and all subsequent embryo transfers, including fresh and/or cryo-thaw transfers. If the published report combined the results of Groups 0, 1 and 2, the authors were contacted by email to collect data for the separate groups. The choice of these groups followed from the finding that most physicians tend to perform endometrial scratching in patients with RIF (Lensen *et al.*, 2016), which is frequently defined as failure of implantation after 2–6 IVF/ICSI treatment cycles (Tan *et al.*, 2005), and from looking at it from the patients' perspective to find out the earliest moment that endometrial scratching would be beneficial.

When statistical heterogeneity was low or moderate (meaning $I^2 \leq 50\%$ (Higgins *et al.*, 2003; Higgins and Green., 2011)), the study results were pooled in forest plots and expressed in risk ratios (RR) with 95% CIs using RevMan 5.3. If statistical heterogeneity was high (meaning $I^2 > 50\%$ (Higgins and Green., 2011)), the study results were not pooled but were summarized in the text. In case of borderline risk of heterogeneity, pooled analysis was performed if a low-moderate chance of heterogeneity was expected based on visually examining the forest plots (i.e. overlapping CIs) (Higgins *et al.*, 2011). A random-effect model was used because it was expected that the true effect size would differ between the studies. However, since random-effect models are less precise when a small number of studies are included, a fixed-effect model was used if less than five RCTs were included (Borenstein *et al.*, 2010). The study data were processed according to the intention-to-treat principle. When results were only reported for patients with an embryo transfer, assumptions were made that the participants who had a cycle cancellation after randomization did not conceive.

Subgroup analysis

Subgroup analysis was performed by excluding studies with a risk of unintentional endometrial injury in the control group. Trials that performed an

intracervical or intrauterine sham procedure, saline infusion sonohysterography, hysterosalpingogram or hysteroscopy in the control group in the cycle prior to the ovarian stimulation cycle were excluded from these analyses. Studies that had performed such procedures during workup were not excluded from subgroup analysis.

Results

Search

After removal of duplicates, the structured literature search resulted in 1326 citations (Fig. 1). After screening the titles, 55 citations remained for abstract and full text reading after which 14 RCTs, comprising a total of 2537 patients, were included. A total of 11 of these 14 RCTs could be included for data analysis. Gibreel *et al.* 2015 and Narvekar *et al.* 2010 only reported data for participants with one or more failed IVF/ICSI cycles and separate data corresponding to our Groups 1 and 2 was not available, so that these studies could not be included in the meta-analysis. Safdarian *et al.* 2011 did not report on the number of previously failed IVF/ICSI cycles so that their data could not be categorized to fit one of our populations. Unfortunately, attempts to contact the authors of these three studies were unsuccessful.

The data of Nastri *et al.* was also not split according to our subgroups, but we received subgroup data through personal communication. However, separate data was only available for Group 0 and Group 1 but not for Group 2 (Nastri *et al.*, 2013). Thus, part of their data could not be included in our analysis.

Risk of bias and methodological quality

The risk of bias of the included RCTs is described in Table II. For seven RCTs, the risk of selection bias was unclear because the method of allocation was not clearly described (Karimzadeh *et al.*, 2009; Baum *et al.*, 2012; Inal *et al.*, 2012; Singh *et al.*, 2015; Mahran *et al.*, 2016; Liu *et al.*, 2017; Maged *et al.*, 2018). Random sequence generation was interpreted to be at a low risk of selection bias because this was performed by computer-generated sequence (Narvekar *et al.*, 2010; Safdarian *et al.*, 2011; Inal *et al.*, 2012; Nastri *et al.*, 2013; Yeung *et al.*, 2014; Gibreel *et al.*, 2015; Singh *et al.*, 2015; Tk *et al.*, 2017; Maged *et al.*, 2018), a table of random numbers (Baum *et al.*, 2012; Liu *et al.*, 2017), sealed envelopes (Mahran *et al.*, 2016) or a bag of papers with equal amounts of intervention and control groups (Karimzadeh *et al.*, 2009).

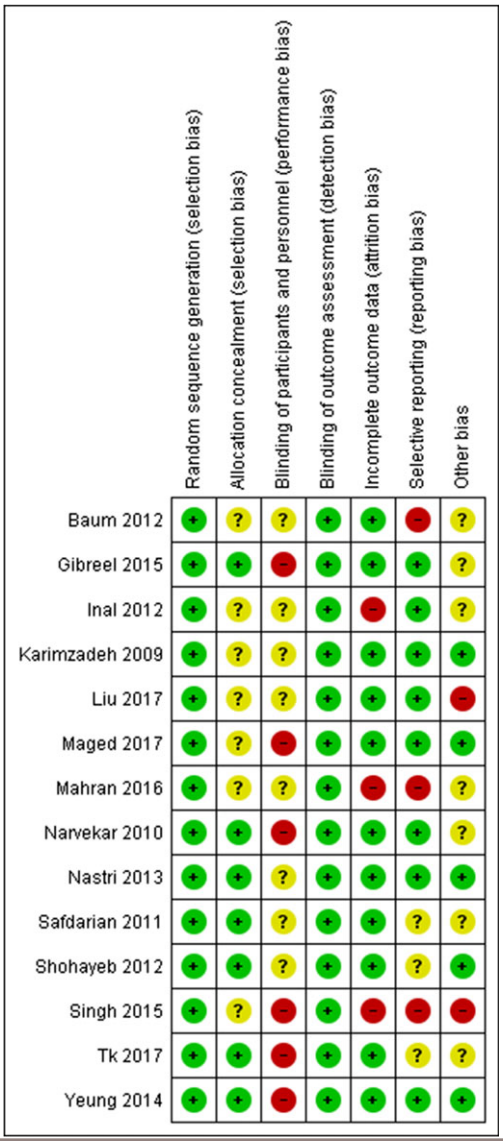
It has been reported that embryo transfer technique can influence pregnancy rates, and therefore not blinding the physician could have affected the study outcomes (Mains and Van Voorhis., 2010). Thus, a high risk of performance bias was suspected in six studies, because the physician was not blinded (Narvekar *et al.*, 2010; Yeung *et al.*, 2014; Gibreel *et al.*, 2015; Singh *et al.*, 2015; Tk *et al.*, 2017; Maged *et al.*, 2018). The remaining eight studies did not describe blinding and were therefore rated as unclear risk of performance bias.

With regard to detection bias, not blinding was thought to have no impact on the outcomes because of the objective nature of the primary and secondary outcomes. Therefore, all studies were thought to have a low risk of detection bias.

A high risk of attrition bias was found in three RCTs because of unclear inclusion and exclusion criteria that could have led to subjective inclusion/exclusion of participants (Inal *et al.*, 2012), not reporting the number of patients lost to follow-up (Inal *et al.*, 2012) or because the outcome measures were not pre-specified (Inal *et al.*, 2012; Singh *et al.*, 2015; Mahran *et al.*, 2016).

Three studies had a high risk of reporting bias because (most) outcomes were reported in percentages instead of numbers of events (Baum *et al.*, 2012; Singh *et al.*, 2015; Mahran *et al.*, 2016) or because there were discrepancies in the reported data (Singh *et al.*, 2015; Mahran *et al.*, 2016). For three studies, reporting bias was interpreted as unclear risk because the

Table II Risk of bias summary for the randomized controlled studies.



The chart presents the authors' judgment of each risk of bias item of each included study.
+, low risk of bias;?, uncertain risk of bias; −, high risk of bias.
Other bias: absence of power calculations, underpowered studies, or performance of procedures other than the studied intervention that could also lead to endometrial injury (i.e. hysteroscopy or sham). Note: Patients and personnel were usually not blinded for the intervention. Not blinding personnel could have led to adjusted embryo transfer techniques and was therefore regarded as 'high risk' for performance bias. On the other hand, outcome measures (live birth and clinical pregnancy rates) were unlikely to be influenced by unblinded personnel, so that all studies were regarded as 'low risk' for detection bias.

intention-to-treat analyses were not reported but had to be deduced from other numbers (Tk et al., 2017; Safdarian et al., 2011; Shohayeb and El-Khayat, 2012). This also applied to the studies that had already been judged to have a high risk of reporting bias.

Other bias consisted of unclear methodological quality (Safdarian et al., 2011; Singh et al., 2015; Tk et al., 2017), cervical sham procedure (Baum et al., 2012; Gibreel et al., 2015), intracavitary sham procedure or hysteroscopy or hysterosalpingography in the month of scratching (Narvekar et al., 2010; Safdarian et al., 2011; Liu et al., 2017). Unclear methodological quality was deduced from, for example, not reaching the calculated sample size or no sample size calculation, or retrospective registration of the trial in a trial register (Safdarian et al., 2011; Singh et al., 2015; Tk et al., 2017).

Several studies reported having performed a hysteroscopy during workup, which was interpreted as a low risk of bias because the hysteroscopies were not performed in the same menstrual cycle as the endometrial injury.

Participants

In line with the inclusion criteria of this review, all participants underwent a fresh IVF/ICSI cycle after randomization. Inclusion and exclusion criteria of each study varied, but factors that are known to affect implantation—such as hydrosalpinx or cavity abnormalities—were excluded in most RCTs. An overview of the study characteristics is shown in Table III.

Each publication provided a baseline table describing at least the mean age and the IVF/ICSI cycle characteristics. None of the studies reported significant differences in demographic baseline characteristics, except for Tk et al. who reported a higher total dose of FSH in the control group than in the intervention group (Tk et al., 2017).

The mean age of the participants was 29–32 years, except for two studies that reported a mean age of 34.6 and 36.5 years, respectively (Baum et al., 2012; Yeung et al., 2014). The mean BMI was 25–26.7 kg/m^2 in seven studies (Narvekar et al., 2010; Safdarian et al., 2011; Shohayeb and El-Khayat., 2012; Gibreel et al., 2015; Singh et al., 2015; Tk et al., 2017; Maged et al., 2018), around 22 kg/m^2 in two studies (Yeung et al., 2014; Liu et al., 2017) and 28.5 kg/m^2 in one study (Mahran et al., 2016). The participants were of diverse ethnic backgrounds.

The mean number of previously failed IVF/ICSI cycles was not always reported (Safdarian et al., 2011; Tk et al., 2017), but it varied from 0 (Mahran et al., 2016; Liu et al., 2017; Maged et al., 2018) up to 8.5 failed cycles (Baum et al., 2012), with the majority of studies including patients with 1–3 failed cycles (Karimzadeh et al., 2009; Narvekar et al., 2010; Inal et al., 2012; Shohayeb and El-Khayat., 2012; Gibreel et al., 2015). The mean number of embryos transferred in the post-randomization cycle was 1.75–2.55, but five RCTs transferred on average 3–3.6 embryos (Narvekar et al., 2010; Safdarian et al., 2011; Shohayeb and El-Khayat., 2012; Singh et al., 2015; Mahran et al., 2016).

Intervention

In most RCTs, endometrial scratching was performed using an endometrial biopsy catheter in the luteal phase of the cycle preceding ovarian stimulation (Karimzadeh et al., 2009; Narvekar et al., 2010; Safdarian et al., 2011; Baum et al., 2012; Inal et al., 2012; Nastri et al., 2013; Yeung et al., 2014; Gibreel et al., 2015; Mahran et al., 2016; Liu et al., 2017; Tk et al., 2017; Maged et al., 2018). Singh and Shohayeb used a Karman's cannula and Novak curette, respectively (Shohayeb and El-Khayat, 2012; Singh et al., 2015).

Six studies performed the endometrial biopsy more than once in the same cycle (Narvekar et al., 2010; Baum et al., 2012; Inal et al., 2012; Gibreel et al., 2015; Liu et al., 2017; Tk et al., 2017). Furthermore, two RCTs performed the endometrial scratching during a standard hysteroscopy which all patients underwent during the follicular phase prior to ovarian stimulation (Narvekar et al., 2010; Shohayeb and El-Khayat., 2012). Narvekar et al. (2010), repeated the endometrial biopsy in the luteal phase without hysteroscopy. Unfortunately, these studies did not report whether intrauterine abnormalities were detected or treated at hysteroscopy.

Four studies were single-blinded (patient only) and performed a sham procedure in the control group by drying the cervix with a gauze (Nastri et al., 2013), introducing an endometrial biopsy catheter or uterine sound up to the internal ostium (Baum et al., 2012; Gibreel et al., 2015), or up to the uterine fundus (Liu et al., 2017).

Outcome measures

LBR

The results of Group 0 could not be pooled due to substantial statistical heterogeneity (I^2 88%). As is shown in Fig. 2a, three out of four studies did not show a significant difference between the control and intervention group (Nastri et al., 2013; Yeung et al., 2014; Liu et al., 2017), with only the study by Mahran et al. showing a significantly improved LBR after endometrial scratching (RR 2.39 [95% 1.87–3.06]) (Mahran et al., 2016). After exclusion of trials with a risk of unintentional endometrial injury in the control group, the results of the two remaining trials, including a total of 227 patients, could be pooled (Nastri et al., 2013; Yeung et al., 2014). No significant difference in LBR was found between the intervention and control groups (RR 0.95 [95%CI 0.64–1.41], total $n = 227$ participants) (Fig. 2a).

In Group 1, results could be pooled (I^2 41%) and showed no significant difference in LBR after scratching as compared to not scratching (RR 1.01 [95%CI 0.68–1.51], total $n = 253$ participants) (Fig. 2b) (Inal et al., 2012; Nastri et al., 2013; Yeung et al., 2014; Tk et al., 2017). None of these trials performed an intracervical or intrauterine procedure in the control group; thus no additional subgroup analysis was performed.

In the group with the highest degree of failed IVF/ICSI cycles, Group 2, the statistical heterogeneity was borderline (I^2 51%), but based on visual assessment of the forest plot we considered the heterogeneity to be low enough to pool the results. No significant difference in LBR was found between the scratch and control group (RR 1.15 [95%CI 0.52–2.55], total $n = 404$ participants) (Fig. 2c). Two of the six included trials performed an intracervical sham procedure or hysteroscopy in the month of endometrial injury (Baum et al., 2012; Shohayeb and El-Khayat, 2012), so that subgroup analysis was performed on four studies (Inal et al., 2012; Yeung et al., 2014; Singh et al., 2015; Tk et al., 2017). Pooling of the results showed no significant difference between the endometrial injury and control groups (RR 1.28 [95%CI 0.67–2.45], I^2 48%) (Fig. 2c).

Of the two studies that could not be included for meta-analysis because their populations could not be categorized into Group 1 or 2, one did not find a significant difference (scratch 47.2% vs. control 38.1%, p 0.08) (Gibreel et al., 2015), and one did find a significantly higher LBR in the intervention group (22.4% vs. 9.8%, p 0.04)

Table III Study characteristics of each RCT included in the review.

Study	Baum 2012	Gibreel 2015	Inal 2012	Karimzadeh 2009	Liu 2017	Maged 2017	Mahran 2016
City, Country	Tel-Aviv, Israel	Mansoura, Egypt	Konya, Turkey	Yazd, Iran	Beijing, China	Giza, Egypt	Minia, Egypt
Patients	Intervention: $n = 18$, Control: $n = 18$	Intervention: $n = 193$, Control: $n = 194$	Intervention: $n = 50$, Control: $n = 50$	Intervention: $n = 58$, Control: $n = 57$	Intervention: $n = 70$, Control: $n = 72$	Intrvention: $n = 150$, Control: $n = 150$	Intervention: $n = 209$, Control: $n = 209$
Intervention/ Comparison	Intervention: Pipelle on CD 9–12 and 21–24 of the cycle preceding the IVF treatment cycle. Control: Cervical introduction Pipelle.	Intervention: Pipelle on CD21 of the cycle preceding the IVF treatment cycle + 2–3 days thereafter. Control: Sound through cervix up to internal os. Note: HY in case of difficulty passing the cervical os.	Intervention: Pipelle twice with 1-week interval during the luteal phase of the GnRH-analog down regulation cycle prior to the ICSI-ET. Antibiotics after procedure. Control: Usual care.	Intervention: Pipelle on CD21–26 of spontaneous menstrual cycle preceding the ART treatment cycle, when GnRh agonist use began. Control: Usual care.	Intervention: Pipelle on CD10-12 or Pipelle on LH + 7–9 days. Control: sham procedure: Pipelle inserted through the cervix without inducing endometrial injury.	Intervention: Pipelle once in midluteal phase of the cycle preceding the ICSI treatment cycle. Control: Usual care.	Intervention: Pipelle once between CD21–24 of the cycle preceding IVF stimulation cycle Control: Usual care All patients received HY at CD2-5 in the cycle preceding ICSI treatment.
In- and exclusion criteria	Inclusion: Age: 18–41. ≥3 failed IVF cycles. Good ovarian response in previous cycles. Exclusion: Uterine malformation. Endometrioma. Hydrosalpinx at US.	Inclusion: Age: <40. ≥1 failed IVF cycle. Exclusion: Poor ovarian response in previous cycles. Endocrinopathy. History of tubal disconnection for hydrosalpinx. History of endometrial curettage <3 months prior to the study. Fibroids, polyps or adhesions.	Inclusion: ≥1 failed IVF/ICSI cycle. Good ovarian response in previous cycles. Exclusion: Thrombophilia. Hydrosalpinx, or submucous myoma at US. Factors found to have a negative impact on implantation.	Inclusion: Age: 20–40. 2–6 failed ART cycles with the transfer of at least 10 high-grade embryos. No history of blood diseases. Exclusion: Poor responders. Uterine malformation. Endometrioma. Hydrosalpinx at US.	Inclusion: Age < 41 years. No history of prior IVF/ICSI treatment. Normal uterine cavity confirmed with SIS US, basal FSH < 12IU/l. Exclusion: hydrosalpinx, endometriosis, endometrial polyp/ fibroid.	Inclusion: Age < 40 years. No history of prior IVF/ICSI treatment. Basal FSH < 10IU/l. >2 basal follicles on US. Normal uterine cavity (confirmed by HY or HSG). Exclusion: endocrine abnormalities, ovarian cysts, hydrosalpinx, endometrial polyps, partner with azoospermia, pre-implantation genetic diagnosis.	Inclusion: Age 20–40 years. No history of prior IVF/ICSI treatment. History of ≥2 good quality embryo's transferred. Normal uterine cavity (confirmed with HY during workup). Exclusion: endometrial fibroid or polyp, Asherman's syndrome, congenital uterine malformations.
Patient selection	Not described	Not described	Not described	Not described	Not described	Not described	Consecutive
Randomization	Table of random numbers	Computer-generated tables of random numbers	Computer-generated random numbers	Drawing a piece of paper from a bag containing equal number of printed paper for each method	Table of random numbers	Automated web-based randomization	Not described
Allocation concealment	Not described	Sealed envelopes	Not described	Not described	Not described	Not described	Sealed envelopes
Blinding	Single	Single	Not described	Not described	Not described	None	Not described
Loss to follow-up	$n = 4$ (no ovarian stimulation/no ET)	$n = 5$ (no ovarian stimulation)	None	$n = 22$ (discontinued IVF/ICSI treatment due to poor response/no ET)	None	None	None
Outcome measures[a]	Implantation, clinical pregnancy (gestational sac with embryonic pool), live birth.	Clinical pregnancy (positive cardiac activity at GA 8 weeks), live birth, miscarriage (after confirmation of a clinical pregnancy), multiple pregnancy (no definition).	Implantation, clinical pregnancy (positive cardiac activity), ongoing pregnancy (GA > 12 weeks), **live birth.**	Chemical pregnancy, clinical pregnancy (positive cardiac activity).	Implantation, biochemical pregnancy, clinical pregnancy (intrauterine gestational sac on US at 6 weeks' gestation),	Clinical pregnancy (gestational sac and positive heartbeat on US at 6 weeks' gestation), multiple pregnancy (> 1 fetuses on US at 6 weeks' gestation), miscarriage	**Live birth**, implantation, clinical pregnancy (positive heartbeat on US), miscarriage (< 24 weeks' of gestation),

Continued

Table III *Continued*

Study	Baum 2012	Gibreel 2015	Inal 2012	Karimzadeh 2009	Liu 2017	Maged 2017	Mahran 2016
					multiple pregnancy, miscarriage, **live birth.**	(spontaneous miscarriage < 12 weeks' gestation).	multiple gestation, pain, bleeding.

Study	Narvekar 2010	Nastri 2013	Safdarian 2011	Shohayeb 2012	Singh 2015	Tk 2017	Yeung 2014
City, Country	Bangalore, India	Sao Paolo, Brazil	Tehran, Iran	Cairo, Egypt	New Delhi, India	Vellore, India	Hong Kong, Hong Kong
Patients	Intervention: $n = 49$, Control: $n = 51$	Intervention: $n = 79$, Control: $n = 79$	Intervention: $n = 50$, Control: $n = 50$	Intervention: $n = 105$, Control: $n = 105$	Intervention: $n = 30$, Control: $n = 30$	Intervention: $n = 55$, Control: $n = 56$	Intervention: $n = 150$[a], subgroup $n = 45$[b], Control: $n = 150$[a], subgroup $n = 46$[b]
Intervention/ Comparison	Intervention: HY + Pipelle on CD 7–10 and CD 24–25 of the cycle preceding the IVF/ICSI treatment cycle. Control: HY on CD 7–10. Antibiotics peri-procedure.	Intervention: Pipelle once 7–14 days prior to ovarian stimulation. Control: sham procedure: drying cervix with a gauze.	Intervention: Pipelle on CD21 of the cycle preceding the IVF/ICSI treatment cycle. Control: Usual care.	Intervention: HY + Novak on CD 4–7 of the cycle preceding the ICSI treatment cycle. Control: HY.	Intervention: Karman's cannula on CD 14–21 of the cycle preceding the IVF/ICSI treatment cycle. Antibiotics after procedure. Control: Usual care. Antibiotics in both groups.	Intervention: Pipelle 2x within 48 h during luteal phase of the cycle preceding the IVF treatment cycle. Control: usual care.	Intervention: Pipelle 7 days after the LH surge in ovulatory women/CD21 in anovulatory women of the cycle preceding the IVF treatment cycle. Control: Usual care.
In- and exclusion criteria	Inclusion: Age: ≤37. ≥1 failed IVF/ICSI cycle. Good ovarian response in previous cycles. Exclusion: History of endometrial TB. Intramural fibroid distorting the endometrial cavity, submucous myoma, Asherman. Hydrosalpinx at US.	Inclusion: Age < 38 years. Planning IVF/ICSI treatment (unselected group).	Inclusion: Age 20–39 years. Good responders to hormone stimulation. Exclusion: FSH > 11IU/l, endometriosis, hypothalamic amenorrhea, azoospermic male.	Inclusion: Age: <39. ≥2 failed IVF/ICSI cycles. Endometrium <5 mm on CD 4. Exclusion: Submucous myoma distorting the endometrial cavity, endometrial polyp, intrauterine synechia, septate or bicornuate uterus at HSG or US.	Inclusion: Age: <35. ≥2 failed IVF cycles. Good ovarian reserve. No uterus manipulation in last 3 months. Exclusion: Grade III&IV endometriosis, abnormal uterine cavity, prior adhesiolysis, endocrinopathy.	Inclusion: Age: <39 years. ≥1 failed IVF cycle with ≥2 good quality embryos and BMI <30 kg/m^2. Exclusion: previous poor response, endometrial pathology, uterine malformations, severe endometriosis, gross adenomyosis, systemic disease.	Inclusion: Normal uterine cavity at SIS/HY. Exclusion: Endometrial polyp or fibroid distorting the endometrial cavity. Hydrosalpinx. IVF treatment carried out for pre-implantation genetic diagnosis. The use of donor oocytes.
Patient selection	Not described	Consecutive	Not described	Not described	Not described	Consecutive	Consecutive
Randomization	Computer-generated random numbers	Computer-generated random sequence	Computer randomization	Table of random numbers	Random allocation software	Computer-generated sequence	Computer-generated randomization list with blocks of 10
Allocation concealment	Sealed envelopes	Sealed envelopes	Not described	Sealed envelopes	Not described	Sealed envelopes	Sealed envelopes
Blinding	No	Double	Not described	Not described	No	None	No
Loss to follow-up	None	None	None	$n = 10$ (discontinued IVF/ICSI treatment due	None	None	None

Continued

Table III *Continued*

Study	Narvekar 2010	Nastri 2013	Safdarian 2011	Shohayeb 2012	Singh 2015	Tk 2017	Yeung 2014
				to poor response/no ET)			
Outcome measures[c]	Implantation, clinical pregnancy (positive cardiac activity), **live birth**, miscarriage (after confirmation of a clinical pregnancy), multiple pregnancy (no definition).	Clinical pregnancy, **live birth** (per woman), multiple pregnancy (>1 fetus with positive heartbeat), miscarriage (spontaneous miscarriage of clinical pregnancy <20 weeks' gestation), pain.	Not pre-specified. Reported: implantation, chemical pregnancy, clinical pregnancy.	Implantation, clinical pregnancy (positive cardiac activity), **live birth**, miscarriage (after confirmation of a clinical pregnancy).	Implantation, ongoing pregnancy (GA > 24 weeks), live birth, miscarriage (no definition).	Clinical pregnancy (gestational sac on US), **live birth**, implantation, multiple pregnancy (>1 gestational sac on US), miscarriage (miscarriage <24 weeks' gestation), preterm delivery.	Implantation, clinical pregnancy (gestational sac at GA 6 weeks), **ongoing pregnancy** (GA > 20 weeks), live birth, miscarriage (GA < 20 weeks), multiple pregnancy (>1 gestational sacs GA 6 weeks).

n, number of participants; CD, cycle day; ET, embryo transfer; GA, gestational age; HY, hysteroscopy; SIS, saline infusion sonography; TB, tuberculosis; US, ultrasound.
[a]In bold the primary outcome measure if described.

(Narvekar *et al.*, 2010). However, when we analysed the results published by Narvekar *et al.*, our calculations showed a non-significant RR of 2.29 [95%CI 0.86–6.11].

CPR

Similar to the results of LBR, the overall results could only be pooled for Group 1 (I^2 77% for group 0, I^2 28% for group 1 and I^2 59% for Group 2).

As is shown in Fig. 3a, five RCTs were included in Group 0 comprising a total of 1087 participants. Three out of five studies showed no significant difference between the endometrial scratching and the control group (RR 0.94; RR 1.09; RR 0.99, respectively) (Nastri *et al.*, 2013; Yeung *et al.*, 2014; Liu *et al.*, 2017). The two largest studies, including a total of 718 participants, reported a significantly higher CPR in the intervention group (RR 2.07 and 1.49, respectively) (Mahran *et al.*, 2016; Maged *et al.*, 2018). Three trials were eligible for subgroup analysis excluding trials with a high risk of unintentional endometrial injury (Nastri *et al.*, 2013; Yeung *et al.*, 2014; Maged *et al.*, 2018). Based on 527 participants, the pooled results of this Group 0 subgroup showed a significant benefit of endometrial injury on CPR (RR 1.28 [95%CI 1.02–1.62], I^2 0%) (Fig. 3a).

The pooled results of Group 1 are shown in Fig. 3b. No significant difference in CPR was found between the endometrial scratching and the control groups (RR 1.04 [95%CI 0.74–1.45]), and as none of the four RCTs had a risk of unintentional endometrial injury in the control group in the month prior to ovarian stimulation, subgroup analysis was not necessary.

Six RCTs in Group 2 reported CPR, evaluating a total of 459 patients. As shown in Fig. 3c, the results of the RCTs varied from non-significant RRs ranging from 0.09 to 1.46 (Baum *et al.*, 2012; Yeung *et al.*, 2014; Tk *et al.*, 2017) to significant RRs ranging from 1.78 to 3.19 (Karimzadeh *et al.*, 2009; Inal *et al.*, 2012; Shohayeb and El-Khayat., 2012). Subgroup analysis with pooling of results was performed on four studies (I^2 30% and n = 223 participants) (Karimzadeh *et al.*, 2009; Inal *et al.*, 2012; Yeung *et al.*, 2014) showing a statistically significant benefit of endometrial injury with a RR of 2.03 [95%CI 1.20–3.43] (Fig. 3c).

Of the three studies that could not be included in pooled analysis because their populations could not be categorized into Group 1 or 2, only Narvekar found a significant benefit on CPR from endometrial scratching (32.7% vs. 13.7%, p 0.01) (Narvekar *et al.*, 2010; Safdarian *et al.*, 2011; Gibreel *et al.*, 2015).

Miscarriage rate and multiple pregnancy rate

A total of nine RCTs reported the miscarriage rate and eight RCTs the multiple pregnancy rate; analysis was performed on the overall groups (i.e. Groups 0, 1 and 2 combined). These rates were calculated relative to the number of clinical pregnancies. Both the miscarriage and the multiple pregnancy rate did not differ between the intervention and control groups, (RR 0.82 [95%CI 0.57–1.17] total n = 825 participants and RR 1.06 [95%CI 0.84–1.35] total n = 775 participants, respectively, Figs 4 and 5)(Narvekar *et al.*, 2010; Shohayeb and El-Khayat., 2012; Nastri *et al.*, 2013; Yeung *et al.*, 2014; Gibreel *et al.*, 2015; Mahran *et al.*, 2016; Liu *et al.*, 2017; Tk *et al.*, 2017; Maged *et al.*, 2018). Subgroup analysis excluding studies with a risk of endometrial injury in the control group for the outcomes miscarriage and multiple pregnancy rate did not lead to different results: RR 1.13 [95%CI

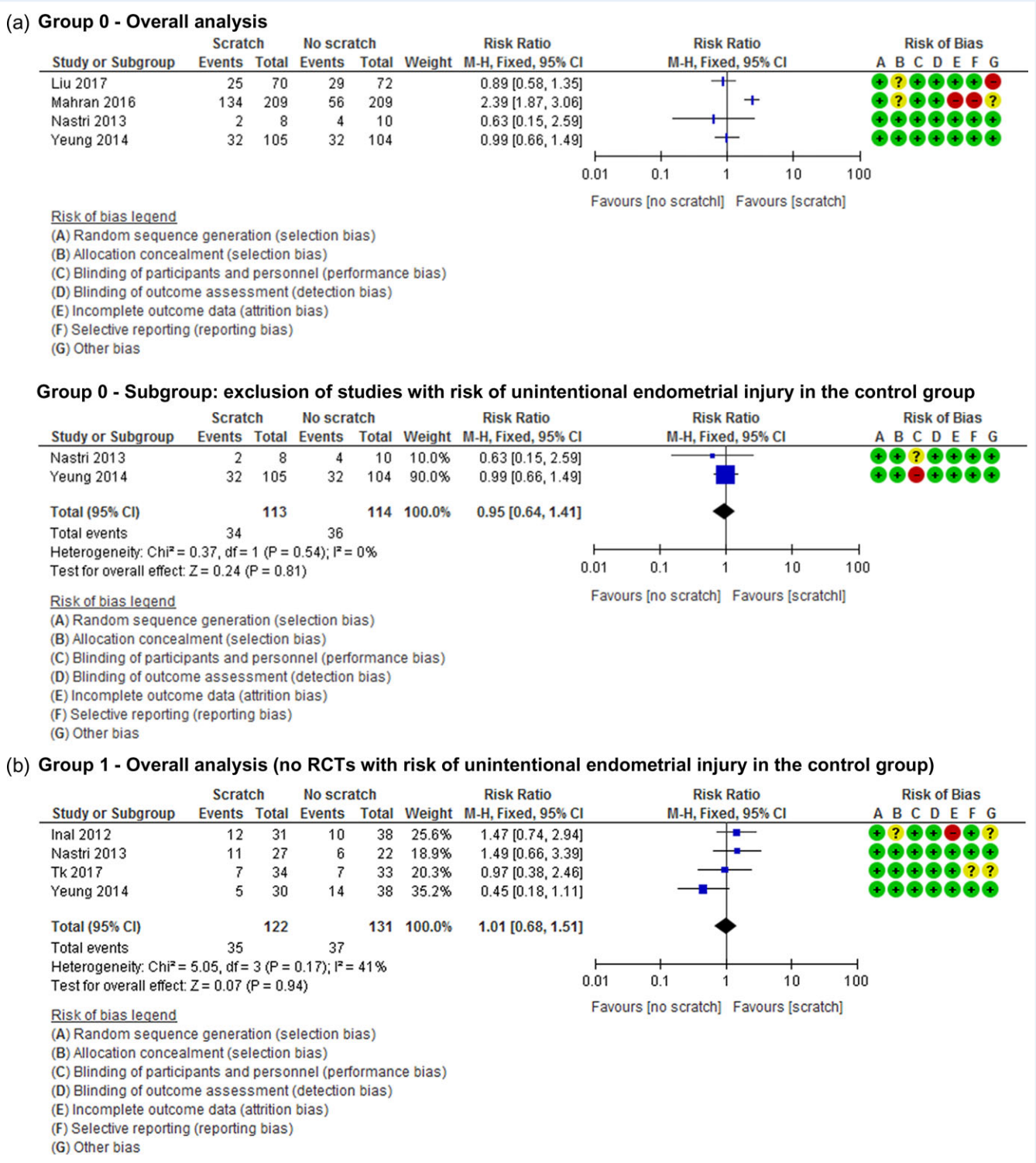


Figure 2 Live birth rate after the IVF/ICSI cycle following randomization. a Live birth rate for participants with no prior IVF/ICSI treatment (Group 0), and subgroup analysis. Events, live birth; M–H, Mantel–Haenszel. **b** Live birth rate for participants with one failed complete IVF/ICSI cycle (Group 1). Events, live birth; M–H, Mantel–Haenszel. None of the trials had a risk of unintentional endometrial injury in the control group. Thus, subgroup analysis was not performed. **c** Live birth rate for participants with two failed complete IVF/ICSI cycles (Group 2), and subgroup analysis. Events, live birth; M–H, Mantel–Haenszel.

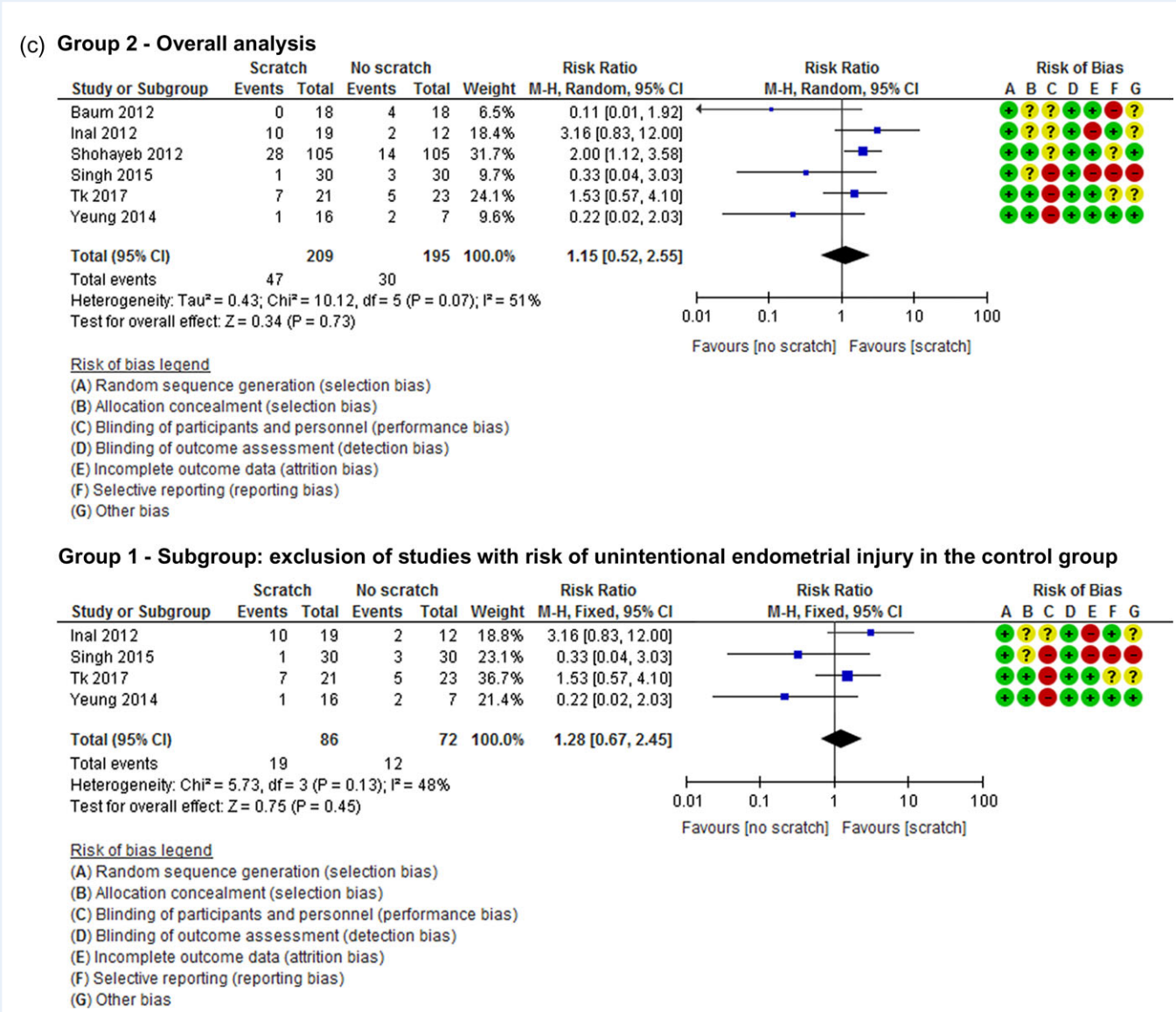


Figure 2 *Continued*

0.73–1.75] and RR 1.03 [95%CI 0.77–1.38], respectively (Nastri *et al.*, 2013; Yeung *et al.*, 2014; Tk *et al.*, 2017; Maged *et al.*, 2018).

Discussion

Summary of the results

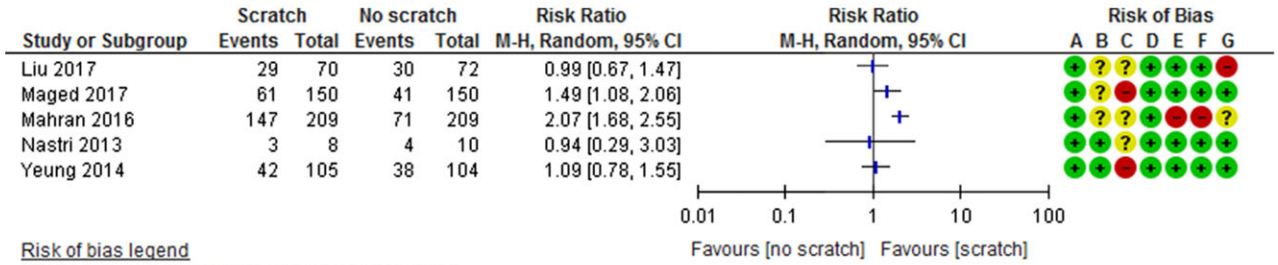
This review evaluates the effect of endometrial scratching in patients naïve to IVF/ICSI treatment and with varying numbers of previously failed IVF/ICSI cycles. Pooled analysis for the outcomes LBR and CPR in women with one previously failed cycle (Group 1) showed no significant difference between the intervention and control groups. The results for women with 0 (Group 0) or two or more failed cycles (Group 2) could not be pooled, and while the RRs varied widely between studies, most RCTs reported no significant difference for LBR and CPR. For CPR in women with two or more failed cycles, the results were more contradictory, with half of the RCTs showing a significantly improved CPR after scratching, and the other half showing no significant difference.

Pooled subgroup analyses including RCTs with no risk of unintentional endometrial injury in the control group showed no statistically significant difference in LBR for women with 0 or 2 or more previously failed cycles, but showed an increased CPR in the intervention group of these same women.

Related literature

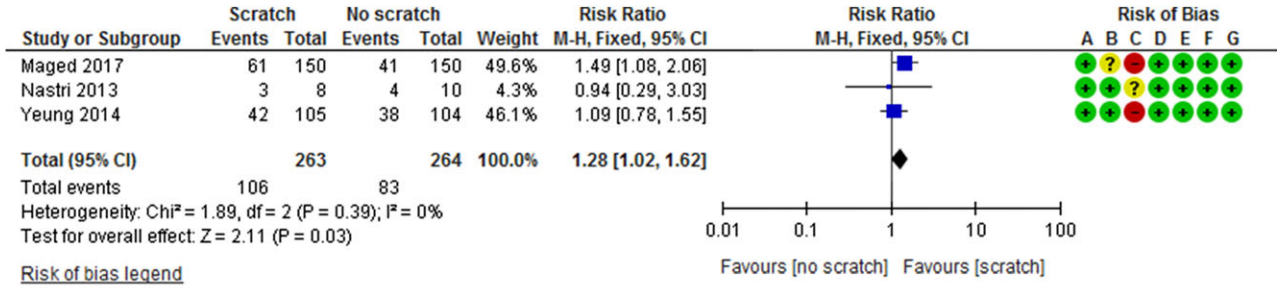
Several reviews on endometrial scratching have been published (El-Toukhy *et al.*, 2012; Potdar *et al.*, 2012; Nastri *et al.*, 2015; Panagiotopoulou *et al.*, 2015; Ko and Ng, 2016; Santamaria *et al.*,

(a) **Group 0 - Overall analysis**



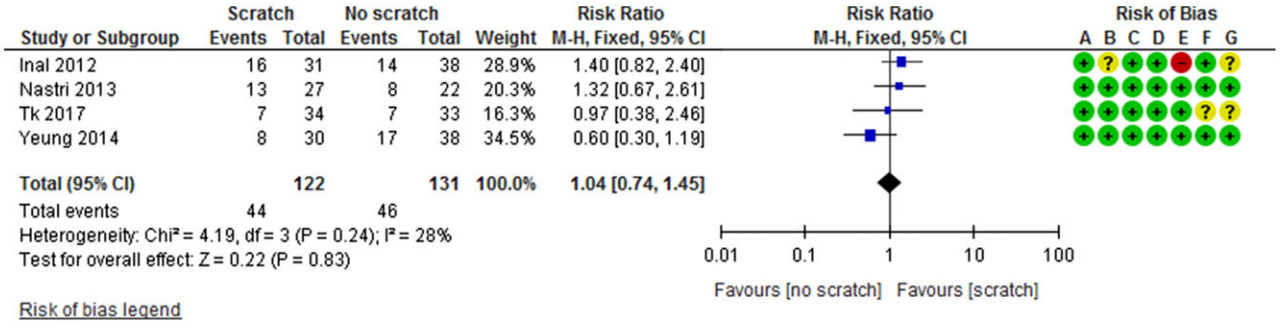
Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Group 0 - Subgroup: exclusion of studies with risk of unintentional endometrial injury in the control group



Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

(b) **Group 1 - Overall analysis (no RCTs with risk of unintentional endometrial injury in the control group)**



Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Figure 3 Clinical pregnancy rate after the IVF/ICSI cycle following randomization. a Clinical pregnancy rate for participants with no prior IVF/ICSI treatment (Group 0), and subgroup analysis. Events, clinical pregnancy; M–H, Mantel–Haenszel. **b** Clinical pregnancy rate for participants with one failed complete IVF/ICSI cycle (Group 1). Events, clinical pregnancy; M–H, Mantel–Haenszel. None of the trials had a risk of unintentional endometrial injury in the control group. Thus, subgroup analysis was not performed. **c** Clinical pregnancy rate for participants with two failed complete IVF/ICSI cycles (Group 2), and subgroup analysis. Events, clinical pregnancy; M–H, Mantel–Haenszel.

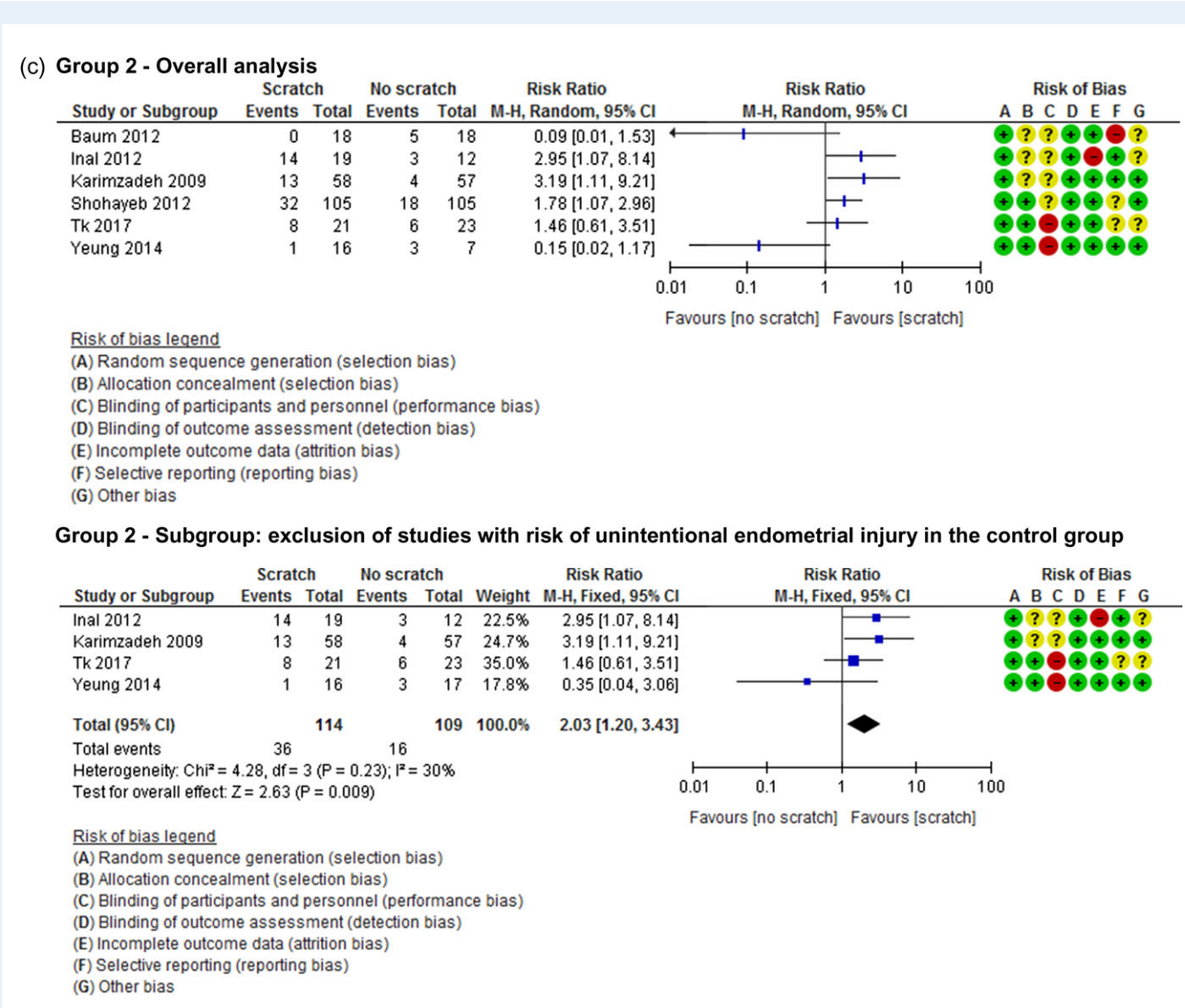


Figure 3 *Continued*

2016; Zygula *et al.*, 2016), of which four did not and three did perform a meta-analysis (El-Toukhy *et al.*, 2012; Potdar *et al.*, 2012; Nastri *et al.*, 2015). All reviews used different inclusion criteria (scratching prior to or during ovarian stimulation, number of previously failed IVF/ICSI cycles, only RCTs versus also non-randomized trials), so that none of the reviews included the same publications. Potdar *et al.* and El-Toukhy *et al.* had their reviews published in 2012. At that time, few RCTs had been published so that both also included non-randomized trials. Potdar *et al.* showed a significantly increased CPR (RR 2.32 [95% CI 1.72–3.13]) and LBR (RR 2.11 [95%CI 1.37–3.25]) after endometrial scratching, but this was based on four studies for CPR (only two were RCTs) and two studies for LBR (one RCT). This led this group to conclude that despite increased pregnancy rates, there are many unanswered questions on the scratching method, stressing the need for large multicenter RCTs (Potdar *et al.*, 2012). Importantly, Potdar and colleagues limited their review to include only patients with RIF, but did not provide the reader with a definition of RIF.

Similarly, based on one RCT and four non-randomized trials, El-Toukhy *et al.* (2012) reported a combined ongoing pregnancy/live birth rate (OPR/LBR) outcome, which was increased after scratching (RR 2.28 [95%CI 1.65–3.14]). This review used different inclusion criteria, which meant it was based on a different set of trials than the review by Potdar and colleagues. El Toukhy *et al.* 2012 state that endometrial scratching may improve the chance of pregnancy, but they also stress the need for robust large randomized trials because of multiple methodological concerns of the included studies.

In 2015, a Cochrane review by Nastri *et al.* 2015 showed no significant difference for women with one or no failed embryo transfers, but a significantly higher chance of clinical pregnancy and live birth after scratching in women with two or more failed embryo transfers (RR 1.63 [95%CI 1.12–2.38] I^2 49% and RR 1.96 [95%CI 1.21–3.16] I^2 37%), but they also state that uncertainties regarding timing, frequency and degree of scratching, and population exist. Our results differ from Nastri's findings; this can be explained by slightly different inclusion

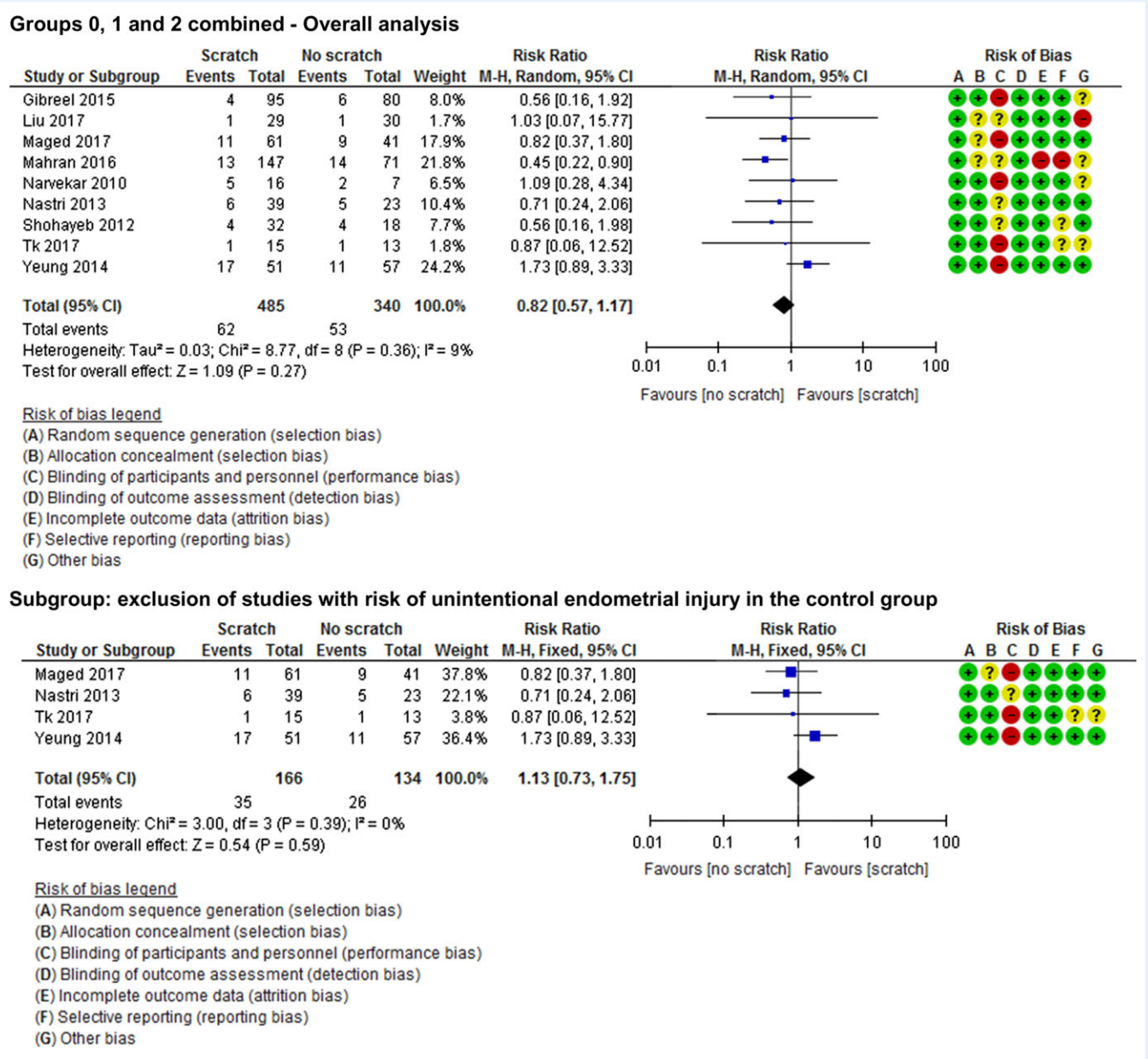


Figure 4 Miscarriage rate for the overall population (Groups 0, 1 and 2 combined), and subgroup analysis. Events, miscarriage; M–H, Mantel–Haenszel.

criteria and different grouping of participants. Nastri included trials that performed endometrial injury during the cycle preceding ovarian stimulation as well as during the ovarian stimulation cycle, while the current review did not include the latter. Also, participants were grouped according to no or one failed embryo transfer versus two or more failed transfers in Nastri's review, while in the current review the grouping is based on failed complete IVF cycles. We have chosen a grouping based on complete failed IVF/ICSI cycles based on the hypothesis that in each oocyte retrieval, only a small number of oocytes is capable of developing into a pregnancy that results in live birth (Patrizio and Sakkas., 2009; Doherty *et al.*, 2014). Thus, a couple with eight prior transfers from one oocyte retrieval will have had less 'true' chances of pregnancy than a couple with eight prior transfers from three oocyte retrievals, and couples will be better comparable in terms of 'exposure to chances of pregnancy' when grouped according to failed complete IVF/ICSI cycles.

Other differences exist in the strategy to cope with statistical heterogeneity: Nastri decided to perform meta-analysis and explore the heterogeneity by performing subgroup analyses. Our strategy was not to perform meta-analysis when high statistical heterogeneity was detected but to perform subgroup analyses including only studies with a low risk of unintentional endometrial injury. Of the reviews not performing a meta-analysis, Panagiotopoulou *et al.* 2015 refrained from doing this because of clinical heterogeneity. They included four RCTs with a total of 416 participants with a history of two or more failed IVF/ICSI cycles, and concluded that the effect of endometrial

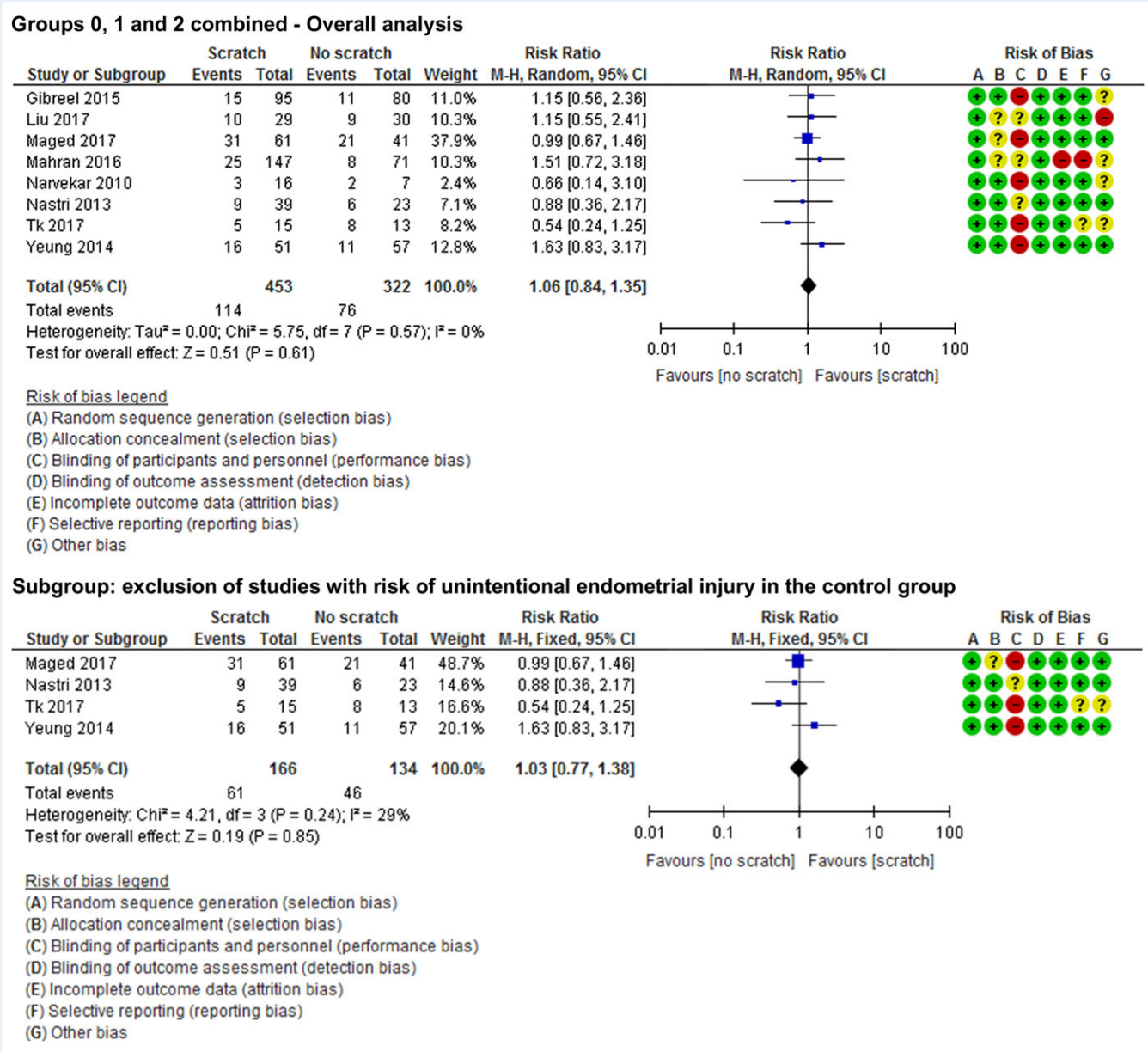


Figure 5 Multiple pregnancy rate for the overall population (Groups 0, 1 and 2 combined), and subgroup analysis. Events, multiple pregnancy; M–H, Mantel–Haenszel.

scratching on pregnancy rates was inconsistent, and that insufficient data was available to advise endometrial scratching in women with RIF undergoing ART (Panagiotopoulou *et al.*, 2015). The reviews by Santamaria *et al.* 2016, Zygula *et al.* 2016 and Ko and Ng, 2016 have a more narrative style, and do not report a clear search or inclusion/exclusion criteria for the RCTs. Santamaria *et al.* 2016 reviewed 16 RCTs with a total of 2570 participants who were either trying to conceive naturally, by IUI or through IVF/ICSI treatment, and specifically focused on the studies published in 2014/2015 in order to explain the different conclusions between the reviews of Nastri *et al.* 2015 and Panagiotopoulou *et al.* 2015. They concluded that the data of methodologically sound RCTs did not demonstrate a beneficial effect, and that a meta-analysis such as in the Cochrane review (2015) leads to incorrect or incomplete conclusions owing to incomplete and insufficient data. Since the underlying biological mechanism is also unknown, they advise that endometrial scratching should not be implemented in daily practice. The reviews by Zygula *et al.* 2016 and Ko and Ng, 2016 included RCTs and non-randomized trials on an unselected group of participants. Zygula concluded that evidence of a positive effect of scratching was lacking, and Ko concluded that no effect was expected in the unselected population, but that scratching could improve pregnancy rates in women with two or more failed IVF/ICSI cycles, but that the evidence was of moderate quality only so that caution should be taken in adhering to this conclusion. As both reviews lack clear inclusion/exclusion criteria, it is difficult to conclude that the reviews were incomplete, but it seems that both are missing three or four

RCTs that had already been published at the time. Moreover, the conclusion of Ko and Ng, 2016 that endometrial injury could improve pregnancy rates in women with two or more failed IVF/ICSI cycles does not seem to be supported by their summarizing table in which three RCTs show a benefit, two RCTs do not show a benefit, and two retrospective studies also do not show a benefit of endometrial scratching.

In summary, several reviews have been published in the past 4 years of which three performed a meta-analysis despite statistical and/or clinical heterogeneity. All reviews either conclude that the data is too diverse to draw a conclusion or that endometrial scratching seems to have a positive effect on CPR and LBR in patients with two or more failed IVF/ICSI cycles but that the data is heterogeneous and studies are of moderate quality at best.

Interpretation of the results

In line with previous reviews, our findings emphasize the statistical heterogeneity between published RCTs. The design of this review—evaluating the effect of endometrial injury according to the number of previously failed IVF/ICSI cycles—ensured that clinical heterogeneity was reduced in comparison to reviews with broader inclusion criteria. Nevertheless, statistical heterogeneity was often too high to perform meta-analysis. This indicates that the results should be interpreted with caution and no firm conclusions can be drawn.

The statistical heterogeneity could, in part, be caused by the variation in study design; for example, the sample size varied widely across studies. Only three RCTs included ≥300 participants (maximally 420 and with varying degrees of IVF/ICSI failure) (Yeung et al., 2014; Gibreel et al., 2015; Mahran et al., 2016), and many of our analyses on the different degrees of IVF/ICSI failure are based on RCTs of <100 or even <50 participants.

Furthermore, despite our attempt to reduce clinical heterogeneity it is still likely to have contributed to the statistical heterogeneity as the method, timing and frequency of the intervention and baseline characteristics (e.g. BMI, age) varied among RCTs. Even in our grouped populations, clinical heterogeneity may still have existed due to varying causes of subfertility—and thus variation in chances of natural conception before the start of ART—and variation in the total number of embryos that had been transferred previously. Clinical heterogeneity also existed in the definitions of the outcome variables. For example, a clinical pregnancy could be defined as a pregnancy on ultrasound or as a gestational sac with positive fetal heartbeat on ultrasound at 6 weeks of gestation. Variation was also present for all the other definitions.

Another problem in interpreting the results is the high or unclear risk of bias that was present in almost all studies. The most frequent causes of high risk of bias were outcome measures that were not pre-specified, pre-specified outcomes that were not reported, discrepancies in the data, and outcomes that were not reported on an intention-to-treat basis. The RCTs by Mahran et al. 2016 and Singh et al. 2015 in particular, and to a lesser extent by Baum et al. 2012 and Inal et al. 2012 were suspected of carrying a high risk of bias. Most of these studies reported a strong and significant effect of scratching, thereby having a large impact on the overall outcomes. Since the study by Mahran et al. 2016 also included the highest number of participants, this study had a major effect on the outcomes in Group 0 by being the only RCT that shows a very strong and significant effect of scratching. Because of the low number of RCTs in each analysis and because many RCTs showed some risk of bias, we refrained from performing a sensitivity analysis including only the methodologically sound studies.

Not all studies reported all outcome measures. Especially with results that vary widely between the different RCTs, this can impact the overall outcome of the analysis. For example, Karimzadeh et al. 2009 and Maged et al. 2018 reported a significant positive effect of scratching but only evaluated CPR and not LBR (Groups 2 and 0, respectively). In contrast, Singh et al. 2015 only reported LBR and not CPR, and they did not find a difference (Group 2). This can explain why some comparisons did report a significant benefit for CPR, but not for LBR.

Strengths and limitations

Strengths of this review are that a comprehensive systematic search was performed and that inclusion was based on well-defined inclusion and exclusion criteria with more similar scratching methods (i.e. only prior to the stimulation cycle). Also, authors were contacted and additional information was collected to perform the best analysis possible. Furthermore, many additional RCTs have been published since the previous reviews. Thus, despite the fact that this review has narrower inclusion criteria regarding the timing of endometrial injury, four additional RCTs with relatively large sample sizes (100–418) have been included in comparison to the Cochrane review by Nastri et al. 2015. Also, this review is the first to group the population into three degrees of previously failed cycles, thereby providing insight into the frequently posted hypothesis that endometrial injury may only benefit patients with a high degree of implantation failure. Strengths of the chosen methods include that no meta-analysis was performed in the case of substantial statistical heterogeneity ($I^2 \geq 50\%$). High statistical heterogeneity can be dealt with in various ways: it can be ignored and a fixed-effect analysis can be performed; if the heterogeneity is unexplained a random-effect model can be used for meta-analysis; a different effect measure can be chosen; studies can be excluded; or one can choose not to perform a meta-analysis (Borenstein et al., 2010; Higgins and Green, 2011). Based on the clinical heterogeneity that became apparent from the trial methods, and based on visual assessment of the forest plots, both ignoring the heterogeneity and performing a random-effect meta-analysis introduced a high risk of misleading results. Since the effect measures were not continuous (LBR, CPR, etc.), and these effect measures were considered relevant outcomes for patients and physicians, changing the effect measure was also not an option. Thus, we chose to perform a subgroup analysis excluding certain trials in order to reduce clinical (and hopefully statistical) heterogeneity, but to also show the overall forest plots without the pooled results in order to still give an overview of all available evidence.

A limitation is that not all RCTs could be included in the analyses because attempts to obtain more information from the authors had failed or because RCTs used different subgroup-criteria so that the results could not be translated to one of our populations. Another limitation is that meta-analysis could not be performed for all outcomes/groups. However, this is inherent to the variation in studies that have been performed and the alternative—performing meta-analysis—would in our opinion have been less desirable. A third limitation is that clinical heterogeneity was still present in the populations analysed in this review, because of varying duration of subfertility with

varying chances of natural conception, and because a failed complete IVF/ICSI cycle can consist of one or multiple failed embryo transfers. This is a problem that applies to all studies in the field of IVF/ICSI and it is difficult to tackle. For this review, the current grouping was the best possible option, but ideally an individual patient data analysis (IPD) will pool the data of women with the same number of previously failed transfers and/or similar duration and cause of subfertility. Also, in order to reduce clinical heterogeneity, we chose to only include trials that performed endometrial injury prior to the cycle in which ovarian stimulation is started. We regard this as a strength because it enables better comparison of the trials, but one could also argue that this is a limitation because some trials were excluded.

Clinical implications and future perspective

Despite inconclusive and sometimes contradicting results as demonstrated in this and other reviews, endometrial scratching is already being advised to women undergoing IVF or ICSI by 83% of clinicians in some countries (Lensen *et al.*, 2016). While endometrial scratching may seem simple and harmless, one should question whether such a procedure should be implemented as standard treatment when the evidence is so weak. Should we not protect patients from treatment that is potentially harmful or at best gives false hope?

The relatively small sample sizes, high risk of bias, methodological flaws and statistical heterogeneity emphasize the need for large, methodologically sound RCTs in more homogeneous populations undergoing ART. Currently, two large trials are being conducted: an RCT conducted by Northwestern University (600 participants) evaluating endometrial scratching in the unselected IVF population (Marsh., 2016), and an RCT on women undergoing their first IVF/ICSI treatment conducted by Metwally *et al.* of the University of Sheffield (1044 participants) (Chatters, 2017). Two large trials have recently completed the inclusion of participants: the Pipelle for Pregnancy (PIP) trial (1300 participants, unselected IVF/ICSI population) and the SCRaTCH study (900 participants, 1 failed IVF/ICSI cycle) (Lensen *et al.*, 2018; van Hoogenhuijze *et al.*, 2017a). The results of the PIP trial have recently been presented and showed no significant effect of endometrial scratching on LBR (odds ratio 1.00 [95%CI 0.78–1.27]) (Lensen *et al.*, 2018). The PIP trial could not be included in this review because it did not fulfill our inclusion criterion that endometrial injury should be performed before the stimulation cycle.

With more high quality RCTs (becoming) available, the effectiveness of endometrial scratching for different subgroups of women or the identification of the optimal scratching method can be evaluated in an IPD. An IPD focusing on the different subgroups of patients that might benefit from scratching is also underway (van Hoogenhuijze *et al.*, unpublished but registered at PROSPERO, 2017b).

Importantly, different study groups worldwide should use standardized definitions of outcomes and outcome measures so that results can be compared and pooled. The COMMIT initiative is now well-underway to achieve this goal (COMMIT).

Conclusion

The current review applied strict inclusion criteria by confining the timing of endometrial scratching to prior to the start of ovarian stimulation, and by defining strict populations (0 vs. 1 vs. 2 or more failed IVF/ICSI cycles). Nonetheless, important clinical and statistical heterogeneity was still encountered making it impossible to draw firm conclusions. This has led us to conclude that despite the publication of multiple RCTs in the past 10 years, it remains unclear whether we should scratch the endometrium prior to IVF/ICSI treatment. The studies that have been performed thus far have small sample sizes, frequently have methodological flaws, include heterogeneous populations, use different scratching techniques and show statistical heterogeneity. This stresses the need for large, high quality RCTs and a well-performed IPD.

Authors' roles

J.K., N.H., H.T. and F.B. participated in the design of the paper. J.K., J.B. and N.H. performed the literature search and data extraction. J.K. and N.H. analysed the data, H.T., F.B. and J.B. contributed in the interpretation of the results. N.H., J.K. and H.T. wrote the manuscript. F.B. and J.B. revised the manuscript. All authors approved the final version of the manuscript.

Funding

No specific funding was sought for the study. The Department of Reproductive Medicine and Gynaecology funds of the University Medical Center of Utrecht were used to support the authors throughout the study period and preparation of the manuscript.

Conflict of interest

N.E. van Hoogenhuijze and H.L. Torrance have received travel grants by Merck. F.J.M. Broekmans, J.C. Kasius and J. Bosteels have no conflict of interest to declare.

References

Barash A, Dekel N, Fieldust S, Segal I, Schechtman E, Granot I. Local injury to the endometrium doubles the incidence of successful pregnancies in patients undergoing in vitro fertilization. *Fertil Steril* 2003;**6**:1317–1322.

Baum M, Yerushalmi GM, Maman E, Kedem A, MacHtinger R, Hourvitz A, Dor J. Does local injury to the endometrium before IVF cycle really affect treatment outcome? Results of a randomized placebo controlled trial. *Gynecol Endocrinol* 2012;**12**:933–936.

Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Hum Reprod* 2007;**6**:1506–1512.

Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 2010;**2**:97–111.

Chatters R Endometrial Scratch Trial (ISRCTN 23800982). 2017; Available at: www.isrctn.com. Accessed January 7, 2019.

De Geyter C, Calhaz-Jorge C, Kupka MS, Wyns C, Mocanu E, Motrenko T, Scaravelli G, Smeenk J, Vidakovic S, Goossens V *et al.* ART in Europe, 2014: results generated from European registries by ESHRE: the European IVF-monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE). *Hum Reprod* 2018;**9**:1586–1601.

Doherty LF, Martin JR, Kayisli U, Sakkas D, Patrizio P. Fresh transfer outcome predicts the success of a subsequent frozen transfer utilizing blastocysts of the same cohort. *Reprod BioMed Online* 2014;**2**:204–208.

El-Toukhy T, Sunkara S, Khalaf Y. Local endometrial injury and IVF outcome: a systematic review and meta-analysis. *Reprod BioMed Online* 2012;**4**:345–354.

Gibreel A, El-Adawi N, Elgindy E, Al-Inany H, Allakany N, Tournaye H. Endometrial scratching for women with previous IVF failure undergoing IVF treatment. *Gynecol Endocrinol* 2015;**4**:313–316.

Gnainsky Y, Granot I, Aldo P, Barash A, Or Y, Mor G, Dekel N. Biopsy-induced inflammatory conditions improve endometrial receptivity: the mechanism of action. *Reproduction* 2015;**1**:75–85.

Gnainsky Y, Granot I, Aldo PB, Barash A, Or Y, Schechtman E, Mor G, Dekel N. Local injury of the endometrium induces an inflammatory response that promotes successful implantation. *Fertil Steril* 2010;**6**: 2030–2036.

Granot I, Dekel N, Bechor E, Segal I, Fieldust S, Barash A. Temporal analysis of connexin43 protein and gene expression throughout the menstrual cycle in human endometrium. *Fertil Steril* 2000;**2**:381–386.

Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA *et al*. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011a;**343**:d5928.

Higgins JPT, Green S. Chapter 9.5: heterogeneity. In: Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 ed* [updated March 2011]. The Cochrane Collaboration, 2011. Available at: www.handbook.cochrane.org. Accessed January 7, 2019.

Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**7414**:557–560.

Inal ZHO, Görkemli H, Inal HA. The effect of local injury to the endometrium for implantation and pregnancy rates in ICSI -ET cycles with implantation failure: a randomised controlled study. *Eur J Gen Med* 2012;**4**:223–229.

Karimzadeh MA, Ayazi Rozbahani M, Tabibnejad N. Endometrial local injury improves the pregnancy rate among recurrent implantation failure patients undergoing *in vitro* fertilisation/intra cytoplasmic sperm injection: a randomised clinical trial. *Aust N Z J Obstet Gynaecol* 2009;**6**:677–680.

Ko JKY, Ng EHY. Scratching and IVF: any role? *Curr Opin Obstet Gynecol* 2016;**3**:178–183.

Lass A, Peat D, Avery S, Brinsden P. Histological evaluation of endometrium on the day of oocyte retrieval after gonadotrophin-releasing hormone agonist/follicle stimulating hormone ovulation induction for in-vitro fertilization. *Hum Reprod* 1998;**11**:3203–3205.

Lensen S, Osavyluk D, Armstrong S, Napier E, Sadler L, Hennes A, Stadelmann C, Hamoda H, Khalaf Y, Webber L *et al*. Endometrial scratching by pipelle biopsy in IVF (the PIP study): a pragmatic randomised controlled trial. *Hum Reprod* 2018;**33**:i62.

Lensen S, Sadler L, Farquhar C. Endometrial scratching for subfertility: everyone's doing it. *Hum Reprod* 2016;**6**:1241–1244.

Li R, Hao G. Local injury to the endometrium: Its effect on implantation. *Curr Opin Obstet Gynecol* 2009;**3**:236–239.

Liu W, Tal R, Chao H, Liu M, Liu Y. Effect of local endometrial injury in proliferative vs. luteal phase on IVF outcomes in unselected subfertile women undergoing *in vitro* fertilization. *Reprod Biol Endocrinol* 2017;**15**:75.

Maged AM, Rashwan H, AbdelAziz S, Ramadan W, Mostafa WAI, Metwally AA, Katta M. Randomized controlled trial of the effect of endometrial injury on implantation and clinical pregnancy rates during the first ICSI cycle. *Int J Gynecol Obstet* 2018;**2**:211–216.

Mahran A, Ibrahim M, Bahaa H. The effect of endometrial injury on first cycle IVF/ICSI outcome: a randomized controlled trial. *Int J Reprod Biomed* 2016;**3**:193–198.

Mains L, Van Voorhis BJ. Optimizing the technique of embryo transfer. *Fertil Steril* 2010;**3**:785–790.

Marsh E Endometrial injury and in vitro fertilization outcomes (NCT 02153814). 2016; Available at: https://clinicaltrials.gov/ct2/show/NCT02153814?term=NCT02153814&rank=1. Accessed January 7, 2019.

Narvekar S, Gupta N, Shetty N, Kottur A, Srinivas M, Rao K. Does local endometrial injury in the nontransfer cycle improve the IVF-ET outcome in the subsequent cycle in patients with previous unsuccessful IVF A randomized controlled pilot study. *J Hum Reprod Sci* 2010;**1**:15–19.

Nastri CO, Ferriani RA, Raine-Fenning N, Martins WP. Endometrial scratching performed in the non-transfer cycle and outcome of assisted reproduction: a randomized controlled trial. *Ultrasound Obstet Gynecol* 2013;**4**:375–382.

Nastri CO, Lensen SF, Gibreel A, Raine-Fenning N, Ferriani RA, Bhattacharya S, Martins WP. Endometrial injury in women undergoing assisted reproductive techniques. *Cochrane Database Syst Rev* 2015;**3**: CD009517.

Panagiotopoulou N, Karavolos S, Choudhary M. Endometrial injury prior to assisted reproductive techniques for recurrent implantation failure: a systematic literature review. *Eur J Obstet Gynecol Reprod Biol* 2015;**193**: 27–33.

Patrizio P, Sakkas D. From oocyte to baby: a clinical evaluation of the biological efficiency of *in vitro* fertilization. *Fertil Steril* 2009;**4**: 1061–1066.

Potdar N, Gelbaya T, Nardo LG. Endometrial injury to overcome recurrent embryo implantation failure: a systematic review and meta-analysis. *Reprod BioMed Online* 2012;**6**:561–571.

Safdarian L, Movahedi S, Aleyasine A, Aghahosaini M, Fallah P, Rezaiian Z. Local injury to the endometrium does not improve the implantation rate in good responder patients undergoing *in-vitro* fertilization. *Iran J Reprod Med* 2011;**4**:285–288.

Santamaria X, Katzorke N, Simón C. Endometrial 'scratching': what the data show. *Curr Opin Obstet Gynecol* 2016;**4**:242–249.

Shohayeb A, El-Khayat W. Does a single endometrial biopsy regimen (S-EBR) improve ICSI outcome in patients with repeated implantation failure? A randomised controlled trial. *Eur J Obstet Gynecol Reprod Biol* 2012; **2**:176–179.

Singh N, Toshyan V, Kumar S, Vanamail P, Madhu M. Does endometrial injury enhances implantation in recurrent in-vitro fertilization failures? A prospective randomized control study from tertiary care center. *J Hum Reprod Sci* 2015;**4**:218–223.

Tan BK, Vandekerckhove P, Kennedy R, Keay SD. Investigation and current management of recurrent IVF treatment failure in the UK. *BJOG* 2005;**6**:773–780.

Tk A, Singhal H, S Premkumar P, Acharya M, S Kamath M, George K. Local endometrial injury in women with failed IVF undergoing a repeat cycle: a randomized controlled trial. *Eur J Obstet Gynecol Reprod Biol* 2017;**214**: 109–114.

van Hoogenhuijze NE, Torrance HL, Mol F, Laven JSE, Scheenjes E, Traas MAF, Janssen C, Cohlen B, Teklenburg G, de Bruin JP *et al*. Endometrial scratching in women with implantation failure after a first IVF/ICSI cycle; does it lead to a higher live birth rate? The SCRaTCH study: a randomized controlled trial (NTR 5342). *BMC Women's Health* 2017a;**17**:47.

van Hoogenhuijze N, Torrance H, Broekmans F, Lensen S, Farquhar C, Raine-Fenning N, Wang R, Mol B, Martins W, Ng E et al. Study protocol: endometrial scratching in women trying to conceive from in vitro fertilization a collaborative individual participant data meta-analysis. PROSPERO 2017 CRD42017079120. 2017b. Available at: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017079120. Accessed January 7, 2019.

Yeung TWY, Chai J, Li RHW, Lee VCY, Ho PC, Ng EHY. The effect of endometrial injury on ongoing pregnancy rate in unselected subfertile women undergoing *in vitro* fertilization: a randomized controlled trial. *Hum Reprod* 2014;**11**:2474–2481.

Zygula A, Szymusik I, Grzechocinska B, Marianowski P, Wielgos M. Endometrial injury for women with previous in vitro fertilization failure—does it improve pregnancy rate? *Neuro Endocrinol Lett* 2016;**6**:419–426.